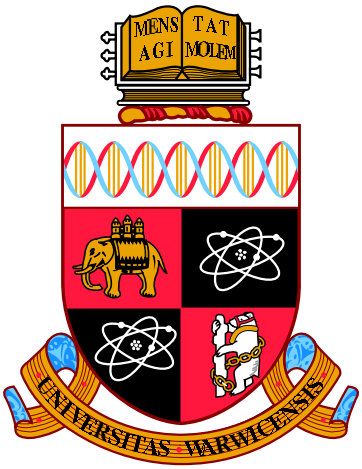
WARWICK
THE UNIVERSITY OF WARWICK



High Dynamic Range Imaging for Face Matching

by

Rossella Suma

Thesis

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

WMG

May 2020

Contents

List of Tables **vi**

List of Figures **vii**

Acknowledgments **ix**

Declarations **x**

Abstract **xi**

Acronyms **xiv**

Chapter 1 Introduction **1**

1.1 Motivation 1
1.2 High Dynamic Range Pipeline 3
1.3 Research Approach 6
1.4 Thesis Structure 7

Chapter 2 Visual Perception and Digital Imaging **9**

2.1 Human Visual System 9
2.2 Visual Adaptation 13
2.3 Characterisation of Visible Light 14
2.3.1 Colorimetry 17
2.4 Introduction to Digital Imaging 18
2.4.1 Sampling and Quantisation 18
2.4.2 Resolution 20
2.4.3 Intensity Transformations 20
2.5 HDR Imaging 21

2.6 HDR Capture 22
2.6.1 Exposure Bracketing 22
2.6.2 Camera Response Function 24
2.6.3 Native HDR Capture 26
2.7 HDR Storage 26
2.8 HDR Display 28
2.9 Tone-Mapping Operators 30
2.9.1 Photographic Tone Reproduction Technique 32
2.9.2 Optimal Exposure Algorithm 33
2.10 Expansion Operators for Low Dynamic Range Images 34
2.10.1 Inverse Tone-Mapping Operator 36
2.10.2 Machine Learning Techniques for Dynamic Range Expansion 36
2.10.3 ExpandNet Architecture 37
2.11 HDR Evaluation 38
2.12 TMO Evaluation: Subjective Experiments 39
2.13 TMO Evaluation Methods: Metrics 40
2.13.1 Tone Mapped Image Quality Index 40
2.13.2 Feature Similarity Index for Tone-Mapped Images 41
2.14 Summary 42

Chapter 3 Face Perception and Recognition 43
3.1 Models of Human Face Processing 43
3.2 Face Recognition: Familiar vs Unfamiliar 45
3.3 Face Matching 48
3.3.1 Facial Stimuli Datasets 48
3.4 Facial Identification: Context and Applications 50
3.5 Factors Affecting Face Recognition from Pictures 51
3.6 Neuropsychological Face Recognition Impairment: Prosopagnosia 54
3.7 Automatic Face Recognition 55
3.8 Summary 57

Chapter 4 Methodology **59**
4.1 Introduction 59
4.2 Critical Analysis and Research Gap 60
4.3 Proposed Approach 66
4.3.1 Warwick HDR Face Dataset 66
4.3.2 Subjective Evaluation of High vs Low Dynamic Range Imaging for Face Matching 66
4.3.3 Backwards-Compatibility HDR 68
4.3.4 Pipeline Comparison 68
4.4 Summary 69

Chapter 5 Warwick HDR Face Dataset **70**
5.1 Introduction 70
5.2 Design 72
5.3 Participants 74
5.4 Procedure 74
5.5 Materials 76
5.6 Dataset Properties 78
5.7 Discussion 79
5.8 Summary 80

Chapter 6 An Evaluation of High vs Low Dynamic Range Imaging for Face Matching **81**
6.1 Introduction 81
6.2 Methodology 82
6.3 Design 83
6.4 Participants 86
6.5 Procedure 86
6.6 Materials 87
6.7 Results for the HDR vs LDR Experiment 88
6.7.1 Multivariate Analysis 89
6.7.2 Univariate Analysis: *accuracy* 89
6.7.3 Univariate Analysis: *time* 91
6.7.4 Multivariate Analysis: *position: same* 91
6.7.5 Multivariate Analysis: *position: different* 92

6.8 Discussion 92
6.9 Summary 94

Chapter 7 Backwards-Compatible HDR Imaging for Face Matching 95
7.1 Motivation 95
7.2 Design 97
7.3 Participants 100
7.4 Apparatus and Materials 100
7.4.1 Tone Mapping Operator Selection 102
7.4.2 Expansion Operator Selection 105
7.5 Procedure 106
7.6 Results: Backwards-Compatibility HDR experiment 107
7.6.1 Multivariate Analysis 107
7.6.2 Univariate Analysis: *accuracy* 108
7.6.3 Univariate Analysis: *time* 108
7.6.4 Multivariate Analysis: *position: same* 109
7.6.5 Multivariate Analysis: *position: different* 110
7.7 Discussion 110
7.8 Summary 111

Chapter 8 Pipeline Comparison 112
8.1 Introduction 112
8.2 Design 113
8.3 Procedure 114
8.4 Results 115
8.4.1 Univariate Analysis: *accuracy* 119
8.4.2 Univariate Analysis: *time* 119
8.4.3 Multivariate Analysis: *DR*: *LDR* 120
8.4.4 Multivariate Analysis: *DR*: *HDR* 120
8.5 Discussion 121
8.6 Summary 122

Chapter 9 Conclusions 124
9.1 Overview 124

9.2 Objectives and Contributions 125
9.2.1 Warwick HDR Face Dataset 126
9.2.2 Subjective Evaluation of High vs Low Dynamic Range Imaging for Face Matching 128
9.2.3 Backwards-Compatibility HDR 128
9.2.4 Pipeline Comparison 130
9.3 Future Work 130
9.4 Final Remarks 132

Appendix A PI20: Self assessment questionnaire on face recognition ability 133

Appendix B Ethical Approvals 135

List of Tables

2.1 Most used radiometric quantities 16
2.2 Main photometric quantities 17

3.1 Factors helping or hindering face recognition. 47

4.1 Comparison of manipulations of image characteristics in stimuli adopted in experiments on unfamiliar face matching. 62

5.1 Average Dynamic Range for the Warwick HDR Face Dataset . 79

6.1 Correct Answers and Reaction Time Descriptive Statistics . . 89

7.1 TMO comparison using the TMQI, the FSTIM and the FSTIM&TMQI metrics 104
7.2 Correct Answers and Reaction Time Descriptive Statistics for the *Backwards-Compatibility HDR* 107

8.1 Descriptive statistic for the *Pipeline Comparison* experiment. . 116

A.1 PI20 test. 134

List of Figures

1.1 HDR Pipeline 4
1.2 Example of under-exposed (left) and over-exposed (right) images. 5

2.1 Diagram of the human eye. 10
2.4 Contrast sensitivity function for luminance and colour. 14
2.5 Grating stimuli adopted for luminance spatial discrimination. . 15
2.6 Photopic luminosity functions $V(\lambda)$ 16
2.7 Example of HDR obtained with exposure bracketing. 23
2.8 CRF recovery weights: Function for Robetson *et al.* vs Debevec and Malick's 25
2.9 HDR Images Visualisation. 28
2.10 Illustration of the HDR display dual-modulation. 29
2.11 Example of the Optimal Exposure algorithm. 33
2.12 Example of overexposed pixels in an LDR image. 35
2.13 ExpandNet architecture diagram. 37

3.1 Example of pixelation effect on faces. 51

4.1 Representation of the main stages of the HDR and LDR pipeline. 61
4.2 Roadmap of the general methodology adopted in this thesis. . 67

5.1 Warwick HDR Face dataset. Five lighting conditions. 73
5.2 Warwick HDR Face Dataset. Details of the experiment setup. 75
5.3 Warwick HDR Face Dataset. Example of the five different exposures. 76
5.4 Xyla-21 Chart 77

6.1 *HDR vs LDR* experiment. Example of the stimuli proposed. . 87

6.2 Schematic representation of the *HDRvsLDR* experiment setup. 88
6.3 Plot illustrating the *HDRvsLDR* experiment results. 90

7.1 Main stages of the HDR and LDR pipelines. 96
7.2 Mixed pipelines compared in the *Backwards-Compatibility HDR* experiment. 97
7.3 *Backwards-Compatibility HDR* experiment. Graph of the results for metrics used in the choice of TMO. 103
7.4 Graph with the results of the *Backwards-Compatibility HDR* experiment. 109

8.1 Diagram illustrating the *Pipeline Comparison* experiment. . . 115
8.2 Graph of the results for the *Pipeline Comparison* experiment. 117
8.3 Results for the *Pipeline Comparison* experiment for *DR*,*pipeline* and *position*. 118

Acknowledgments

This PhD journey has been one of the most rich and fulfilling experiences of my life and I am grateful to all that have made it possible.

I am incredibly thankful to Alan Chalmers, without whom I would most probably never have embarked in this adventure. I admire his cheerful attitude, his unflinching optimism and his sharp-wittedness, I am honoured to be his fiftieth PhD student. I would like to thank Kurt Debattista, for his constant guidance and support. I feel very privileged to have had such an honest and dedicated mentor. I would also like to thank Derrick Watson for his rich insight and guidance, every time I talk to him I learn something new! The list of my supervisors would not be complete without Liz Blagrove, her smile and energy have always been comforting especially through rough times.

What has made this journey really special are all the lovely people I have met along the way. A special thanks to Pinar, for her sincere friendship that has helped me to improve and grow; to Demetris for his good humour and a touch of sarcasm that has made the office life more bearable; and to Dani, for her enthusiasm, tenacity and the many adventures. Also, a big thank you goes to Debmalyia, Maria, Lvyin, Hells, Tom, Carlo, Aru, Kairul, Daniel, Mark, Amar, John, Josh, Tim, Magda, Ellie, Gina, Diego, George and Anaïs: it would not have been the same without you!

Finally, a heartfelt thanks to my parents and my brother for their unwavering support, love, generosity and for never failing to trust my capabilities. Last but not least, Manoj, his honesty, patience and warmth are what makes my heart feel light and brings a smile to my face.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.
Parts of this thesis have been previously published by the author in the following:

- Suma, R., Debattista, K., Watson, D. G., Blagrove, E., and Chalmers, A. Subjective evaluation of high dynamic range imaging for face matching. *IEEE Transactions on Emerging Topics in Computing*, December 2019

Abstract

Human facial recognition in the context of surveillance, forensics and photo-ID verification is a task for which accuracy is critical. In most cases, this involves unfamiliar face recognition whereby the observer has had very short or no exposure at all to the faces being identified. In such cases, recognition performance is very poor: changes in appearance, limitations in the overall quality of images - illumination in particular - reduces individuals' ability in taking decisions regarding a person's identity.

High Dynamic Range (HDR) imaging permits handling of real-world lighting with higher accuracy than the traditional low (or standard) dynamic range (LDR) imaging. The intrinsic benefits provided by HDR make it the ideal candidate to verify whether this technology can improve individuals' performance in face matching, especially in challenging lighting conditions. This thesis compares HDR imaging against LDR imaging in an unfamiliar face matching task. A radiometrically calibrated HDR face dataset with five different lighting conditions is created. Subsequently, this dataset is used in controlled experiments to measure performance (i.e. reaction times and accuracy) of human participants when identifying faces in HDR.

Experiment 1: HDRvsLDR ($N = 39$) compared participants' performance when using HDR vs LDR stimuli created using the two full pipelines. The findings from this experiment suggest that HDR (μ =90.08%) can significantly ($p<0.01$) improve face matching accuracy over LDR (μ =83.38%) and significantly ($p<0.05$) reduce reaction times (HDR 3.06s and LDR 3.31s).

Experiment 2: Backwards-Compatibility HDR ($N = 39$) compared participants' performance when the LDR pipeline is upgraded by adding HDR imaging in the capture or in the display stage. The results show that adopt-

ing HDR imaging in the capture stage, even if the stimuli are subsequently tone-mapped and displayed on an LDR screen, allows higher accuracy (capture stage: μ =85.11% and display stage: μ =80.70%), (p <0.01) and faster reaction times (capture stage: μ =3.06s and display stage: μ =3.25s), (p <0.05) than when native LDR images are retargeted to be displayed on an HDR display.

In *Experiment 3:* the data collected from previous experiments was used to perform further analysis ($N = 78$) on all stages of the HDR pipeline simultaneously. The results show that the adoption of the full-HDR pipeline as opposed to a backwards-compatible one is advisable if the best values of accuracy are to be achieved (5.84% increase compared to the second best outcome, p <0.01).

This work demonstrates scope for improvement in the accuracy of face matching tasks by realistic image reproduction and delivery through the adoption of HDR imaging techniques.

Sponsorships and Grants

This work was partially supported by:

- EU 7th Framework Programme for research technological development and demonstration under grant agreement No. 608013;
- WMG Doctoral Scholarship;
- the Rabin Ezra Scholarship Trust.

Acronyms

CIE Commission Internationale de l'Éclairage.

CMF Colour Matching Function.

CRF Camera Response Function.

CSF Contrast Sensitivity Function.

DP Developmental Prosopagnosia.

EM Elecro Magnetic (Radiation).

ERP Event-Related Potentials.

fMRI functional Magnetic Resonance Imaging.

FSTIM Feature Similarity for Tone-mapped IMages.

HDR High Dynamic Range.

HVS Human Visual System.

LDR Low Dynamic Range.

LFR Live Facial Recognition.

NIST National Institute of Standards and Technology.

PCA Principal Component Analysis.

PET Positron Emission Tomography.

RPE Retinal Pigment Epithelium.

SCR Skin Conductance Response.

TMQI Tone Mapping Quality Index.

Chapter 1

Introduction

Facial recognition is an activity routinely carried out by people on a daily basis. Police officers checking a passport, witnesses observing mug-shots or bank cashiers assessing identity, are all required to perform a face comparison task [Johnston and Bindemann, 2013; White *et al.*, 2014] with significant security implications. Although increasing technological support is provided, the ultimate decision-making relies on the human observer [Spaun, 2011]. When this task is performed in a context where the recognition is fulfilled or mediated by the adoption of imaging systems such as screens and cameras, the characteristics and limitations of the imaging adopted can be detrimental to the viewer performance [Norell *et al.*, 2015; Noyes and Jenkins, 2017]. Striving to achieve the highest accuracy possible in reproducing reality becomes paramount to avoid misclassified identification.

High dynamic range (HDR) imaging technology provides higher fidelity reproduction of real-world lighting and colours than the technology currently employed in surveillance and forensics contexts. In this thesis the capability of HDR to support and improve unfamiliar face recognition is explored.

1.1 Motivation

The problem of face recognition and, more specifically, unfamiliar face matching, is of great interest, not just from a psychological and perceptual point of view, but also from a broader perspective. CCTV is part of our daily experience, and it is common for witnesses or experts to be called to perform such identification

from video recordings [Norell *et al.*, 2015; Towler *et al.*, 2017b]. The social cost of erroneous identification can be serious, resulting in, for example, the wrongful imprisonment of innocent people [Norell *et al.*, 2015].

For human observers familiarity with a face depends on frequent and prolonged exposure to it, including viewing under different lighting conditions and from differing viewpoints. Additionally, the observer has generally seen that face displaying a variety of emotional expressions, has seen it in motion and has associated semantic information to it (e.g. name) [Hole and Bourne, 2010]. When the observer has had minimal or no "experience" of a face, this is termed *unfamiliar*. The literature on face perception, discussed in Chapter 3, shows how this phenomenon of unfamiliarity is susceptible to poor identification performance [Towler *et al.*, 2017a].

Psychological experiments on face recognition adopt the generic term performance to qualify whether the recognition has been successful or not. More specifically, this is generally evaluated in terms of two parameters: *accuracy* and *reaction time*. Unfamiliar and familiar face recognition task have been studied with a variety of experiments trying to identify the most perceptually influential information for the purpose of recognition. Within this corpus of literature, a subset of experiments has emerged: face matching, as opposed to face recognition. The latter involves memory recollection (e.g. eyewitness testimony), while face matching consists in comparing simultaneously-presented facial stimuli, if the faces are unknown to the viewer this is termed unfamiliar face matching.

Some studies have pointed out how, in specific scenarios, even professionals, such as trained or experienced police officers are not any better than non-specialists [White *et al.*, 2014] in performing a face matching task. This suggests an intrinsic limitation in human performance related to this activity which is further exacerbated by factors such as poor quality of the imaging or variations in lighting conditions [Braje *et al.*, 1998; Longmore *et al.*, 2008; Bindemann *et al.*, 2013; Noyes and Jenkins, 2017].

Specifically, many researchers have underlined the huge impact that illumination has on performance [Johnston *et al.*, 1991; Longmore *et al.*, 2008; Hermens and Zdravković, 2015]. To the best of our knowledge, however, no study to-date has investigated the impact of the image's dynamic range on

face matching. The following section defines this concept in detail.

One of the most common trends in recent years in face recognition is the adoption of machine learning and computer vision technology. Yet, this technology is not considered fully mature and its effective adoption in real-life scenarios remains controversial. For example, Fussey and Murray [2019] recently released a report on test deployments by the London Metropolitan Police Service (MPS) trialling live facial recognition (LFR). They shadowed the police force during trials in which computer vision systems were devised to alert the operator of the presence of individuals of interest. Fussey and Murray's report advocates for the halt to the use of this technology, emphasising how the police deployment of LFR technology was unlikely to comply to human rights' laws and would *"be held unlawful if challenged before the courts"*. Only a fifth of the matches found by the system turned out to be correct. Moreover, they stressed the importance of the "human factor" as stated in the Surveillance Camera Code of Practice [Great Britain, 2013] :

> *"any use of facial recognition [...] should always involve human intervention before decision are taken that affect an individual."*

According to Spaun [2009], in the case of facial biometrics, the request for specialised professional examiners to supervise the automated process will be increasing in the next few years, especially in the forensic context, similarly to what has happened in the field of fingerprint biometrics, where the community has established human examiners to be the ultimate verifiers of the output of automated systems. Hence, supporting the facial recognition process by humans with the best technology available becomes increasingly important.

1.2 High Dynamic Range Pipeline

The dynamic range of an image is defined as the difference between the largest and the smallest value of luminance[1] [Mantiuk *et al.*, 2016]. Depending on the application several measures are used.

The human visual system is capable, without adaptation, of perceiving a luminance range of $10^4 : 1$ [Banterle *et al.*, 2017]. The perceivable dynamic range

[1]Luminance is a photometric quantity that defines the brightness of light emitted or reflected off a surface that is perceived by the human eye [Eppig, 2016].

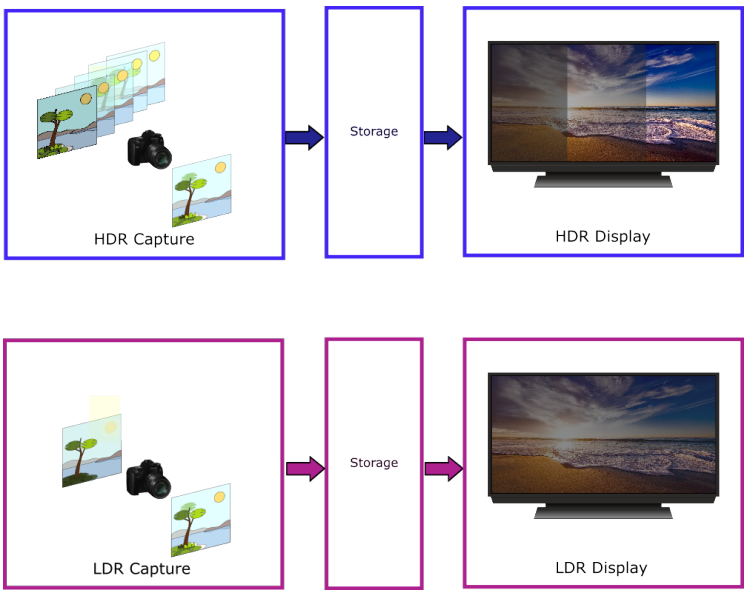


Figure 1.1: Simplified representation of the LDR and HDR pipeline. The technique represented in the HDR Capture stage is called exposure-bracketing, see Chapter 2 for further details.

is specific to the scene being observed. Multiple studies have shown people's preference for images with a dynamic range higher than 1000:1 [Mantiuk *et al.*, 2016].

The imaging pipeline currently adopted in the context of unfamiliar face matching both in lab experiments and real-life scenarios, represented in Figure 1.1), is termed Low (or Standard) Dynamic Range (LDR). The pipeline is subdivided in three stages: capture, storage and display.

In the capture stage, LDR devices capture images that are non-linearly related to photometric or colorimetric values. More importantly, these capture devices are inadequate to capture the whole range of luminance in the scene, producing images that are over- or under- exposed [Mantiuk *et al.*, 2016] (see Figure 1.2). Moreover, when it comes to displays, the maximum brightness available with most modern off-the-shelf LCD screen is around 500 cd/m^2, which is not nearly comparable to the levels of lighting we experience in our daily life. For example, outdoor lighting levels can be significantly in excess of 10,000 cd/m^2 [Banterle *et al.*, 2017]. The limitations intrinsic to LDR imaging thus make it susceptible to poorly reproduce reality and possibly affect performance when users are called to perform a recognition task.

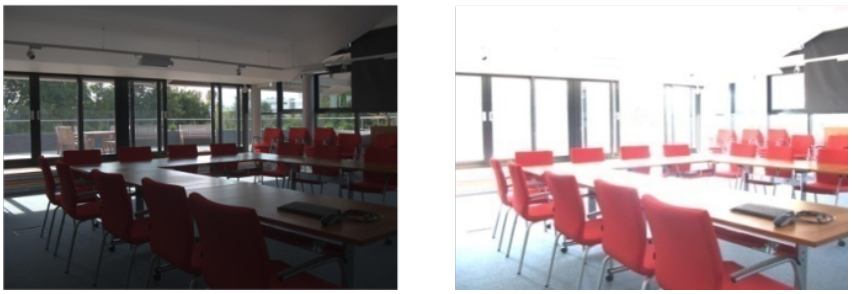


Figure 1.2: Example of under-exposed (left) and over-exposed (right) images.

HDR imaging allows encoding all the colours and brightness levels visible to the human eye [Myszkowski *et al.*, 2008]. This makes HDR an ideal candidate to investigate whether it could be capable of providing observers more perceptually relevant information, and thus enable them to perform better at an unfamiliar face matching task.

The main stages of the HDR imaging pipeline are also illustrated in Figure 1.1. For real-world capture, the capture stage of the HDR pipeline, is achieved by combining multiple exposures from a standard LDR camera or by using specialist HDR cameras. The whole luminance range in the scene is captured and the values encoded in the image are linearly related to the luminance values. In addition, the pipeline allows for visualisation of the content on screens that can achieve brightness levels of up to 10,000 cd/m^2, which corresponds to levels of an average outdoor scene [Hatchett *et al.*, 2019].

The superiority of the HDR pipeline is not just limited to the width of the luminance range that is possible to capture, store and display. Its true potential lies in the capability of representing a far higher number of luminance levels than the LDR pipeline. This is due to the way the images are encoded. The HDR formats allow for the use of real numbers ($\mathbb{R} : [0, \infty]$) as opposed to integers over the limited range ($\mathbb{N} : [0, 255]$) which is the case for LDR formats. Some aspects in modern display technology exceed visual acuity, for example, in terms of refresh rate and spatial resolution [Chalmers and Debattista, 2017], but standard screens still struggle with reproducing shades of grey and brightness to that which we experienced in our daily life.

HDR imaging has a wide range of applications, such as feature detection and people tracking [Agrafiotis *et al.*, 2015] or photogrammetry [Suma *et al.*, 2016], as well as rocket launches [Karr *et al.*, 2016]. Furthermore, the application

of HDR is not limited to general photography. It has shown great potential, for example, in enhancing medical or biological imaging capture techniques, such as the HDR-OPT system developed by Fei *et al.* [2012], or the enhancement of real-time laser scanning fluorescence microscopy [Vinegoni *et al.*, 2016]. Furthermore, Marchessoux *et al.* [2016] have shown the benefits of the adaption of HDR displays in radiology applications, as the extended dynamic range allows doctors to identify difficult breast lesions. The impact of HDR imaging on face matching is explored for the first time in this thesis.

1.3 Research Approach

In order to address the challenges related to face recognition and the detrimental impact of poor lighting reproduction, this thesis investigates whether advanced imaging technology can significantly improve recognition performance. In particular, this research work attempts to answer the following research question:

Does HDR imaging improve unfamiliar face matching performance?

Given the complexity of the topic, this research was undertaken in several steps. Specifically, five main objectives were identified:

1. Review of the literature regarding both the HDR technology pipeline and the processes underlying face perception (Chapter 2 and Chapter 3);

2. Creation of a radiometrically calibrated HDR faces dataset: the Warwick HDR Face Dataset (Chapter 5);

3. Use of the Warwick HDR Face Dataset stimuli in a face matching experiment aimed at measuring accuracy and reaction times performance of HDR vs LDR imaging (Chapter 6);

4. Adoption of the Warwick HDR Face Dataset stimuli in a face matching experiment aimed at measuring accuracy and reaction time performance of an enhanced LDR pipeline through the application of advanced imaging algorithms - i.e. Expansion Operators and Tone Mapping Operators (Chapter 7).

5. Comparison of the results obtained in the Objectives 3 and 4 to establish if the adoption of the full HDR pipeline or a partial upgrade of the LDR pipeline is adequate for sufficiently improving performance (Chapter 8).

1.4 Thesis Structure

This thesis has been structured as follows:

Chapter 2 introduces the basic processes related to visual perception, presents elements of digital imaging and explains in detail each of the components of the HDR imaging pipeline.

Chapter 3 describes the literature related to face perception. The impact of the characteristics of the imaging stimuli is highlighted and the elements relevant to identifying significant gaps in the knowledge are outlined.

Chapter 4 describes the choices made to address the highlighted limitations in the current body of knowledge, linking the perceptual investigations with the impact of imaging characteristics of the stimuli adopted.

Chapter 5 describes the methodology adopted for the creation of the first HDR faces dataset and illustrates the approach followed during the preparation and image acquisition stages.

Chapter 6 recounts the steps taken for the preparation and execution of a subjective experiment. The experiment, aimed at evaluating the potential of the HDR pipeline in addressing human limitations in performing a face matching task, is described. The performance achieved using the HDR and the LDR pipeline are compared.

Chapter 7 reports a new subjective experiment aimed at comparing partial upgrades of the LDR pipeline to HDR respectively in the capture and display stage.

Chapter 8 describes another experiment based on the analysis of the results obtained from the experiments described in Chapter 6 and Chapter 7. This experiment intends to answer the question of whether a complete upgrade of the imaging pipeline to HDR is necessary, or a partial upgrade would suffice to improve performance in the unfamiliar face matching task.

Chapter 9 concludes the main contribution of this work, outlines the limitations and provides an overview of the possible future developments of this research work.

Chapter 2

Visual Perception and Digital Imaging

To address the research question, it is necessary to have an understanding of the mechanisms that characterise vision, the digital imaging pipeline and the cognitive face recognition processes. Chapter 3 will provide a detailed review of the literature on face recognition, while this chapter will present aspects of visual perception and will provide an introduction to digital imaging. Emphasis will be given to the HDR imaging pipeline, focusing on elements that will be relevant for their application in the following chapters.

2.1 Human Visual System

One of the way humans perceive their surroundings is through their eyes. These are very complex organs through which light from the environment gets collected, transformed and transmitted to the brain through the optic nerve.

The eye structure can be simplified by defining it using two segments: the anterior and the posterior segment [Cholkar *et al.*, 2013]. The anterior segment consists of the cornea, conjunctiva, aqueous humour, iris, ciliary body and crystalline lens, while the posterior segment includes sclera, choroid, Bruch's membrane, retinal pigment epithelium (RPE), neural retina and vitreous humour.

The cornea (see Figure 2.1) is a transparent tissue and the most external part of the eye. Together with the sclera, it provides protection and structural

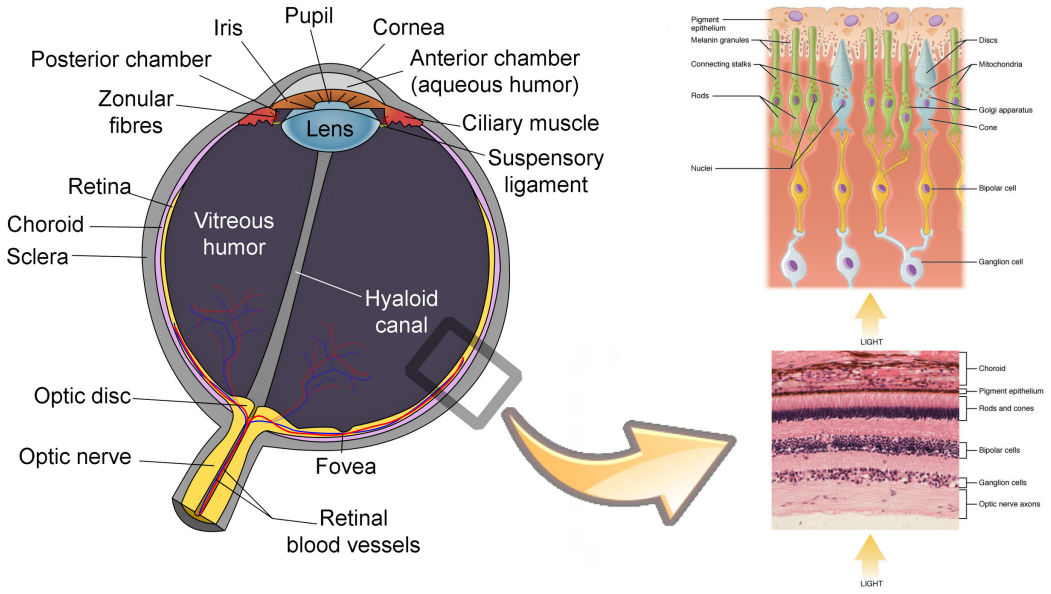


Figure 2.1: Diagram of the human eye (left) and of the transversal section of the retina (right) (Wikimedia Commons).

integrity [Treuting and Dintzis, 2011]. The cornea acts as an interface between the air and the eye. Its refractive index characteristics can lead to sight problems such as myopia, astigmatism or hyperopia [Fairchild, 2013].

The iris can be considered, in the structure of the eye, the equivalent of a camera diaphragm. The characteristic iris's opening, separating the anterior and the posterior segments of the eye, is called pupil. The size of the pupil regulates the amount of light entering the eye [Forrester *et al.*, 2015]. Its size is determined by the incoming light, but can also be influenced by non-visual phenomena, such as arousal [Fairchild, 2013].

The lens, also called crystalline, lies behind the iris and in front of the vitreous body. Its main role is to alter the refractive index of light entering the eye and focus it on the retina by changing shape under the influence of the ciliary muscle. As age progresses the crystalline loses its ability to change shape making it more difficult to focus on nearby objects, or it becomes more opaque due to UV or infrared damage causing loss of clear vision. The retina is a thin layer of cells located in the focal plane of the eye's optical system and converts the incoming light into electrical impulses that are transmitted to the brain. The two regions in which the retina is usually subdivided are called

peripheral and central. In particular, the peripheral area is mostly oriented towards movement detection and rough shapes perception, while the central area is highly specialised for visual acuity [Remington, 2011].

Figure 2.1 illustrates representation of the radial section of a portion of the retina. Light must travel through the thickness of the retina before striking and activating the rods and cones, which are photoreceptor cells located in the outer retina.

Photoreceptor cells are responsible for measuring the amount of light reaching the eye. Depending on the intensity and on the wavelength of the incoming light each class of photoreceptor has a higher or smaller probability that an isomerisation[1] of the 11-cis retinal molecule might happen [Stockman and Brainard, 2015]. Said isomerisation is the chemical trigger for the generation of neural signals that will form the sensation of vision in the brain.

There are three different classes of light-sensitive (i.e. photoreceptor) cells called rods, cones plus the ipRGCs (intrinsically photosensitive retinal ganglion cells). The photosensitive ipRGCs receptors are a rather recent discovery [Hattar *et al.*, 2002]. They are excited by light even when all influences from cones and rods is inhibited. This class of photoreceptors plays an important role in vital functions, such as regulation of the circadian rhythm [Stockman and Brainard, 2015].

The other two classes of photoreceptors that contribute to conscious vision are named rods and cones due to their prototypical shape. They transform electromagnetic energy in the visible range (380-830nm) into an electrochemical signal. This signal will cascade into electrochemical activity throughout the optic nerve and the rest of the visual system, generating what we call vision [Gegenfurtner and Ennis, 2015].

The rods and cones disposition in the retina is not uniform. The majority of cones is located in the fovea, a rather central circular area of the retina covering an area of about 1.5mm diameter, while the rods occupy a more peripheral area.

The rods are sensitive to dim light, providing a black and white vision called *scotopic*. Their sensitivity has its peak at wavelengths (indicated with λ)

[1] *isomerisation* is the transformation of a molecule into another molecule with the same atoms, but having them differently arranged

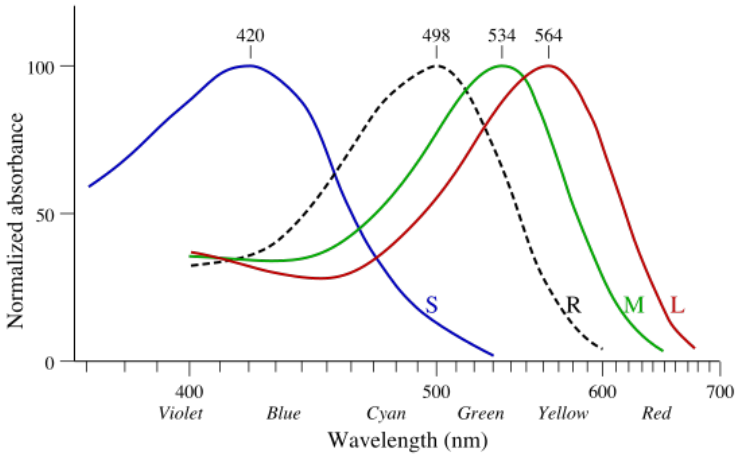


Figure 2.2: Simplified human rods and cone response curves. Each spectral absorption curve represents respectively the short (S), medium (M) and long (L) wavelength pigments in human cone and rod (R) cells. Image courtesy of Maxim Razin (Wikimedia Commons), based on Bowmaker and Dartnall [1980].

of approximately 510nm. In typical daylight intensities (from dawn until dusk) the rods are completely saturated, leaving only to the cones the task of vision and perception of colour. The activation of cones is called *photopic* vision. We call *mesopic* light levels when both rods and cones are active.

At the foveal level, three classes of cones are present. Each one possessing different sensitivity to light and determining what is called trichromatic colour vision. The different classes are referred as long (L), medium (M) and short (S) cones, depending on their peak spectral sensitivity [Stockman and Sharpe, 2000] (See Fig. 2.2). The cones act as transducers, transforming the light that has entered through the pupil into the three-dimensional LMS space: colour vision is therefore said to be trichromatic, as any spectral composition can be matched by a mixture of three primary lights [Gegenfurtner and Ennis, 2015].

The spectral sensitivity function describes the relationship between wavelength and the cones' sensitivity. As stated before, the three classes of cones are named to indicate their peak sensitivity within the visible spectrum. As shown in Figure 2.2, the breadths of the three sensitivity curves overlap. This causes different combinations of light information to appear identical. Such pairs are known as *metamers* [Stockman and Brainard, 2015].

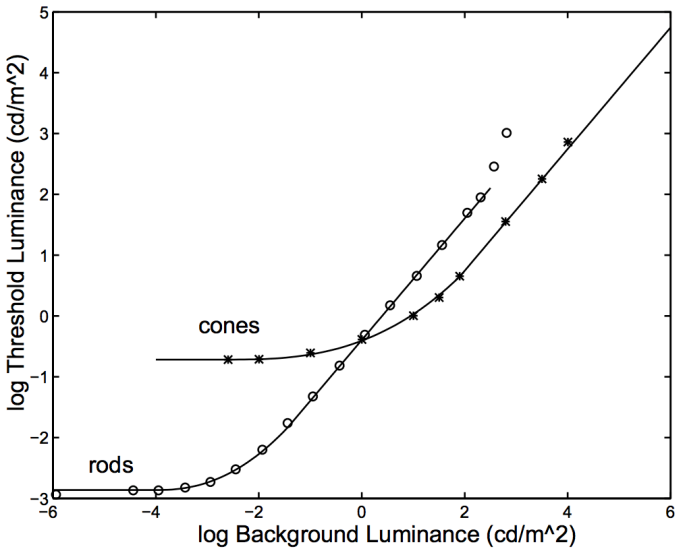


Figure 2.3: Rods and cones threshold versus intensity (TVI) response. Image courtesy of Ferwerda *et al.* [1996]

2.2 Visual Adaptation

The human visual system is able to deal with a luminance range of nearly 14 log units through a process called *adaptation* [Ferwerda *et al.*, 1996]. Depending on the illumination level and the time-frame, the adaptation will entail specific quality of threshold visibility, colour appearance and visual acuity.

The diagram of threshold visibility in Figure 2.3 shows the relationship between background luminance levels and the minimum smallest detectable increment. The experiments performed to obtain these levels are based on the detectability of light spots while looking at a uniformly lit background. For a wide range of background luminance levels, the size of the minimum increment $\Delta L = L_{disk} - L_{background}$ is proportional to the ambient luminance L, following *Weber's law*: $\Delta L/L = k$, where k is a constant term.

However, the simple design of the brightness discrimination experiments makes it hard to generalise in the case of complex scenes. In turn, this has inspired experiments adopting sinusoidal grating patterns, aimed at discovering spatial and temporal properties of the visual system.

Visual sensitivity is defined as the inverse of the threshold contrast necessary to elicit a response in the visual system: $S = \Delta L/L$. The Contrast

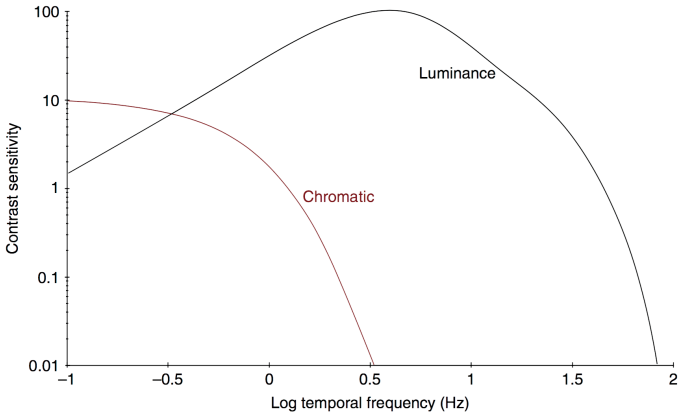


Figure 2.4: Threshold contrast sensitivity function for luminance and colour grating stimuli. Adapted from Fairchild [2013]

Sensitivity Function (CSF) describes the relation between visual sensitivity as a function of a harmonic stimulus with variation in time or space [Westland *et al.*, 2006; Fairchild, 2013].

If the stimulus is temporal, the experiment consists in the detection of a flickering signal (sinusoidally modulated) over a constant average light intensity background. Different responses are registered depending if the stimulus is black and white or chromatic, as illustrated in Figure 2.4.

In the spatial case, the experiments consist of the regular alternation of bright and dark bars (see example in Figure 2.5). The threshold is established by adjusting a grating input of a given spatial frequency until it becomes barely detectable [Campbell and Robson, 1968]. The inverted U-shaped upper margin is the observer contrast sensitivity function. The exact location of the peak depends on the viewing distance [Campbell and Robson, 1968]. The reason for this band-pass behaviour is neural in nature for the lower frequencies, while, for high spatial frequencies its decline is due to the eye's optics spatial limitations in the cone receptors mosaic [Fairchild, 2013].

2.3 Characterisation of Visible Light

Visible light is a section of the electromagnetic spectrum and it is the means through which the human visual system perceives its surroundings. The range

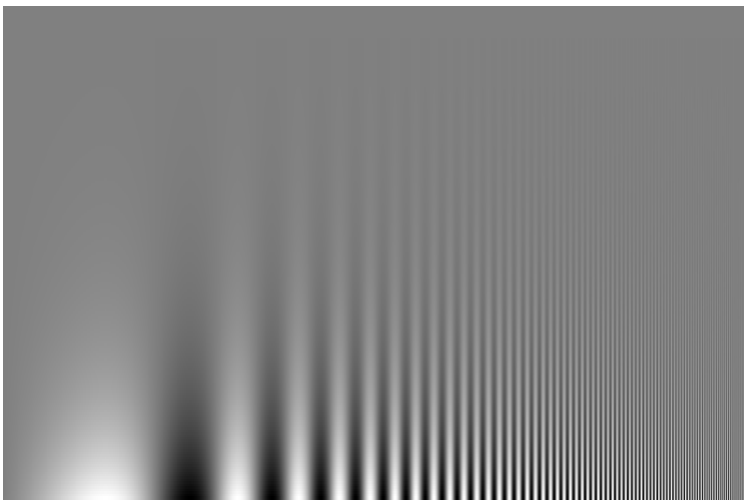


Figure 2.5: Example of the stimuli used in luminance spatial discrimination experiments, the reader will realise that the shape of the curve will differ, depending of the distance at which the image is observed. Image courtesy of Campbell and Robson [1968].

of perceivable wavelengths (λ)is between 380nm and 830nm [Walraven and Walraven, 1999].

What can we measure? Radiometry is the science that measures electromagnetic (EM) radiation, including visible light. These measurements are independent of the properties of the HVS (Human Visual System). Radiant energy is the energy of EM radiation and is measured in joules (J). Table 2.1 presents the definitions of the most commonly used radiometric quantities.

Photometry and colorimetry, on the other hand, focus on quantifying the luminous stimulus taking into account its effect on the HVS. Photometry is a special subset of radiometry where the radiometric quantities are weighted by a standard-human-eye response [Shevell, 2003], while colorimetry describes the EM radiation in terms of its effect of retinal cone photoreceptors (short, medium and long wavelength). This topic is further discussed in Section 2.3.1.

The photopic luminous efficiency curve function (also called $V(\lambda)$) has been defined by the CIE (Commission Internationale de l'Eclairage) in 1931

Table 2.1: Most used radiometric quantities

Radiometric Unit	Definition	Unit
Radiant Energy	EM radiation, Ω_e	$[J]$
Radiant Power	$P_e = \frac{\Omega_e}{dt}$	$[W] = [J/s^{-1}]$
Radiant Intensity	$I_e = \frac{dP_e}{d\omega}$	$[W/sr]$
Irradiance	$E_e = \frac{dP_e}{dA_e}$	$[W/m^2]$
Radiance	$L_e = \frac{d^2 P_e}{dA_e \cos\theta d\omega}$	$[Wm^2/sr]$

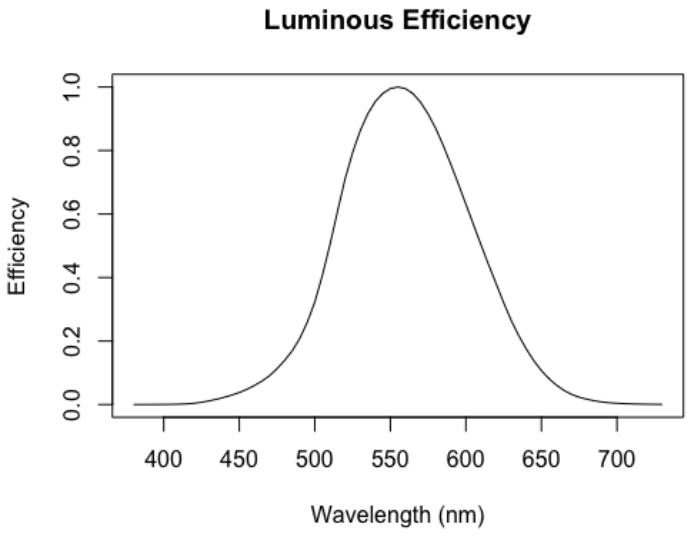


Figure 2.6: Photopic luminosity functions $V(\lambda)$

[CIE, 1931] and subsequently supplemented and revised [Walraven and Walraven, 1999]. The curve shown in Figure 2.6 is a transfer function which approximates the behaviour of the average human eye depending on the wavelength (λ) of the stimulus. Luminance, which is a photometric quantity, is obtained following the equation [Reinhard *et al.*, 2010]:

$$L_v = \int_{380}^{830} L_{e,\lambda} V(\lambda) d\lambda, \tag{2.1}$$

where the wavelength (λ) ranges from 380nm and 830nm. Table 2.2 presents the definitions of the main photometric units, where the main radiometric quantities are weighted by $V(\lambda)$.

Table 2.2: Main photometric quantities, they are the equivalent of radiometric quantities weighted by the CIE standard observer luminance response.

Photometric Unit	Symbol	Definition	Unit
Luminous Energy	Q_v	Amount of Light	$[lm\ s]$
Luminous Power	$\Phi_v = \frac{Q_v}{dt}$	Amount per unit time	$[lm]$
Luminous intensity	$I_v = \frac{\Phi_v}{d\omega}$	Amount per unit time per solid angle	$[cd] = [lm/sr]$
Luminance	$L_v = \frac{\Phi_v}{dA_e}$	Amount per unit time per unit area	$[cd/m^2]$ also $nits$

2.3.1 Colorimetry

Colorimetry is the discipline concerned with the measurement of colour [Fairchild, 2013] and can predict whether two visual stimuli of different spectral distributions will match when presented to an observer. What humans perceive as colour is an attribute of the visual sensation generated by the concomitant presence of EM radiation and the physical and chemical properties of the object upon which the EM radiation bounces [Fairchild, 2013]. Therefore, colour is not directly accessible for measurement, the only measurable element is the stimulus that triggers a specific colour perception. CIE colorimetry is a

standard providing a metric for psychophysical colour stimulus. The standard defines the additive RGB colour model, where, every visible colour is defined as the linear combination of the standardised three primaries: Red, Green and Blue, hence the acronym RGB. This colour model is adopted by display systems, although Fairchild [2007a] points out that this representation needs to be combined with a device characterisation or profile to be considered accurate.

2.4 Introduction to Digital Imaging

The digital image acquisition is the process through which the incoming light from a scene is measured, generally conveying it through the use of a lens, by an imaging sensor. The lens allows for the collection of the incoming light and the focusing of the rays onto a focal plane. The sensor, located on the focal plane of the lens, integrates the incident radiation into a voltage value. The sensor consists of a 2D array of CCD (Charged Coupling Device) or CMOS (Complementary Metal Oxide Semiconductor).

The intensity of light collected form each element of the array is, in turn, digitised (generally into 8 to 14 bits) through sampling and quantisation [Gonzalez and Woods, 2018]. Each one of these signals will then be effectively considered the value for the single element of the array, called pixel, $p(i, j)$ where i and j are the spatial coordinates. Each CCD or CMOS has a minimum and a maximum threshold which will guide its response to light. Increasing or decreasing the exposure time allows controlling the number of photons being collected. The appropriate use of filters on the sensor array will allow the light to be differentiated into the three colour channels RGB - the most used filter is called Bayer filter [Bayer, U.S. patent 3971065].

2.4.1 Sampling and Quantisation

Sampling and quantisation are the main processes involved when trying to capture images with a digital device [Gonzalez and Woods, 2018]. The definitions presented in this section are carried out for a single channel or greyscale image, but they can be translated easily into a colour version by replicating the procedure for each of the single colour channel (i.e. RGB).

Each of the sensors of the 2D array will produce an output voltage which is proportional to the amount of incoming light. The sampling process s consists of mapping the continuous values of x and y (spatial coordinates) of the voltage signal $f(x, y) \in \mathbb{R}^2$ into a discrete array. The array indices $\hat{x} = 0, 1, 2, ..., M - 1$ and $\hat{y} = 0, 1, 2, ..., N$ with M and N are positive integers as in the following equation:

$$s : (x, y) \in \mathbb{R}^2 \Longrightarrow (\hat{x}, \hat{y}) \in \mathbb{N}_0^2. \tag{2.2}$$

The quantisation process q involves the discretization of the intensity levels f for each of the spatial tuple $f(\hat{x}, \hat{y})$ as follows:

$$q : f(\hat{x}, \hat{y}) \in \mathbb{R} \Longrightarrow \hat{f}(\hat{x}, \hat{y}) \in \mathbb{N}_0, \tag{2.3}$$

where the number of intensity levels is typically an integer power of 2, due to the nature of the hardware (i.e. $L = 2^k$). L is an indicator of the Dynamic Range of the imaging system, which will be defined in Section 2.5.

A digital image is a 2D-array of size $M \times N$ containing the result of the sampling and quantisation processes. An intensity level proportional to the incoming light for each distinct pair of spatial coordinates (x, y) is assigned to each element of the 2D-array (i.e. pixel $p(i, j)$).

If the incoming light causes the maximum charge transfer capacity of the CCD to be reached, the resulting channel value (i.e. RGB) will take the maximum intensity level (i.e. the sensor has reached saturation). When several pixels reach saturation in all three channels, the area of the image is called *overexposed* and if all the channels are overexposed the pixels will appear pure white [Robins and Bean, 2003]. If, on the other hand, the incoming light is not enough to be detected by the CCD (i.e. it is below the noise threshold), the resulting channel is called *underexposed* and the pixels appear completely black. In both cases, the original detail of the image is lost.

LDR images use 8-bits per colour channel (i.e. $k = 8$), allowing to describe a single pixel through the linear combination of RGB values. Each channel is therefore characterised by 256 levels (i.e. $L \in [0, 255]$).

2.4.2 Resolution

The spatial resolution of a digital image is generally intended as an indicator of its smallest perceptible detail. Although commonly the resolution of an image is provided as the number of pixels that form it, this is not a well-posed statement. While it allows for relative comparison between images, it does not provide any reference to spatial units. If the images are printed, a common measure of the image resolution is provided as *dots (pixels) per inch* [Gonzalez and Woods, 2018].

In regard to displays, the term *resolution* will indicate the number of pixels along each dimension. In addition to the resolution, some important elements to consider are the display size (generally expressed as diagonal size) and the viewing distance. These two elements will determine how "densely packed" the pixels appear on the screen from a certain distance. The Recommendation ITU-R BT.2022 defines the optimal viewing distance as "the distance at which two adjacent pixels subtend an angle of 1 arc-min at the viewer's eye".

2.4.3 Intensity Transformations

Digital images can be processed in many ways. Amongst those, intensity transformations are some of the simplest. Some of them operate globally on all pixels transforming their intensity level.

Image Negatives The aim of a negative transformation t' is to invert the intensity level t for the pixel $p(i, j)$. If the image has k intensity levels within the range $[0, L-1]$, the transformation is expressed as:

$$t' = L - 1 - t. \tag{2.4}$$

The output of this processing is the equivalent of a photographic film negative and is usually applied in cases where it is desirable to enhance small light grey details that are immersed in dark regions. This kind of manipulations have also been used in face perception experiments (see Chapter 3).

Power-Law (Gamma) This type of transformation assumes the form:

$$t' = c\, t^{\gamma}, \tag{2.5}$$

where c and γ are positive constants and t and t' are in the range $[0, 1]$. As previously illustrated, the human visual system has a non-linear response to light stimuli. When encoding an image, a gamma transformation might be applied ($\gamma < 1$), so that the usage of bits is optimised to a perceptually uniform domain [Poynton, 1998], this is termed encoding gamma. When displaying these images an inverse gamma correction has to be applied to compensate for gamma correction. Typically, standard display apply γ =2.2 or 1.8, this is termed decoding gamma.

2.5 HDR Imaging

High Dynamic Range (HDR) imaging is a complex pipeline that enables to capture, store, transmit and display images described by 32 bits floating point per colour channel, offering the ability to represent real world colours and lighting with higher accuracy reproducing more accurately the reality as perceived by the HVS [Mantiuk *et al.*, 2016].

As described in 2.4.1, dynamic range is defined by the ratio between the highest and the smallest value of luminance that is possible to encode in an image, although, depending on the context, other definitions are used in the literature [Myszkowski *et al.*, 2008]. In particular, it is called Contrast Ratio when measuring displays' luminance values and is defined as:

$$CR = \frac{L_{peak}}{L_{noise}} \tag{2.6}$$

where L_{peak} represents the maximum and L_{noise} the minimum luminance value, defining it by the ratio between brightest (white) and darkest(black) reproducible colour. In this context the term saturation indicates the upper limit, while noise determines the lower limit.

For HDR photography, dynamic range (also called log exposure range) is measured in stops (i.e. f-stops), which is the CR expressed in a logarithmic

scale as:

$$DR = log_2(L_{peak}) - log_2(L_{noise}) \tag{2.7}$$

where L_{peak} and L_{noise} are, respectively, the maximum and the minimum pixels' intensity levels.

2.6 HDR Capture

Two major sources of HDR content are available today: computer-generated scenes and photographically captured real scenes. Since the objective of this investigation is strictly related to the reproducibility of reality, this thesis will address only the latter. The photographic result can be achieved by using traditional LDR cameras, or *ad hoc* HDR cameras.

2.6.1 Exposure Bracketing

Exposure bracketing, as proposed by Mann and Picard [1995], is one of the most commonly adopted technique to create HDR images. It requires, usually, a tripod and an LDR camera and consists of taking a sequence of pictures from the same spot with identical settings except for the exposure values. In this way, it is possible to capture the entire scene luminance range. Depending on the dynamic range of the scene, a number (N) of exposures are required (see Figure 2.7), $ldr_i, i \in \{1, .., N\}$.

Following the capture stage, the single different exposures will be merged into a single HDR image. Assuming that the camera sensor behaves linearly, the most simplistic approach to generate an HDR image is to bring the different exposures into the same domain by dividing each pixel $p_k(i, j)$ by the appropriate exposure time. Each pixel $p_{HDR}(i, j)$ in the final image will be calculated following the equation:

$$p_{HDR}(i, j) = \frac{\sum_{k=1}^{N} \frac{p_k(i,j)}{\Delta t_k}}{N} \quad , \tag{2.8}$$

where $p_k(i, j) \neq p_{max} \wedge p_k(i, j) \neq p_{min}$. The pixel $p_{HDR}(i, j)$ is equal to a weighted average of the N different exposures, where Δt_k is the exposure time for the k-th exposure. Overexposed (p_{max}) and underexposed (p_{min}) pixels are

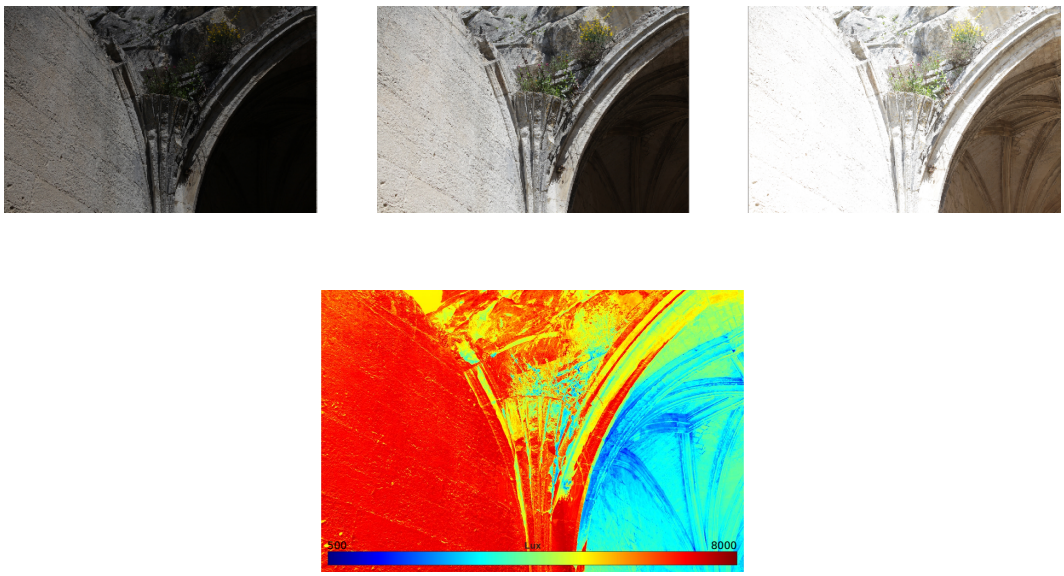


Figure 2.7: The Château des Baux, France. Three different exposures of the scene and a false colour luminance map. The blue pixels in the map represent low luminance values, while red indicates the maximum luminance in the scene. HDR is useful in case of harsh lighting conditions, where strong shadows and bright sunlight coexist in the same scene.

not taken into account when averaging.

However, several limitations and non-linearities occur when this process is executed in reality. For example, corrections introduced by the camera manufacturer in order to obtain images more appealing to the user (if the RAW[2] format is not accessible), noise on the sensor due to heat or lenses' artefacts or even just micro-movements between the different shots can generate a mismatch in the pixel coordinates [Reinhard *et al.*, 2010]. Therefore, when generating an HDR image all of the following procedures are applied: feature extraction, feature matching, alignment, camera response function derivation, ghost removal, flare removal, noise removal.

The feature extraction, feature matching and alignment are steps aiming at identifying corresponding pixels among the different exposures - in case of misalignment. The ghost removal, flare removal and noise removal are procedures aiming at the removal of artefacts caused by small movements, lenses distortion respectively or sensor's noise. A complete review of these

[2]The proprietary RAW format contains the camera sensor's data with minimal or no manipulation by the onboard software.

topics can be found in [Reinhard *et al.*, 2010].

2.6.2 Camera Response Function

Deriving the camera response function (CRF) allows recovering the radiance measured by the camera depending on the exposure time. Since almost no instrument has a perfectly linear response recording the scene radiance [Mitsunaga and Nayar, 1999], several solutions exist in literature.

Mann and Picard [1995] offered a simple solution where the CRF assumes the form of a gamma curve, however this method does not support most real CRFs [Banterle *et al.*, 2017].

Debevec and Malik [1997] improved on the solution proposed by Mann and Picard assuming a semi-monotonic trend of the CRF. Considering several samples (i.e. pixels), the algorithm finds the curve that minimises the mean-square error, weighing the contribution of different exposures proportionally to the distance from the mean. Higher weight is attributed to input data closer to the mean of the input pixel range (128 for 8-bit data), and lower weight is assigned to the input data that is near the extremes of the input pixel range (close to 0 and 255 for 8-bit data). This algorithm is well suited when the pixels are not noisy and precise exposure times are available [Ward, 2003].

Mitsunaga and Nayar [Mitsunaga and Nayar, 1999] proposed a parametric model to approximate the CRF with a polynomial function of arbitrary order P and coefficients c_k, $k \in 0, ..P$. Calibration, could be viewed as determining the order P in addition to the coefficients. This solution is more tolerant to noisy images and to unavailability of exposure times.

Another algorithm commonly adopted to recover the CRF is the one presented by Robertson *et al.* [1999] and later improved by Robertson [2003]. In this algorithm the CRF is modelled as a random Gaussian variable, while fixing *a priori* the weighting function - similarly to Debevec and Malik [1997] - in order to reduce the complexity (see Figure 2.8). Lower confidence is given to the accuracy of pixel values at the beginning or end of the range (e.g. near 0 and 255) while higher confidence is attributed to mid range pixels' value accuracy. This algorithm is the one adopted in the data collection described in Chapter 5. Grossberg and Nayar's [2003] approach uses Principal Component

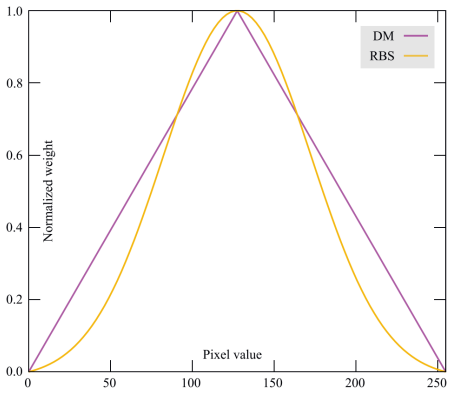


Figure 2.8: Robertson vs Debevec and Malik's camera response function recovery weights. Image courtesy of Akyüz and Gençtav [2013].

Analysis (PCA) to find the best parameters, rather than assuming a specific shape of the camera response curve. The authors performed a PCA analysis on 201 real-world CRF (as well as their inverses) in order to determine the best basis. To determine the CRF of a specific camera, the algorithm approximates the CRF with a polynomial function that adopts the said basis.

A comparison paper by Akyüz and Gençtav [Akyüz and Gençtav, 2013] illustrates strengths and weaknesses of these methods using the following metrics:

- consistency: the repeatability of the recovered CRF for a specific camera with different input images;
- accuracy: the closeness of a recovered CRF to the ground truth (established by capturing a colour chart of known reflectance);
- robustness: quality of the results when the quality of the imaging becomes poor.

Despite variations between the response curves, each algorithm produced relatively similar results for different scenes, with Mitsunaga and Nayar [1999] and Grossberg and Nayar [2003] being the most consistent. The comparison concludes that there is no statistically significant difference in terms of accuracy among the algorithms illustrated above when their variances are considered for different scenes. However, Grossberg and Nayar [2003] and Robertson [2003]

achieved the highest results, while Debevec and Malik [1997] and Mitsunaga and Nayar [1999] are shown to be the most robust under increasing levels of noise.

2.6.3 Native HDR Capture

Native image capture, where a specialised sensor or a camera optical setup are used, has the advantage of capturing all exposures simultaneously, hence avoiding temporal artefacts such as ghosting and motion blur, typical of the multiple exposures technique [Banterle and Unger, 2017].

Recently, many consumer camera producers (e.g. Canon, Nikon) started releasing models with sensors that achieve up to 14 f-stops RAW format, increasing the possibility of capturing static images with a wide dynamic range in a single shot.

The professional video sector offers cameras that can cover up to 16 f-stops (e.g., the Red MONSTRO and Arri Alexa Series). Cameras, where the sensor has a logarithmic response, are also available, but unfortunately, these perform poorly (i.e. high level of noise) in dark regions and also have a relatively low resolution [Banterle and Unger, 2017].

An alternative approach is to use multiple sensors with different sensitivities and adopt a prism or a mirror in order to split the image incoming from the lens. This is considered the best alternative [Kronander *et al.*, 2014], but suffers, nevertheless, from high sensitivity of the instrument in terms of alignment [Banterle *et al.*, 2017].

2.7 HDR Storage

LDR imaging typically stores images with 8-bit per pixel per colour channel offering a significantly reduced palette, in terms of both dynamic range and accuracy. Moreover, many applications commonly adopt *.jpeg* format [Wallace, 1991]. However, the use of this format hides many caveats: it is a lossy format where chrominance values downsampling and quantization are performed on the original RAW data and it is impossible to reverse the procedure.

HDR imaging captures and stores real-world radiance values. Each pixel

is encoded with a single precision floating point value per channel (R,G,B), thus enabling the representation of values in the range of $[0, 2^{32} - 1]$. The accuracy provided by uncompressed HDR images have a considerable cost in terms of storage space. If a single HDR pixel is stored uncompressed, it will occupy 96-bit per pixel (bpp) in memory, as opposed, for example, 24 bpp occupied by an uncompressed LDR pixel encoded in the *.png* RGB format. Therefore, several techniques have emerged in the literature to address this issue.

Different formats are available to store HDR images [Reinhard *et al.*, 2010]. The two most common are *.hdr* and *.exr* proposed respectively by Ward [1991] and by Industrial Light & Magic [Kainz *et al.*, 2009].

The Radiance HDR (*.hdr*) format, proposed by Ward [1991], uses an 8-bit mantissa for each primary followed by a single 8-bit exponent. This approach allows for the preservation of the most significant bits, which is the general goal of any floating point format. The use of a single exponent takes advantage of the fact that generally the three colour channels intensities are correlated. A single pixel encoded in this format occupies 32 bpp. A critique towards this encoding is the fact that it favours the largest primary value at the expense of accuracy in the other two primaries, although generally the largest value dominates the displayed pixel colour and therefore the other primaries become less noticeable, resulting in a perceptually smaller error [Mantiuk *et al.*, 2016].

A more compact representation of the HDR radiance values is represented by the OpenEXR format (*.exr*) by Industrial Light & Magic [Kainz *et al.*, 2009]. This encoding adopts half-floating point precision (48 bpp) offering the advantages of a floating point representation, with a reduced storage cost. The only limitation is that HDR images given in absolute luminance or radiance units often need to be scaled down by a constant factor before storing them, as some of the luminance values of outdoor scenes for example, can easily exceed the maximum value representable in 16-bit [Mantiuk *et al.*, 2016]. This format allows also for several lossless (e.g. Run Length Encoding) and lossy (e.g. PXR24) compression methods allowing for reduction in the total size of the image.

Moreover, the increased popularity of HRD images has promoted the

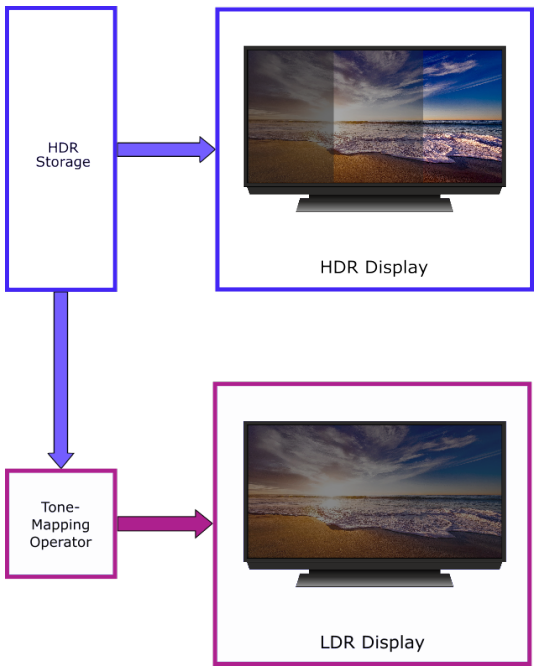


Figure 2.9: The HDR images can be visualised natively on an HDR display, retargeted to be visualised on an LDR display through the application of a tone-mapping operator.

creation of the JPEG XR standard, which is an image compression format that allows for the encoding of images with an arbitrary number of colour channels and different bit-depths [ISO/IEC 29199-2:2009]. Also, the development of a new decoding standard, the JPEG XT [ISO/IEC 18477-1:2015], that is structured in layers and allows for the handling of both LDR images and HDR images shows the growth of consumer HDR imaging applications. Although this is outside of the scope of this thesis, it is worth mentioning that there is a considerable body of literature dedicated to the problem of effectively storing and transmitting HDR images and videos. A thorough analysis is presented by Banterle *et al.* [2017].

2.8 HDR Display

The stored HDR images can be either visualised natively by using an HDR display, or they can be adapted to be visualised on an LDR display through the application of a tone-mapping operator (Figure 2.9). The first will be discussed

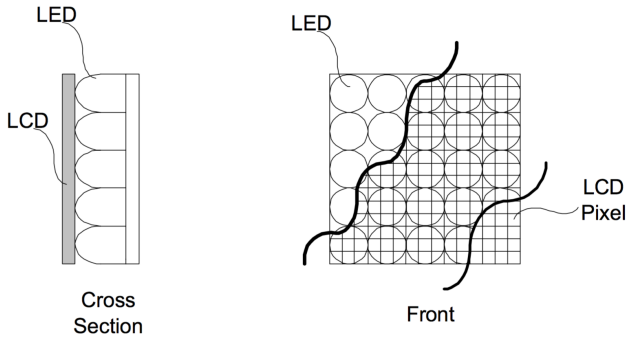


Figure 2.10: Dual-modulation HDR display, the input HDR image is redirected towards two components: a low-resolution version of the image created by the LED array is then projected through a colour LCD, which displays a similar, but higher resolution, version of the image. Courtesy of Seetzen *et al.* [2003]

in this section while the latter will be introduced in the next section.

The most desirable feature when visualising images - especially for scientific purposes - is to be able to reproduce reality as accurately as possible. Modern display technology exceeds visual acuity, for example, in terms of refresh rate and spatial resolution [Seetzen *et al.*, 2003], but reaching levels of brightness comparable to what we perceive in our daily life still remains a challenge. However, in recent years, HDR screens have become very popular on the consumer market, from being a niche product. In September 2016, SIM2 in partnership with the University of Warwick, TrueDR Ltd and Vicomtech-IK4, showcased the world's first 10,000 cd/m^2 display at the IBC Conference [Chalmers and Debattista, 2017].

The dominant technology for HDR screens is based on the use of a serial combination of a standard LDR panel with a spatial light modulator, typically an array of LEDs with coarser resolution [Trentacoste *et al.*, 2007], as illustrated in Figure 2.10.

In standard LCD screens, the backlight panel intensity can only be adjusted globally, for each frame, whereas, in the case of HDR screens, each LED is individually controllable allowing for a finer control. Several problems are the subject of current research: energy management; the back panel image generation - due to the many-to-many mapping between pixels and back panel LEDs; flickering issues for videos; necessity of algorithms capable to adapt the content depending on the display's luminance limits and saturation errors

[Dufaux *et al.*, 2016].

HDR display technology that adopts organic LEDs (OLEDs) as pixels has also been used by manufacturers. Since each OLED pixel is a light-emitting unit, backlights are not needed, and also high peak intensity can be reached with less power. However, the maximum brightness is around $540cd/m^2$ and black levels are about $< 0.0005cd/m^2$, reaching a contrast ratio of 1,080,000:120 f-stops, while LED screens are capable of contrast ratio 20,000:114.3 f-stops [Chalmers and Debattista, 2017]. Moreover, production cost and lifetime make OLED technology not competitive yet [Dufaux *et al.*, 2016].

2.9 Tone-Mapping Operators

HDR images can store physical radiance, but it is not always possible to display them natively on HDR screens. Tone-mapping is a technique that aims at reducing contrast and brightness in HDR images, so they can be displayed on LDR media (i.e. LDR screens or print).

The algorithm performing this transformation is called a tone-mapping operator (TMO). The main concept behind TMOs is the following: in the HDR file there is the availability of more information that can be handled by the display device, and therefore it is necessary to compress or discard some of it. Assuming an 8-bit precision, which is the most commonly used LDR format, the HDR pixel values are mapped into the interval $[0, 255]$. Specifically, the majority of TMOs operate on the luminance channel, while generally the colours are unprocessed [Banterle *et al.*, 2017]. The HDR luminance[3] channel L_{HDR} is mapped as follows:

$$L_d = TMO(L_{HDR}),\ where\ TMO : \mathbb{R}^+ \rightarrow [0, 255]\ . \tag{2.9}$$

The literature presents a vast quantity of tone-mapping operators, each one with different characteristics, often related to their target application. Depending on the operative aspect of the analysis, several classifications have been proposed [Čadík *et al.*, 2006; Reinhard *et al.*, 2010; Banterle *et al.*, 2017]. Eilertsen *et al.*

[3]The luminance channel can be calculated for example from ITU-R BT.709-6 [2015] RGB as: $L = 0.2126R + 0.7152G + 0.0722B$.

[2013] suggest the following classification based on the end-application:

- Visual system simulators (VSS): simulating the characteristics and limitations of the HVS;
- Scene reproduction (SRP) operators: pursuing the maximum fidelity to the actual scene appearance;
- Best subjective quality (BSQ) operators: aiming at satisfying the users preference.

Another categorisation is the one proposed by Banterle *et al.* [2017] based on the kind of operation performed:

- Global: the TMO performs the same operation on every pixel;
- Local: the TMO operation is performed locally to a pixel's neighbourhood.
- Frequency/gradient: the TMO operates on specific image frequencies - usually low frequencies (i.e. coarse details), while preserving high frequencies (i.e. fine details).
- Segmentation: the image is coarsely subdivided in distinct segments and different TMOs are applied to each of them.

Moreover, within these four categories, it is possible to further refine the TMOs according to their "design philosophy" [Banterle *et al.*, 2017]:

- Perceptual: they aim at emulating specific features of the HVS;
- Empirical: they aim at achieving aesthetically pleasing results.

The literature on the topic is substantial, with each TMO processing mainly focusing on some particular features [Čadík *et al.*, 2006]. Although every TMO is designed to preserve as many image attributes as possible, quite often the pursue of a specific characteristic results in detrimental effect on other desirable features.

For example, the algorithms proposed by Jack and Holly [1993] or Krawczyk *et al.* [2006] focus on preserving the perceived luminance, while the

work proposed by Larson *et al.* [1997] or Ashikhmin [2002] address contrast preservation - with some variations on its definition. Pattanaik *et al.* [1998] main attention was directed at accurate colour reproduction, while Fattal *et al.*'s work 2002 aimed at details preservation. In the following sections the work presented by Reinhard *et al.* [2002] and Debattista *et al.* [2015] will be described in detail, as these are used in Chapter 6 and Chapter 7 of this thesis.

2.9.1 Photographic Tone Reproduction Technique

Reinhard *et al.* [2002] proposed an Empirical Global and Local TMO inspired by Ansel Adams' photography technique [Adams and Baker, 1995c,b,a]. This operator is designed to emulate the Zone System widely used in analogue photography (i.e. film photography): a photographer collects the luminance values of the area in the scene that is perceived as having middle brightness and maps it to middle-grey values of the print (Zone V). After measuring the dynamic range of the scene, if this exceeds the representable range, some areas (i.e. underexposed or overexposed) are mapped to pure black or white during the development process. If specific areas of the image are to be preserved it is necessary to apply a "Dodging-and-burning" technique[4].

Reinhard *et al.* [2002] implemented this technique by first calculating the luminance for the pixel of coordinates (x,y) and then scaling it proportionally to the log-average luminance of the whole image using the following equation:

$$L(x,y) = a \, \frac{L_w(x,y)}{\frac{1}{N} \, exp(\sum log(\delta + L_w(x,y)))} \, , \tag{2.10}$$

where N is the number of pixels in the image, $L_w(x,y)$ is the "real-world" luminance value for the single pixel, δ is a constant set to avoid singularities and a is a scaling factor that will control for the key of the image.

The key of an image indicates whether the scene is subjectively judged as bright, normal or dark. For example, low-key photographs are characterized by striking contrasts, dark tones, and shadows. The following step consists of

[4]Technique adopted in the darkroom when printing from film negative: dodging decreases the exposure for areas desired to be lighter, while for areas that should be darker, burning increases their exposure.

Figure 2.11: Example image tone-mapped using Optimal Exposure [Debattista *et al.*, 2015]. Left. HDR Tone-mapped image, bit-depth: 8 bits. Right. Histogram of the log-luminance values, red indicates the range selected by the algorithm, blue indicates the information discarded due to the specified bit depth. Image courtesy of Fairchild [2007b].

tone-mapping the luminance value by calculating $L_d(x, y)$ as:

$$L_d(x, y) = \frac{L(x, y)(1 + \frac{L(x,y)}{L^2_{white}})}{1 + L(x, y)} , \tag{2.11}$$

where L_{white} is set to represent the smallest luminance value that will be mapped as white. If the maximum value of luminance in the scene is selected, then there will be no burn-out (i.e. overexposed areas).

In its local version, this operator calculates L_d for the largest local neighbourhood without sharp edges [Banterle *et al.*, 2017]. Specifically, the area is automatically detected by using a measure of local contrast [Peli, 1990], more precisely, using circularly symmetric Gaussian profiles at different scales. The neighbourhood size is adjusted by controlling the sharpening parameter ϕ: the higher ϕ value the larger the scale of the pixel's neighbourhood defined. Together with the parameter a, ϕ allows the user to control for the selection of the key and sharpening respectively [Salih *et al.*, 2012; Mantiuk *et al.*, 2016].

2.9.2 Optimal Exposure Algorithm

When dealing with images exceeding the possible dynamic range displayable by an output device, an alternative option to tone-mapping is an optimisation algorithm like the Optimal Exposure Algorithm proposed by Debattista *et al.*

[2015].

This algorithm emulates the behaviour on the cameras' onboard software when trying to identify the best exposure given a specific scene luminance. The algorithm calculates the histogram of the input image log-luminance and selects the exposure that preserves the maximum possible amount of information from the original HDR image given a specific bit depth. The number of bins of the histograms is calculated with the Freedman-Diaconis rule.

2.10 Expansion Operators for Low Dynamic Range Images

This class of operators, first introduced by Banterle *et al.* [2006], allows the expansion of LDR content into HDR content. This kind of transformation is extremely useful when facing the problem of showing legacy content on HDR displays or using LDR images for applications such as image-based lighting [Eilertsen, 2018]. These operators are frequently also called Inverse Tone-Mapping Operators (iTMOs). Although strictly speaking iTMOs are a subset of EOs, the two names will be used interchangeably throughout this thesis Banterle *et al.* [2017]. The generic Expansion Operator EO over an LDR image I_{LDR} can be expressed as:

$$I_{HDR}(x, y) = EO(I_{LDR}(x, y)) : [0, 255] \rightarrow \mathbb{R}^{+} , \tag{2.12}$$

where the original LDR image I_{LDR} defined over the domain $[0, 255]$ is expanded into an HDR image domain I_{HDR}, which in the case of single precision floating point is $[0, 3.4 \times 10^{38}]$. As explained in section 2.9, when storing radiance values in an LDR image there is a loss of information that cannot be easily reconstructed without any knowledge of camera characteristics and settings used for capturing and without any information about the captured scene [Mantiuk *et al.*, 2016]. In an image like the one in Figure 2.12, for example, where the sky has overexposed/saturated pixels it will be impossible to determine if the lost information belonged to a cloud or to a blue sky.

Ideally, the expansion operator should be able to reconstruct the CRF and therefore re-establish the original photometric quantities. The first step,

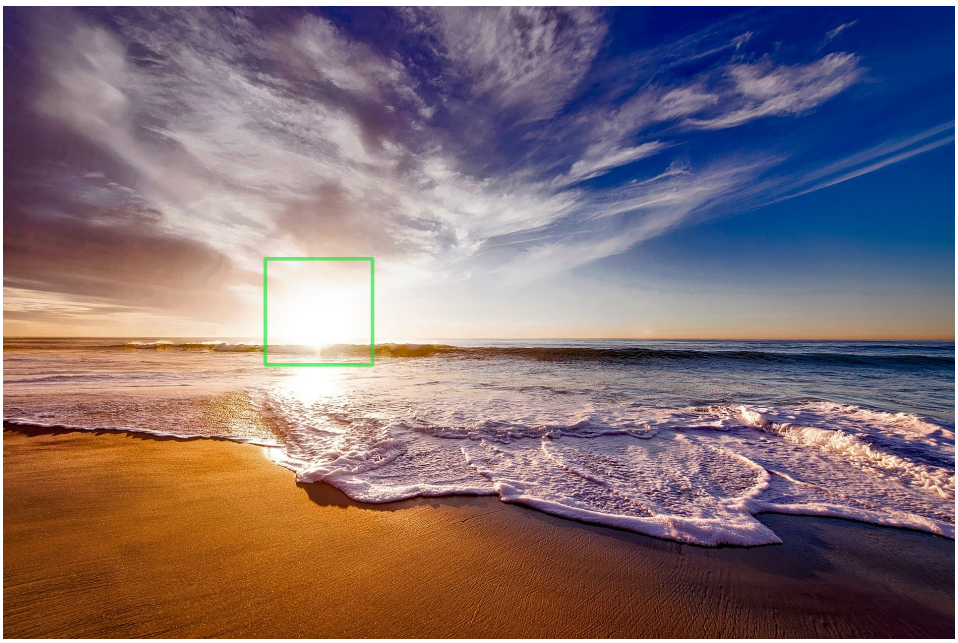


Figure 2.12: Example of LDR image where the scene dynamic range exceeds the capability of the camera sensor. The LDR image captured has saturated pixels (green square), whose original information is lost and cannot be recovered.

which is generally adopted by the majority of the EOs, is to invert the non-linearity introduced by the gamma-correction curve applied when the LDR image was encoded. The second step is the expansion of the dynamic range of the well-exposed pixels values to the dynamic range supported by the HDR displays via a specific function operating locally or globally depending on the algorithm. The computation is performed on the linearised single-channel luminance. The third step is characterised by the reconstruction of the over- and under-saturated pixels.

However, this not an easy problem to solve because the lost information cannot be recovered. Mantiuk *et al.* [2016] suggest that the best results are obtained by in-painting and texture synthesis techniques, which, require user interaction, such as [Wang *et al.*, 2007]. In the next section the first EO operator proposed by Banterle *et al.* [2006] will be described. Section 2.10.3 will offer a description of ExpandNet by Marnerides *et al.* [2018]. This is an example of a recent class of EOs that adopts machine learning techniques to reconstruct the original DR.

2.10.1 Inverse Tone-Mapping Operator

Banterle *et al.*'s [2006] inverse Tone-mapping operator (iTMO) follows the following four steps:

1. the LDR input image is expanded by inverting the Photographic Tone Reproduction by Reinhard *et al.* [2002];
2. the Median Cut Algorithm is applied [Debevec, 2008] to identify areas of high luminance;
3. the *Expand Map* containing high luminance areas, found by the Median Cut Algorithm is created;
4. the original LDR image and the HDR image, generated at step 1, are linearly interpolated by using the Expand Map values as weights.

The solution of step 1 is achieved given some assumptions:

- the geometric average of the luminance in the resulting HDR image is equal to the LDR one ($\overline{L}_{HDR} = \overline{L}_{LDR}$)
- the luminance for each pixel in the Reinhard formula ($L_d(x, y)$) is assumed to be equal to the luminance value calculated in the LDR image as a linear combination of the three RGB channels as $L = 0.213R + 0.715G + 0.072B$.
- they introduce a variable selected by the user L'_{max} which links the two parameters in Reinhard's formula α and L_{white}.

2.10.2 Machine Learning Techniques for Dynamic Range Expansion

One of the most recent and successful trends to recover missing information is to use machine learning (ML) techniques in order to avoid the need for human expertise or heuristics. Amongst the most recent works Eilertsen *et al.* [2017] and Endo *et al.* [2017] values, for example, use Convolutional Neural Network (CNN) to reconstruct pixels' value. Discussing CNN is outside the scope of this thesis, Goodfellow *et al.* [2016] offer an extensive review of the topic.

Eilertsen *et al.* [2017] were the first to propose the use of deep learning algorithms for a dynamic range expansion application. The algorithm applies an inverse CRF (using a logarithm approximation) in areas where the pixels are well defined. Whereas, in over-exposed only areas, a deep convolutional neural network (CNN) specifically designed for HDR value predictions is used. The training stage is done on a large dataset of HDR images, also augmented by simulating sensor saturation for a range of cameras. This algorithm does not deal with the recovery of dark regions of an image. In the following section a detailed description of the CNN architecture developed by Marnerides *et al.* [2018] is provided as this technique is used in the experiment presented in Chapter 7 of this thesis.

2.10.3 ExpandNet Architecture

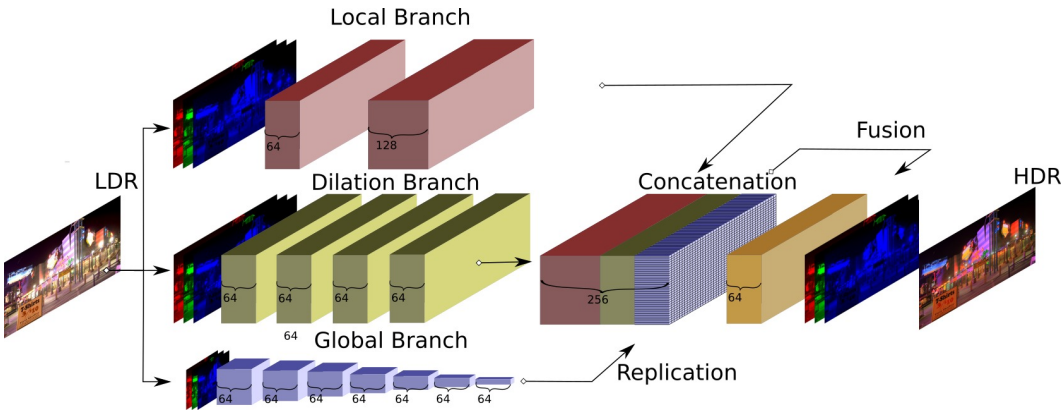


Figure 2.13: ExpandNet architecture. Courtesy of Marnerides *et al.* [2018].

The first end-to-end architecture aiming at dynamic range expansion that uses a CNN was proposed by Marnerides *et al.* [2018]. The basic characteristic of ExpandNet is the presence of three branches, each one responsible for the extraction and preservation of specific features of the image at different levels of detail (illustrated in Figure 2.13). The local branch deals with local detail, the dilation branch accounts for medium-level detail, and a global branch is responsible for preserving higher-level features. The three branches' outputs are eventually concatenated and processed by a final convolutional layer that outputs the final HDR image. The CNN does not make any prediction about absolute luminance, therefore the images are normalised in the range $[0, 1]$.

The global branch is the only one that downsamples the image reducing the dimensionality of the input. It expects an LDR input of 256×256 pixels maximum, and progressively downsamples it to a feature of size 1×1 over seven layers.

The local branch is responsible for the preservation of fine details. It extracts features using two layers with a kernel of 3×3, a stride of 1 and padding of zeros of size one[5]. It produces 64 and then 128 feature maps respectively.

The dilation branch has a wider receptive field: 17×17 pixels and adopts a specific type of convolution called dilated convolution [Yu and Koltun, 2015]. This allows for the aggregation of multi-scale contextual information without losing resolution. Four dilation layers are used, each with 64 features. This branch allows for capturing features in the mid-range frequencies.

The results obtained by each of the three branches are eventually concatenated and fused together. The output of the dilation and the local branch do not need upsampling (they have the same height and width as the input) and they are simply concatenated, while the global branch output (i.e. a 64 features vector) is replicated along the width and height dimensions. The resulting output is a 256 features map. The concatenation is followed by two convolutional layers: the first, with a size 1×1 kernel, fuses the global feature vector with each individual pixel of the local and dilated features; and the second, with a 3×3 kernel and stride 1, produces the final HDR output image.

2.11 HDR Evaluation

The main two approaches for evaluating the quality of the images created with the previously described algorithms are [Banterle *et al.*, 2017]:

- using metrics: to compare image similarity, generally based on image statistics or simulations of the HVS;
- through subjective experiments: visual comparisons of the stimuli performed by human participants.

[5]Zero padding is a common technique employed in image processing consisting in surrounding the original image with extra zero-rows and zero-columns. This allows to slide the kernel centre over pixels located at the edge of the image.

If humans are the end-users of the HDR imaging, then subjective experiments are certainly more indicative of its quality and, consequently, the effectiveness of the adopted algorithms. However, running subjective experiments is costly and time-consuming.

Some of the metrics available in the literature are derived from signal processing theory, for example, peak signal-to-noise ratio (PSNR), while others are based on the simulation of the HVS behaviour. Mantiuk *et al.* [2011] proposed the HDR visible difference predictor (HDR-VDP) metric. This metric, through a process that emulates the HVS response, produces a probability map showing how noticeable the differences between two HDR images is.

With regard to TMOs and the fact that, depending on the input image, their performance is different, it is often necessary to evaluate how the tone-mapped image compares to the original HDR. Ledda *et al.* [2005], for example, conducted an experiment where participants were asked to compare tone-mapped scenes to the original HDR reference image. The experiment described in Chapter 7 of this thesis requires the comparison of several TMOs, therefore in the following sections an in-depth review of TMO evaluation methods, with a focus on facial stimuli, will be provided.

2.12 TMO Evaluation: Subjective Experiments

The literature presents many perceptual experiments in which researchers have tried to establish the best TMO, similarly to the experiment performed by Ledda *et al.* [2005]. The majority of the literature engaging with comparisons amongst different TMOs has established that the choice of tone-mapper is application-oriented and it depends on specific characteristics of the images [Čadík *et al.*, 2008; Parraga *et al.*, 2018].

Čadík *et al.* [2006] investigated the natural reproduction of real-world scenes (i.e. indoor, outdoor, night urban) while also proposing a methodology for evaluating TMOs using common image attributes (i.e. brightness, contrast, detail and colour reproduction, and presence of artefacts). They compared 14 TMOs. Those that had the best overall performance, with no significant

difference among them, were all global TMOs with the exception of by Reinhard *et al.* [2002] (described in Section 2.9).

Salih *et al.* [2012] conducted a similar subjective study comparing six TMOs. In all cases Reinhard *et al.* [2002] (local and global) obtained the highest score. Several other studies on the topic have been conducted, but, generally, the evaluation involves generic images. The only study that has evaluated tone-mapping preferences on facial stimuli is Korshunov *et al.* [2015].

Korshunov *et al.* [2015] compared five different TMOs for the purpose of a face-matching task. They tested 860 participants, 756 of whom were considered trustworthy, through a crowdsourcing platform[6]. The experiment was conducted as a one-to-many type of paradigm, where the participant was asked to match a face image, captured natively as LDR, to a set of HDR tone-mapped images. The chosen operators were Drago *et al.* [2003], Reinhard *et al.* [2002], Mai *et al.* [2011], and Mantiuk *et al.* [2006a], in addition to a simple a gamma clipping operator. Their study proved the advantage brought by tone-mapped images over LDR images, but there was no significant difference among the TMOs tested.

2.13 TMO Evaluation Methods: Metrics

The adoption of metrics allows speeding up the process of evaluating a specific TMO without having to run a subjective experiment and also allow automating the process of tuning specific TMO' parameters [Banterle *et al.*, 2017]. Several metrics are available in the literature. This section describes two metrics in more detail, as they are used in Section 7.4.1.

2.13.1 Tone Mapped Image Quality Index

The Tone Mapped Image Quality Index (TMQI), presented by Yeganeh and Wang [2012], is a method that allows evaluating the quality of an HDR image compared with its tone-mapped version. This algorithm takes as input an HDR (I_{HDR}) and its tone-mapped (I_{LDR}) version, and returns an index between 0

[6]With crowdsourcing there is the risk of participants not answering the test genuinely, so quite often "honey traps" are inserted to ensure the participants are correctly engaging with the experiment.

and 1. The index is specified as:

$$TMQI(I_{HDR}, I_{LDR}) = a\ S(I_{HDR}, I_{LDR})^{\alpha} + (1-a)N(I_{LDR})^{\beta}, \tag{2.13}$$

where $0 \leq a \leq 1$. The first component, S is a measure that accounts for the structural fidelity of the tone-mapped image, while the second component N, is a measure that accounts for the naturalness of the image. The parameter a controls for the importance of structural fidelity and naturalness, while α and β are parameters that control sensitivity (the suggested parameters are $a = 0.8012$, $\alpha = 0.3046$ and $\beta = 0.7088$).

In 2015 Ma *et al.* proposed TMQI-2 which offers an improvement on the statistical naturalness components and structural fidelity of TMQI.

2.13.2 Feature Similarity Index for Tone-Mapped Images

Nafchi *et al.* [2014] proposed a new technique called Feature Similarity Index for Tone-Mapped Images (FSTIM). This method is based on a phase-derived feature map. The algorithm acts in the Fourier domain and uses the image phase angle map.

The FSTIM on a single channel C (with $C = \{R, G, B\}$) for two images I_{HDR} and I_{LDR} is calculated as:

$$FSTIM^{C} = \alpha F^{C}(I_{HDR}, I_{LDR}) + (1-\alpha)F^{C}(log(I_{HDR}, I_{LDR}), \tag{2.14}$$

where $0 \leq a \leq 1$. F^{C} is a similarity index based on locally weighted mean phase angle for the two input images. This provides a good representation of the image features, including edges and objects' shapes, while being independent from noise. Moreover, it assesses colour changes, which is a frequent shortcoming of many TMOs. The parameter α controls for the influence of I_{HDR} and $log(I_{HDR})$. The authors also suggest a combined index (FSTIM_TMQI) given by averaging the sum of FSTIM and TMQI, which has the effect of moderating similarity estimation errors of the two metrics.

2.14 Summary

Answering the research question of this thesis involves the understanding of two different elements: the digital imaging technology and the process of face perception and recognition. This chapter has briefly introduced the mechanisms governing the process of vision and has reviewed the literature related to digital imaging and the HDR imaging pipeline in particular. In the next chapter the psychological and perceptual literature will be reviewed and the problems related to the unfamiliar face matching will be highlighted.

Chapter 3

Face Perception and Recognition

Faces are central to human interaction as they convey a great deal of information. Simply looking at someone's face triggers all sorts of mechanisms such as defining their identity (deciding whether we know the person or not), determining their sex or approximate age or emotional state. Also, in case of positive recognition, we will recall the person's name and various knowledge related to them, including one's emotional response to that individual [Haxby and Gobbini, 2012].

The literature on this topic is quite vast, as conquering a deeper understanding of the conscious processes behind face recognition has been a challenge that many scientists of different disciplines have taken up. This chapter will have an outlook on the topic mostly focusing on the psychological perspective (both cognitive and perceptual). Furthermore, a review of factors affecting face recognition will be provided with particular focus on the effect of imaging characteristics on recognition. Finally, a brief overview of the automatic face recognition is provided.

3.1 Models of Human Face Processing

Theoretical models are used to organise and synthesise many scattered empirical data. One of the most historically popular models related to human face processing has been proposed by Bruce and Young [1986].

This model (called *Bruce&Young*) describes the modalities of face perception as a sequence of stages. Each one of these stages enables the generation and access to a piece of information - called *code*. According to their model, there are seven different kinds of *codes* (pictorial, structural, identity-specific semantic, visually derived semantic, name, expression and facial speech). Each *code* is the product of a separate processing module. This model has been considerably revised and updated since its first publication in 1986, with, for example, the elaboration of the Interactive Activation and Competition (IAC) model by Burton *et al.* [1990], but it has always occupied a central position in the cognitive literature in the last 30 years, either to confirm or refute it. However, neither *Bruce&Young* nor the IAC model address precisely which information is extracted during face processing [Hole and Bourne, 2010]. To date, a definitive answer has not been found.

The Distributed Neural System for Face Perception is one of the most recent and accredited models, and it has been elaborated starting from a great number of neurophysiological observations conducted by means of fMRI (functional Magnetic Resonance Imaging) and PET (Positron Emission Tomography). Proposed by Haxby *et al.* [2000], similarly to *Bruce&Young*, it suggests the dissociation between the two mechanisms for the visual analysis of facial expression and identity [Calder and Young, 2005]. Differently to *Bruce&Young*, this model for face processing combines both a cognitive conceptualisation and a hypothesis for a distributed neural system organisation.

The coordinated recruitment of several areas produces what we call face perception. The model consists of two main hierarchically organised blocks:

- the Core System: responsible for the visual analysis of a face;
- the Extended System: responsible for processing the information collected to infer a meaning.

The Core System consists of three main regions identified by the adoption of neuroimaging techniques, as areas that respond to viewing faces (i.e. the activation is the highest when compared to any other object): the occipital face area, the lateral fusiform area and the posterior superior temporal area.

These three areas correspond to two classes of face perception operations:

- identity recognition: identifies invariant features across facial movements and any variation in viewing conditions enabling the viewer to distinguish between individuals;

- facial expression perception: groups together variability in facial movements, regardless of identity, enabling to distinguish, for example, a smile from an angry pout.

The Extended System comprises of areas outside the visual extrastriate cortex that are dedicated to familiar face recognition. These areas are activated by seeing someone's face and are responsible for the retrieval of details about this individual, such as biographical information or episodic memories.

The Extended System includes areas related to emotion expression and eye gaze: seeing someone else's emotional expression evokes activity in regions associated with emotion, and with the motor areas for producing the same movements in ourselves. Although the majority of experimental data from functional imaging and neuropsychology studies are commonly interpreted within this framework, there are still reservations about this model [Calder and Young, 2005].

3.2 Face Recognition: Familiar vs Unfamiliar

Although the common assumption is that adults are "face-experts" [Carey, 1992], this is generally true only in the case of familiar faces [Young and Burton, 2017]. Familiar face recognition is a quite intuitive concept and in the literature is defined as the recognition of famous or personally familiar faces, as well as, the recognition of faces that were previously unfamiliar and went through an extensive learning phase [Johnston and Edmonds, 2009]. The extended exposure to a face allows the observer to build representation robust enough to become less sensitive to changes (e.g. viewpoint, lighting conditions). Humans show good expertise with faces that fall into categories to which the individual is constantly exposed - for example, people's own racial group (i.e. own-race effects) [Ritchie *et al.*, 2015] and they become progressively less accurate as the degree of other-race contact in their daily experience decreases.

Unfamiliar faces are faces that the person has not encountered before or to which has had very little exposure to. When encountering an unfamiliar face, recognition might be challenging and often a source of misidentification errors [Young and Burton, 2017]. The difficulty is partially related to the capability of coping with a person's changes in appearance [Ritchie *et al.*, 2018]. Besides the obvious interference due to changes in external features (e.g. hair, hats, glasses), in many situations, the stimuli characteristics have a significant impact (i.e. photos, video, live observation).

There is neurophysiological research [Eimer, 2000; Herzmann *et al.*, 2004] using event-related potentials (ERP) in support of the fact that familiar faces trigger different processes from unfamiliar ones. Results have shown that the very early stages of face perception, 130-200 ms after the stimulus, do not seem to be influenced by familiarity (N170), while a later event-related potential elicited between 200 ms and peaking at 250-300 ms (N250) seems to be larger for familiar faces compared to unfamiliar faces [Schweinberger *et al.*, 2004]. Moreover, several studies on skin conductance responses (SCRs) were also able to replicate these findings reinforcing the hypothesis that humans have different representations for familiar and unfamiliar faces [Ellis *et al.*, 1999].

Some cognitive psychology theories point towards the idea that familiar faces are processed primarily through the use of structural codes [Bruce and Young, 1986; Johnston and Edmonds, 2009] (see Section 3.1): for each face encountered individuals create an abstract memory representation that is progressively refined through subsequent encounters. Each time knowledge of specific facial features (i.e. constancies) will be reinforced across different views. Unfamiliar face recognition is, on the other hand, highly image-specific (relying mainly on pictorial codes). When faces are learnt from photographs, people do not readily form a viewpoint-invariant model, but instead recognition is mediated by image-based codes which may be either viewpoint-dependent (e.g. changes in distance and angle from the camera) or pictorially-dependent (e.g. light changes). Longmore *et al.* [2008] observed that the variation in lighting resulted in a significant performance drop, compared to viewpoint variations, suggesting that image-specific pictorial codes have a predominant impact.

[0]This technique consists of the extraction of the neural responses to specific events the EEG, see Luck [2014] for an introduction on this topic.

Unfamiliar and familiar face recognition task have been studied with a variety of experiments trying to identify the most perceptually influential information for the purpose of recognition. These experiments aimed at isolating specific factors such as lighting, facial expression, viewpoint or negation (i.e. images presented with inverted colours). Johnston and Edmonds [2009] in their review on the topic offer a summary of the various effects and whether these factors help or hinder recognition, here illustrated in Table 3.1.

Unfamiliar	**Familiar**
View	View
Expression	Expression
Context	Context
Lighting	Lighting
Negation	Negation
Inversion	Inversion
Movement	Movement
Distinctiveness	Distinctiveness
Caricaturing	Caricaturing
Salience of Internal Features	Salience of Internal Features

Table 3.1: List of factors influencing face recognition for familiar and unfamiliar faces. The green, red and white blocks identify respectively a factor that helps, hinders or has no impact on the recognition process. Summary adapted from Johnston and Edmonds [2009].

Within this corpus of literature, a subset of experiments has emerged: face matching, as opposed to face recognition (i.e. from memory) tasks [Burton, 2013]. The following section will describe face matching in detail.

[0]Although negation is not something that would happen when performing a recognition, studies have used its effects when exploring underlining face encoding.

3.3 Face Matching

Matching unfamiliar faces is a difficult problem, even when high-quality images are used [Bruce *et al.*, 1999]. Although strictly related to face recognition, unfamiliar face matching has emerged in the last two decades as a problem worthy of investigation in his own right from both its theoretical and practical implications. On the one hand, it allows to expand our understanding of this cognitive processes and, on the other hand, it is applicable in a number of security-related situations in which photo-to-photo matching is required, like, for example, police investigators matching CCTV footage to an image of a suspect.

3.3.1 Facial Stimuli Datasets

A typical aspect of studying face matching depends on the use of appropriate stimuli to draw conclusions that could be generalised. Several facial stimuli datasets are available in the literature for unfamiliar face matching experiments; amongst the most popular is the Glasgow Face Matching Test [Burton *et al.*, 2010] and the recently released Kent Face Matching Test [Fysh and Bindemann, 2018]. Unfortunately, these stimuli set frequently present technical limitations especially in terms of image resolution, lack of information regarding camera calibration, and colour accuracy. The literature also offers other datasets [Langner *et al.*, 2010; Strohminger *et al.*, 2015; Ma *et al.*, 2015a], but they are aimed at emotion recognition and therefore do not typically include several representations of the same face in the same pose - an essential feature for side by side comparison which avoids mere picture comparison [Young and Burton, 2017].

Other attempts involve computer graphics generated stimuli, but this might raise the objection of drifting even further away from ecologically valid stimuli [Logie *et al.*, 1987]. There is also a large availability of face datasets oriented at computer vision applications, but these are rarely used in perceptual psychology experiments, often due to intrinsic differences in the construction of the stimuli although there have been a few attempts [Phillips *et al.*, 2012; Burton, 2013; White *et al.*, 2015].

When it comes to experiments on face matching the Glasgow Face

Matching Test (GFMT) [Burton *et al.*, 2010] is among the most popular datasets, having been featured in over 30 face-matching studies. It consists of medium quality (300-horizontal pixels per face, see Section 3.5) frontal facing faces with a neutral expression. It comprises same-day photographs taken with two different cameras. Although camera settings and characteristics are mentioned as relevant factors by the authors, no other technical detail is provided. Its main limitations are as follows:

- low image resolution - for modern standards;
- image compression - the images are stored in *.jpeg* format;
- absence of information relative to lighting conditions - it just mentions "good lighting";
- absence of any information related to the characteristics of the two cameras adopted: besides mentioning the brand and the resolution, no mention is made about camera settings or white balance.

In their recently released Kent Face Matching Test [Fysh and Bindemann, 2018] the authors focus mainly on the time interval occurring between data capture, as experiments, like the ones by Megreya *et al.* [2013] have shown impairment due to changes in time of the face. The dataset is presented in two versions: one consisting of 40 Caucasian identity pairs (20 males, 20 females) and one comprising 220 face pairs (166 females, 54 males). In both versions, the two pictures for each participant are: a student-ID card portrait and a portrait taken with an off-the shelf bridge camera (Fujifilm FinePix S2980).

This dataset, however, shows limitations similar to the GFMT:

- relatively low resolution for modern standards (283×332 pixels for one set of images and 142×192 pixels for the student-ID one);
- ID photographs are not constrained by expression, pose, or image-capture device

[0]White balancing is part of colour grading standard procedures in professional digital imaging. These allow for the accurate reproduction of the colours in the captured scene, rather than relying on the camera inbuilt proprietary software whose aim is to create visually pleasing pictures, not necessarily accurate ones.

- no information related to the characteristics of the two cameras adopted: besides mentioning the brand and the resolution, no mention is made about settings or colour grading.

3.4 Facial Identification: Context and Applications

When forensic facial examiners are called to make a judgement, although superimposition and photo-anthropometry have been used in the past, their low reliability has been now clearly demonstrated [Valentine and Davis, 2015]. Morphological analysis is the only methodology that is considered reliable as it is less susceptible to the camera's optical properties and viewpoint. Nevertheless, operators need to be aware of its limitations (see Valentine and Davis [2015] for a detailed analysis). Usually, the only manipulation possible on the picture is related to luminance adjustments [Steyn *et al.*, 2018], showing clearly how a more accurate lighting reproduction could bring an advantage to the forensic practice.

Forensic facial reviewers are a group of experts enrolled in the law enforcement, whose task is to perform fast and less rigorous identifications while assisting in criminal cases [Phillips *et al.*, 2018]. This group of observers is even more susceptible to misidentification errors because of the time constrains. Therefore, technological support can be advantageous especially for tackling data limits (e.g. illumination, viewpoint, image degradation, within-target variation)[Fysh and Bindemann, 2017].

When non-experts are called to judge the representation of a face (e.g. a bank cashier checking a document-ID or a juror watching surveillance footage) their accuracy is on average lower than experts [Norell *et al.*, 2015] and especially affected by poor image quality (i.e. more false positive and more false negative). False positives, in case of a crime witness, could determine the prosecution of an innocent person [Norell *et al.*, 2015], therefore, the impact brought by advanced imaging could be valuable.

Figure 3.1: Examples of facial images with different quantisation. From left to right: full resolution (over 500 pixels/face), 26 pixels/face, 18 pixels/face.

3.5 Factors Affecting Face Recognition from Pictures

Face matching experiments allow evaluating the image properties of the proposed stimulus and their impact on the task. While familiar viewers show high performance, unfamiliar viewers are greatly affected by factors not just related to the basic characteristics of the stimuli (e.g. pigmentation, distortion [Hancock *et al.*, 2000] or configuration [Burton, 2013]), but are also affected by the properties of the technology adopted to conduct the experiment. As Young and Burton [2017] have pointed out, sometimes the extent to which image variability, for example, due to lighting, camera and lens characteristics, influences performance is overlooked.

In the context of face image processing, pixelation consists of the subdivision of the face portrait into a grid. For each square of the grid, the average luminance (or colour values) of that region in the original image is calculated. The resulting image resembles a mosaic. Depending on the coarseness of the sampling grid, disruption of the original image can be more or less noticeable. The number of horizontal pixels at nose level per face (pixel/face or pix/f) is the unit of measure most commonly adopted in the literature to describe such images (see Fig 3.1).

One of the first researcher to systematically investigate the impact of pixelation on face perception was Bachmann [1991]. In his experiment, eight

levels of pixelation (i.e. 15, 18, 21, 24, 27, 32, 44, 74 pixel/face) and a set of 48 pixelated images depicting male faces were used in a forced-choice experiment where the pixelated image had to be identified by referring to the full resolution one. Six exposure durations were used (i.e. 1, 4, 8, 20, 40, 100 ms). The results showed a sharp decrease in the identification rate at 15 pixel/face, establishing the minimum amount at 18 pixel/face.

Bindemann *et al.* [2013] have shown that resolution has an effect on face matching bringing performance almost to chance when just a few pixels are available for the representation. They identified that a 20-pixels horizontal resolution (i.e. 20 pixel/face) brings down the accuracy of identification from 85% at full- resolution (i.e. 380-pixel/face) to just 66%, reaching values as low as 48% if the resolutions scales down to 14-pixel/face. Although it can be shown that applying a Gaussian blur on the image can help performance, this still raises questions about the reliability of certain CCTV footage.

Ritchie and Burton [2017] brought attention to the impact of lighting and camera parameters variation on the recognition process. Noyes and Jenkins [2017], proposed an experiment where they changed the camera-to-subject distance (0.32m versus 2.70m) and recorded a high impairment of perceptual matching of unfamiliar faces, even if images' size was kept uniform, while familiar face matching was unaffected. In their research, it is underlined how no governmental rules are in place to tackle such a fundamental issue especially considering the popularity of official forms of identification based on photos.

Megreya and Burton [2008] presented a study in which participants were asked to match live actors to LDR passport size pictures. In their analysis, the poor performance is solely attributed to human perceptual limitations especially in connection with familiarity. However, no consideration was made relatively to the quality of the imaging offered to the participants for the comparison (a $5x7$cm 8-bit greyscale *.jpeg* compressed image - 72 dpi).

Russell *et al.* [2006] conducted experiments on the effect of pigmentation. Showing that it is extraordinarily difficult to recognize a face in an image with negated contrast, as in a photographic negative. Negation disrupts the observers' ability to use pigmentation cues and the authors suggest that, by extension, this shows that pigmentation contributes to the neural face representation [Hill and Bruce, 1996]). Other studies have, however, shown little effect of

using colour rather than greyscale images [Hancock *et al.*, 2000]. Stephen and Perret's study [2016] seems to indicate that colours influence social judgements as they are a cue to attractiveness and health (e.g. skin's increased redness enhances the apparent femininity and attractiveness).

An interesting result from Bhatia *et al.* [1995], is the introduction of greyscale levels besides pixelation. The experiment showed that the higher the number of grey-levels, the smaller the number of pixels were needed to perform the forced-choice task. The limitation of this study, however, lies in their focus solely on the discriminability of faces: participants were asked to discriminate between face vs non-face.

Other studies have attempted a different kind of manipulation, consisting of performing spatial filtering on the images. This operation is performed in three steps: 1) the image is transformed using the Fourier transform, 2) specific frequencies are selected 3) the image is transformed back using the Inverse Fourier Transform. Several studies have shown that there is a specific range of frequencies (i.e. 6.8 - 17.6 cycles/face) that are optimal for recognition performance in terms of accuracy and response times [Jeantet *et al.*, 2018]. The "optimal range", is nevertheless affected by other factors such as, for example, the specific task or the stimulus eccentricity or the exposure duration or the stimulus distance [Jeantet *et al.*, 2018].

Several studies have focused on the effect that changes in lighting has on face perception [Hill and Bruce, 1996; Longmore *et al.*, 2008; Hole and Bourne, 2010; Norell *et al.*, 2015; Fysh and Bindemann, 2017], on the one hand, lighting is beneficial, as shadows contribute to the understanding of the three-dimensional nature of objects [Ramachandran, 1988], on the other hand, overcoming these changes is essential to be able to correctly perform a face recognition. Neurological studies by Hietanen *et al.* [1992] have suggested that the object recognition processes are view-specific. Braje *et al.* [1998] found that changes in illumination had a detrimental impact on face matching. Later subjective experiments by Braje [2003] had participants performing an unfamiliar face matching task while illumination and position of the stimuli (i.e. eliciting central vs peripheral vision) were varied. This experiment confirmed the impact of illumination on central viewed faces suggesting that

[0]i.e. if the stimulus is processed in foveal projection or in the peripheral region.

face representation retains illumination information, but, interestingly, when the position was varied (i.e. stimuli viewed using peripheral vision rather than central vision illumination), illumination did not affect the recognition. Favelle *et al.* [2017] performed a study where the interaction between lighting position and face rotation was examined. Their results indicate that face perception processes involve the creation of a surface shape representation which is especially susceptible to top lighting illumination but not susceptible to left or right illumination. One limitation of this work is the use of digitised faces, and the limited range of lighting angles variation (i.e. yaw axis: 0, 45, -45 degrees).

In summary, research is still required in order to have a precise understanding of how humans generalize faces from known to unknown lighting conditions [Tarr *et al.*, 2008]. The adoption of HDR imaging with a higher fidelity in the reproduction of lighting might contribute to this discourse.

3.6 Neuropsychological Face Recognition Impairment: Prosopagnosia

The term coined by Boadamer in 1947 (reprint by Ellis and Florence [1990]), comes from the Greek words *prosopon*, meaning face, and *agnosia*, meaning not-knowledge. One of the cases reported by Bodamer was, for example, a 24-year old man who, as a result of a brain injury, had been suffering from a number of neuropsychological impairments. Among these, the patient had become unable to recognise familiar faces or process their facial expressions. The patient would rely only on non-facial features, such as hairstyle or glasses, to being able to identify other people.

Studies have estimated that prosopagnosia affects between 1.9 % and 2.5% of the population [Kennerknecht *et al.*, 2006, 2008]. It is usually divided in two categories [Duchaine *et al.*, 2007]: *acquired prosopagnosia*, where the impairment appears after the person suffers brain damage, and *developmental prosopagnosics* (DPs) where there is a lack in the development of normal face recognition abilities. Some researchers suggest this phenomenon to be specifically associated with face processing, while others believe it to be more subtly

connected to general processing of the visual stimuli [Hole and Bourne, 2010]. Regardless of its origin, identification of people affected is not straightforward [Gainotti, 2010] as no formal diagnostic criteria exist.

Generally, computer-based tests of face recognition ability are used, among these the Benton Facial Recognition Test [Kreutzer *et al.*, 2011], or the Cambridge Face Memory Test [Duchaine and Nakayama, 2006] and the Cambridge Face Perception Test [Duchaine *et al.*, 2007]. Additionally, self-report evidence is used to supplement computer-based tests as, for example, prosopagnosics patients may score considerably well in these tests as they have developed compensatory strategies.

Shah *et al.* [2015] have proposed the 20-item prosopagnosia index (PI20), a self-report questionnaire that allows the identification of developmental prosopagnosia. It consist of a standardised set of questions to reinforce the assessment of DP. The test consists of a 20-item set of questions, where the respondent is required to indicate a value between 1 and 5 The questionnaire is reported in Appendix A.1. The total scores of the PI20 test were correlated with participants' performance on objective tests. Scores above 65 are classified as potential DP, specifically, mild, moderate and severe DP with ranges within the intervals $[65, 74]$, $[75, 84]$ and $[85, 100]$, respectively.

3.7 Automatic Face Recognition

Along with the advancement of technology and machine-learning applications, research comparing human and machine performance on face identification has emerged [Phillips and O'toole, 2014]. The National Institute of Standards and Technology (NIST) periodically launches competitions where a benchmark dataset is released and algorithms' performance are measured on identity matching tasks for pairs of still images and videos. Images vary from low-quality handheld pictures (i.e. simulating cell phone captures) to high-quality ones. The idea is to emulate comparison tasks of forensic facial examiners.

The term recognition implies familiarity with a specific face, however, this concept is different in case of algorithms. Familiarity for a human entails the repeated experience of several and diverse images of a person (possibly including changes in viewing angle, lighting and so on) and the creation of

a stable representation of the individual identity, allowing for high flexibility when it comes to poor image quality or appearance transformation [Hole and Bourne, 2010]. For an algorithm, familiarity generally implies the presence of at least one image of the same individual in the gallery set, allowing for a subsequent increased matching accuracy. Face recognition algorithms are not built to identify specific people, their architecture consists of a face detector (i.e. an algorithm capable of identifying a group of adjacent pixels that resembles a face), followed by a feature extraction algorithm that "translates" the input image into a vector of features that relate to the person's identity [Grother *et al.*, 2019b]. When an input is provided to the algorithm it will extract a feature vector. The recognition consists of an estimate of the closeness between the feature vector obtained from the input and any of the feature vector previously stored, if this exceeds a specific threshold, the algorithm will return a match.

These systems are characterised by two types of errors:

- false positives: when an input image not enrolled in the gallery is incorrectly associated with one of more gallery images;
- misses: the algorithm does not return a match when an input image of someone that is present in the dataset is provided as input.

Additionally, there is also the failure of the algorithm to identify a face in the input image (i.e. is not able to detect a face). All of these elements have to be considered and evaluated when establishing the performance of a face recognition algorithm.

The accuracy of facial recognition algorithms varies greatly across the literature. Nevertheless, in recent years this has highly increased reaching misses rates of just 0.1%. These error rates are two orders of magnitude below what they were in 2010 [Grother *et al.*, 2019a], due to the replacement of the old algorithms with those based on (deep) convolutional neural networks (CNNs). However, these low misses rates are attained mainly using high quality cooperative live-capture mugshot images, while poor quality webcam images or unconstrained capture tend to lead to higher false positives and misses rates. According to the latest NIST report [Grother *et al.*, 2019a] (September 2019),

[0]A gallery is defined as a set of features, each tagged with an identity label.

for frontal images with optimal illumination, in at least the 5% of "successful identifications", the similarity scores were so weak that true and false matches would become indistinguishable, requiring a subsequent human adjudication.

Moreover, several experiments have shown that face recognition algorithms tend to over-tune towards the dataset used in the training stage leading to intrinsic bias. An example is the other-race effect discovered for algorithms participating in the NIST Face Recognition Grand Challenge (FRCG) competition [Phillips *et al.*, 2011]: East Asian faces were classified better by algorithms developed by East Asian researchers and, similarly, Caucasian faces were better classified by algorithms developed in Western research centres. The most probable explanation is an over-fitting of the algorithms on locally available data sets [White *et al.*, 2017; Grother *et al.*, 2019b].

The adoption of facial recognition algorithms offers the undoubted advantage to search databases containing millions of images very quickly, although, at this stage of their development, human supervision is still fundamental. For example, Rice *et al.* [2013], selected cases in which state-of-the-art face-recognition algorithms scored 100% incorrectly (i.e. the similarity-score attributed to the image was lowest in same-face pairs or highest in different-face pairs) and submitted these images to human observers. Humans performed well above chance on a one-to-one matching task when considering face internal features only. Furthermore, eye tracking data revealed that, when other information was available, like for example external facial features or other body information, humans were subconsciously exploiting them to support their decision and were consequently achieving higher scores. Whereas algorithms had failed to use them (i.e. they were not programmed to extract and use the extra information).

This is a fast-growing research topic, future development might involve annotated face datasets [Jain *et al.*, 2012], allowing to combine human and algorithm feature-to-feature similarity judgments [Towler *et al.*, 2017b].

3.8 Summary

This chapter has described the cognitive and perceptual understanding of the face recognition process. In particular, the problem related to unfamiliar face matching has been outlined and the factors affecting humans performance in

this task have been shown. The next chapter will provide a critical review of the literature, highlighting how the adoption of advanced imaging techniques such as HDR could be beneficial for performance in the unfamiliar face matching task. A methodology is presented and several research objectives are introduced.

Chapter 4

Methodology

The previous two chapters have provided an in-depth review of the relevant literature. The current chapter considers the main factors relevant to establish the research gaps and outlines the general methodology adopted to addresses the research question.

4.1 Introduction

The accuracy of determining whether two distinct instances represent the same or different individual remains a crucial issue in surveillance and identity verification. Although face recognition algorithms are becoming more efficient, their effective adoption in real-life scenarios still remains controversial [Fussey and Murray, 2019]. For example, the demand for human examiners to perform face identification in the legal system is still extremely high [Spaun, 2011].

Most of the research described in Chapter 3 indicates how the unfamiliar face matching task is susceptible to a number of factors, including imaging quality. Several studies have also identified lighting as a central factor affecting recognition [Braje, 2003; Longmore *et al.*, 2008; Norell *et al.*, 2015; Fysh and Bindemann, 2017; Favelle *et al.*, 2017], however no study has yet explored the impact of dynamic range.

Chapter 2 has described the properties and processes of HDR imaging technology and the added value brought by correctly reproducing images with higher fidelity. This thesis investigates whether HDR can outperform traditional imaging in face matching tasks, especially considering its accuracy

in reproducing lighting in terms of image and display quality [Mantiuk *et al.*, 2016]. Moreover, the work described in this thesis tackles the issue of changes in illumination, as it has been demonstrated that illumination data across both images in a face pair is one of the key elements affecting performance [Hietanen *et al.*, 1992; Braje, 2003; Jenkins *et al.*, 2011; Fysh and Bindemann, 2017]. Given the complexity of human perceptual mechanisms, several steps are required to answer the research question: *"Does HDR imaging improve unfamiliar face matching performance?"*. This chapter will provide a critical analysis of the literature together with an overview of the methodology proposed to tackle the question.

4.2 Critical Analysis and Research Gap

The problem related to face perception is of high complexity, as already illustrated in Chapter 3. In order to test whether HDR imaging can be beneficial to the unfamiliar face matching task, it is essential to examine more closely the factors that can affect performance.

The task of face matching involves both processes of visual perception and face recognition. In the copious body of literature on face perception and face identification it is possible to find experiments focusing solely on manipulation of image characteristics (e.g. resolution [Bindemann *et al.*, 2013], camera-to-subject distance [Noyes and Jenkins, 2017]) or specific factors related to facial identity constancy (e.g. face expression [van der Schalk *et al.*, 2011], time between captures [Norell *et al.*, 2015]), as well as experiments that presented a mix of these factors (e.g. time between captures and different cameras [Fysh and Bindemann, 2018; Jenkins *et al.*, 2011]). These two aspects (i.e. manipulation of image characteristics and facial identity constancy) are orthogonal to each other, therefore, in order to evaluate the HDR technology it is necessary to first isolate imaging manipulation from the facial identity constancy factors. If the technology is proven to be effective, it would be then possible to add more variability.

Historically, experiments on face perception have involved the use of positive-contrast film slides (e.g.[Bachmann, 1991]), or portraits captured with film cameras and printed on paper (e.g. [Bruce *et al.*, 1999; Valentine and

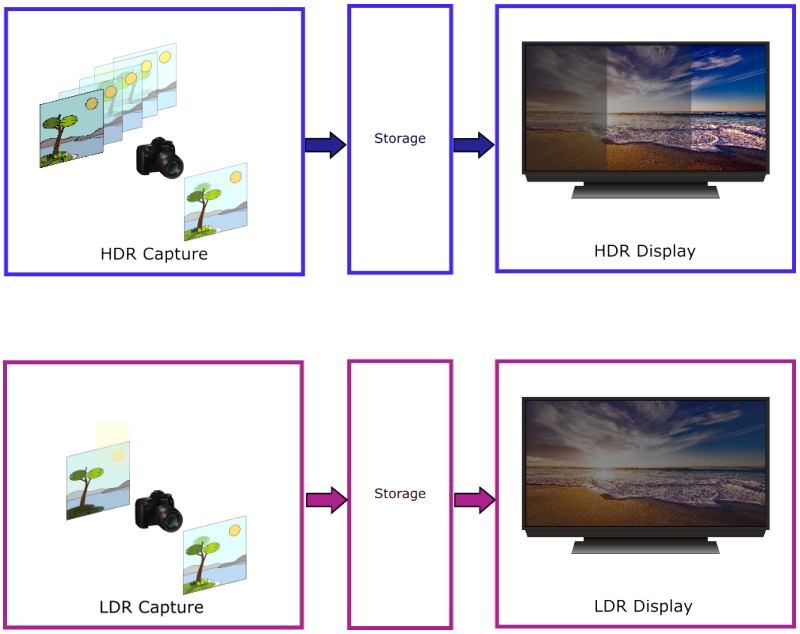


Figure 4.1: Representation of the main stages of the HDR and LDR pipeline.

Davis, 2015]). With the advent of digital cameras and CRT displays first and LCD displays later, the modalities of stimulus presentation have shifted towards the adoption of the LDR pipeline (e.g. [Bindemann *et al.*, 2013]).

Figure 4.1, illustrates the general structure of the HDR and LDR pipelines. These two pipelines have several fundamental differences: the capture stage allows in the case of HDR images to record values that are, in general, linearly related to real world luminance, while pixel values in LDR images are non-linearly related to luminance, and *device-referred* [Mantiuk *et al.*, 2016]. During the storage phase, HDR images are encoded with floating-point accuracy per each colour channel, while LDR only allows the traditional 8-bit encoding. Similarly, for the visualisation step, LDR allows for limited reproduction of colours and luminance values, while new HDR screens can reproduce images at higher luminance (up to 10,000 cd/m^2) and wider colour gamut [ITU-R BT.2100, 2019].

This thesis attempts to answer the question:

> *Does HDR imaging improve the performance of unfamiliar face matching?*

The answer to this research question requires to compare participants'

Image Manipulation	*Bruce1999*	*Braje2003*	*Russell2006*	*Longmore08*	*Glasgow2010*	*Bindemann13*	*Norell2015*	*Korshunov15*	*Noyes2017*	*Kent2018*	*Contribution*
Pixelation - Resolution	Printed 720 dpi	150 pixels/face	Max 700 vertical pixels	NO	Max 350pixels/face	Variation (3 levels)	Set 1: 386 pixels/face Set 2: 114 pixels/face Set 3: 68 pixels/face	Variable: max 500 pixels/face min 100pixels/face	Variable: Constant interocular distance of 150 pixels	Set 1: 283 pixels/face Set 2: 142 pixels/face	High Resolution (1200 pixels/face)
Lighting Conditions	Controlled	Controlled	Uncontrolled	Controlled	Controlled	Controlled	Controlled vs Uncontrolled	Uncontrolled	NO	Uncontrolled	Controlled
Changes in Lighting Direction	NO	YES: 45° top +45°, -45° left/right	NO	YES	NO - Frontal	NO - Frontal	Not Specified	Variable - uncontrolled	NO	Not Specified	YES: 90°,-90 Top/below +90°, -90° left/right
Subject-to-camera distance	FIX	NO	Variable - uncontrolled	FIX	FIX	FIX	YES: 1st: 3 m 2nd: 5 m	Variable - uncontrolled	YES: 1st 0.32 m 2nd 0.32 m	Variable - uncontrolled	FIX
Camera Calibration	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES
High Dynamic Range	NO	NO	NO	NO	NO	NO	NO	NO - TMO only	NO	NO	YES
Viewpoint	Frontal and 30 degree angle	7 and 11 degree with respect to frontal view	Frontal	4 different	Frontal	Frontal	Frontal	Frontal	Frontal	Frontal	Frontal

Table 4.1: This table compares the image characteristics of the stimuli used in experiments on unfamiliar face matching. The last column represents the current work contribution.

performance using native LDR and HDR stimuli when displayed on HDR and LDR displays. Only two HDR facial databases are available in the literature, the one presented by Ige *et al.* [2016] and the one created by Korshunov *et al.* [2015]. Unfortunately, they were not considered suitable for use in a controlled experiment: Ige *et al.* [Ige *et al.*, 2016] due to the lack of information regarding camera radiometric calibration and Korshunov *et al.* [Korshunov *et al.*, 2015] due to the uncontrolled environment and unspecified lighting conditions. Moreover, as seen already in the literature, it is important to control for factors such as camera-to-subject distance and colour calibration [Noyes and Jenkins, 2017; Russell *et al.*, 2006; Johnston *et al.*, 1991; Norell *et al.*, 2015; Ritchie and Burton, 2017; Jenkins *et al.*, 2011] as they can have a considerable impact. Neither of the above dataset controlled for such factors. Therefore, the need for the creation of an HDR facial database was identified as a first and fundamental step. The Warwick HDR Face Dataset presented in Chapter 5 aims to address these limitations.

Illustrated in Table 4.1 are the most salient image features that have been manipulated in unfamiliar face matching experiments. It is clear that technical aspects such as camera radiometric calibration and the use of the HDR imaging pipeline end-to-end have not been explored yet.

Image resolution in facial stimulus datasets has increased over the years, due to improvement in camera technology and decreasing costs. Arguably, the image resolution has not been regarded as fundamental perhaps due to experiments performed, amongst others, by Bachmann [1987] and Bindemann *et al.* [2013], where recognitions rates ranged above 75% with resolutions of 96 pixels/face. However, the resolution factor is not the only one to be taken into account. Almost all of the datasets available in the literature present images encoded in *.jpeg* format. This is a lossy format where a heavy quantisation is performed (depending on the desired quality), possibly resulting in further loss of image quality, especially in the range of higher frequencies (i.e. loss of fine detail) [Wiseman, 2015].

The *.jpeg* format, is often produced, due to bandwidth limitations, by cheaper camera modules. These cameras essentially perform on board:

- some form of tone-mapping to produce gamma-corrected pixel value;

- *.jpeg* compression.

Both these operations are irreversible and have the unfortunate disadvantage of introducing distortions and reducing the dynamic range of the images [Mantiuk *et al.*, 2016].

The camera used for the research in this thesis is a DSLR camera which offers the option of capturing images in RAW format. This is the digital equivalent of a film negative, presenting the values registered by the CCD sensor. This allows for higher fidelity and more freedom in the subsequent processing and manipulation of the images without loss of detail. In this work, face stimuli are available in high resolution (min 1200 pixels/face) and in RAW format, for the individual exposures, and in *.hdr* (a lossless format), for the merged version[1], preserving extra detail as well as the luminance and the chrominance information.

Although the effect of different lighting direction on the perception of faces has been explored in previous experiments (see for example [Braje, 2003; Longmore *et al.*, 2008]), the use of HDR imaging minimises the presence of saturated and underexposed pixels, allowing for a more accurate reproduction of the scene lighting. Moreover, the radiometric calibration process performed allows for the encoding of luminance values that are *scene-referred*, as opposed to the *device-referred* values available in other datasets.

The next phase of the work is the comparison between the HDR and the LDR pipeline. Stimuli from the newly created Warwick HDR Face Dataset were used to compare the HDR and the LDR pipeline from capture to display. The procedure adopted in the experiment *HDRvsLDR* was similar to other experiments on unfamiliar face matching such as those used by Bindemann *et al.* [2013] or Burton *et al.* [2010].

A further step of this research consisted in comparing the two most sensitive stages of the two pipelines: the capture and the visualisation stages. If a mixed solution is to be proposed as a transitional stage before the whole LDR pipeline is upgraded to HDR, the question of which part of the pipeline it would be more important to update is answered with the experiment *Backwards-Compatibility HDR*. On the one hand, it would be preferable to preserve old

[1]The HDR data capture has been performed using the bracketing technique described in Section 2.6.1

LDR datasets and "upgrade" them to HDR, however, on the other hand, it might be desirable to visualise HDR images on LDR displays when HDR is still not available. In essence we ask the question: *"if the budget does not allow for a complete renovation of the pipeline, which part should be prioritised in the upgrade process?"*

As a final step, the results of the *HDRvsLDR* and *Backwards-Compatibility HDR* experiments are compared to answer the question as to whether the full HDR pipeline is necessary or a partial upgrade of the LDR pipeline with specific HDR elements is sufficient to improve performance on the unfamiliar face matching task.

To summarise, the specific objectives of the research presented in this thesis can be described as follows:

1. Review of the literature regarding both the HDR technology pipeline and the processes underlying face perception (completed in Chapter 2 and Chapter 3);

2. Creation of a radiometrically calibrated HDR faces dataset: the Warwick HDR Face Dataset (Chapter 5);

3. Use of the Warwick HDR Face Dataset stimuli in a face matching experiment (the *HDRvsLDR* experiment) aimed at measuring accuracy and reaction times performance of HDR vs LDR imaging (Chapter 6);

4. Adoption of the Warwick HDR Face Dataset stimuli in a face matching experiment (the *Backwards-Compatibility HDR* experiment) aimed at measuring accuracy and reaction times performance of an enhanced LDR pipeline through the application of advanced imaging algorithms - i.e. Expansion Operators and Tone Mapping Operators (Chapter 7);

5. Comparison of the results obtained in the Objective 3 and 4 to establish if the adoption of the full HDR pipeline or a partial upgrade of the LDR pipeline is adequate for sufficiently improving performance (Chapter 8).

4.3 Proposed Approach

The following sections provide a detailed description of the specific methodology adopted to achieve each one of the specific objectives (see Figure 4.2 for an illustration of the methodology).

4.3.1 Warwick HDR Face Dataset

Chapter 5 describes the creation of an HDR face stimuli dataset, the Warwick HDR Face Dataset, for use in digital forensics, psychological or visual perception studies. This dataset is the first of its kind. The dataset offers a set of 170 images radiometrically calibrated HDR portraits. Seventeen male subjects are portrayed. For each person portrayed five lighting conditions are available and for each lighting conditions two different image captures are made available. As HDR provides more accuracy in reproducing light, in addition to the traditional full lit image used in other similar datasets for face matching (see for example the Glasgow Face Matching Dataset [Burton *et al.*, 2010]), four additional lighting conditions (i.e. light from right, left, top and below) are created in the studio at the moment of the capture. The availability of several lighting conditions allows for the full exploitation of the HDR imaging characteristics in accurately capturing light, avoiding over- and under-exposure.

4.3.2 Subjective Evaluation of High vs Low Dynamic Range Imaging for Face Matching

Chapter 6 details how the Warwick HDR Face Dataset is used to compare the HDR and LDR imaging pipelines in a subjective experiment: the *HDRvsLDR* experiment. The availability of this new stimulus set allowed the evaluation of performance (i.e. reaction time and accuracy) with LDR vs HDR images in a matching task. Moreover, it allowed the effects of different types of lighting on face processing and the potential interaction between differences in lighting and LDR/HDR imagery to be assessed, [Braje, 2003; Longmore *et al.*, 2008].

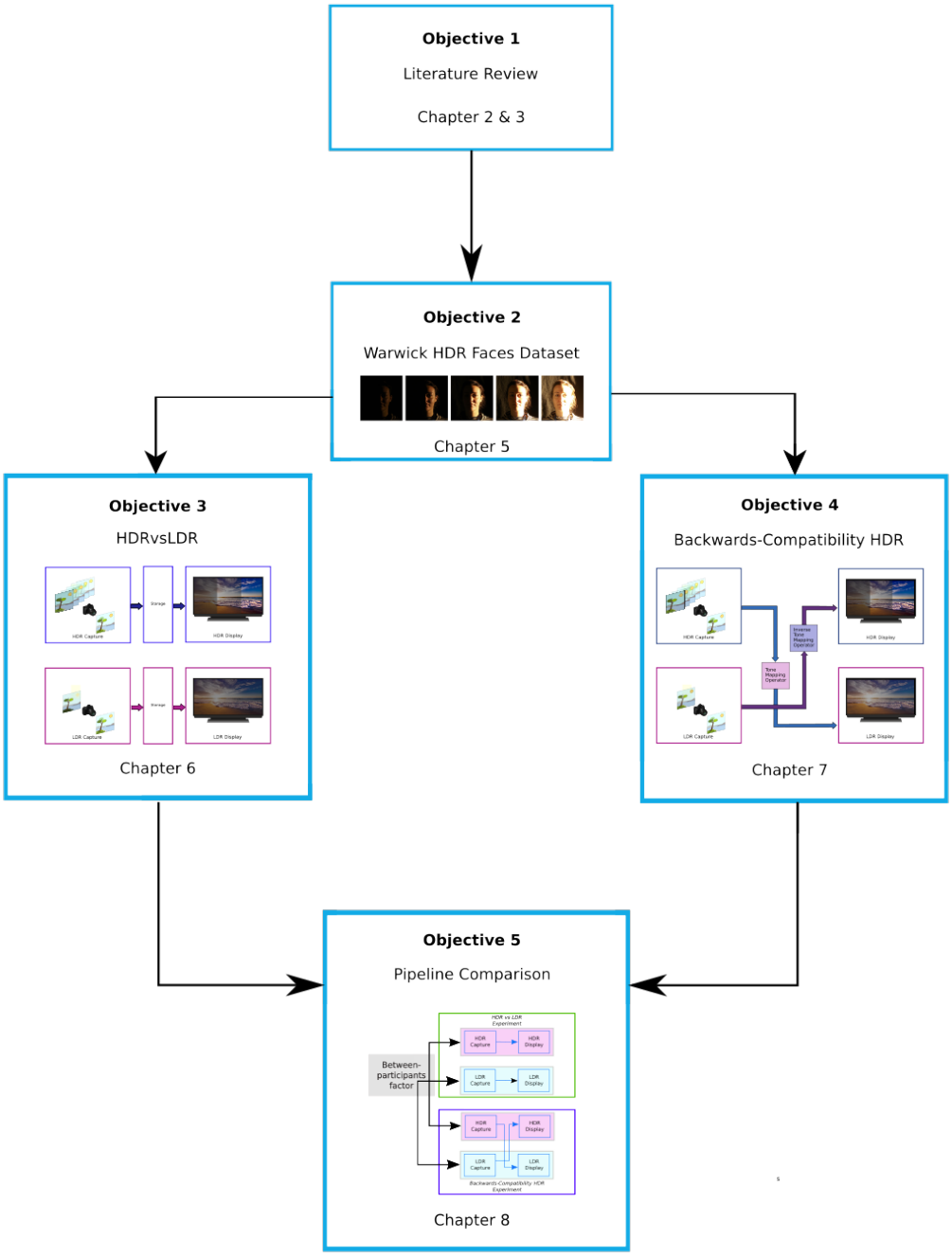


Figure 4.2: Illustration of the methodology adopted in this thesis. Chapter 5 will describe the creation the an HDR facial stimuli set. Chapter 5 will describe the *HDRvsLDR* experiment while Chapter 6 will describe the *Backwards-Compatibility HDR* experiment. Both experiments make use of the newly created Warwick HDR Face Dataset in subjective experiments. Chapter 8 will compare across the results of the two experiments.

4.3.3 Backwards-Compatibility HDR

In view of the positive results obtained from the experiment illustrated in Chapter 6, a new experiment was designed to investigate backward compatibility of HDR imaging for face matching tasks: the *Backwards-Compatibility HDR* experiment. Specifically, if this process has to be done in stages, which part of the pipeline would be more beneficial to upgraded; capture vs display.

Considering the fact that HDR is a mature albeit expensive technology, it is realistic to speculate on the possibility of complementing the LDR pipeline using specific HDR elements (i.e. cameras, displays) with the purpose of improving face matching performance. The investigation introduced in Chapter 7 attempts to answer the question of whether it is more effective to upgrade existing LDR stimuli by adopting an Expansion Operator and visualise the facial stimuli on HDR displays, or if it is more advantageous to upgrade the capture stage in the pipeline and adapt HDR imaging to be visualised on LDR screens by adopting an appropriate TMO algorithm.

Analogously to the *HDRvsLDR* experiment, in the *Backwards-Compatibility HDR* experiment performance (reaction time and accuracy) in a face matching task was evaluated when HDR and/or LDR face stimuli are provided. The HDR and LDR images were obtained respectively using the an Inverse TMO and a TMO algorithm. The availability of face stimulus portraits under different lighting conditions allowed, once again, the assessment of the effects of differences in the facial illumination.

4.3.4 Pipeline Comparison

Chapter 8 contains a statistical analysis where the results presented in Chapter 6 and Chapter 7 are compared using a mixed-method approach. The results obtained from the *HDRvsLDR* and *Backwards-Compatibility HDR* have demonstrated improvement of participants' performance due to the use of HDR imaging while performing unfamiliar face matching. A further necessary step is to establish a comparison across the two pipelines modality. Chapter 8 answers the question of whether a full HDR pipeline is required in order to achieve high performance in the face matching task, or a partial upgrade of the LDR pipeline with HDR elements is sufficient to achieve optimal results.

4.4 Summary

This chapter focused on the motivation for this research and highlighted important gaps in the literature. In order to answer the question related to the use of HDR vs LDR pipeline for the unfamiliar face matching task several steps are necessary. This chapter provided a description of the proposed methodology, offering a roadmap for the work described in the following chapters and briefly introduced the content and aim of each of them.

Chapter 5

Warwick HDR Face Dataset

The previous chapter has introduced the general methodology for this thesis. In order to answer the research question, the first objective is the creation of an HDR facial stimulus set. The present chapter describes the steps followed for the creation of the Warwick HDR Face Dataset.

5.1 Introduction

Facial stimuli are essential for face perception experiments and a vast *corpus* of literature has been created over time. Each dataset presents peculiar characteristics which make it more or less suitable for a specific experiment. One of the typical experimental designs adopted in face matching is the two-alternative forced-choice, where participants are presented with two facial stimuli and requested to answer to the question *"Is this the same person or not?"* .

When investigating face matching, it is of utmost importance to elicit face-related processing, whereas it is almost never appropriate to mix it with image comparison processes [Burton, 2013], which can be applied to any stimulus[1]. For this specific reason, it is necessary to have at least two instances (i.e. portraits) for each subject present in the dataset.

As already seen in Chapter 3, although many faces dataset are available

[1]Whether faces are a "special" kind of object eliciting innate specific cognitive processes or not it is still being debated, however, they certainly occupy a privileged position in our life regardless if this is an acquired expertise or not [Hole and Bourne, 2010].

in the literature, quite often they lack the availability of two pictures portraying the same face and same pose, especially those focusing on facial expressions (see for example Lyons *et al.* [1998] or Ma *et al.* [2015a]). Moreover, other limitations are related to:

- absence of colour grading;
- absence of details related to camera and lenses characteristics;
- unknown lighting conditions present in the scene at the moment of the capture;
- unknown face-to-camera distance.

Many researchers have indeed oriented their research interests towards the use of "ambient images" [Ritchie and Burton, 2017] (i.e. images captured in uncontrolled conditions), suggesting that this is key to better recognition of a novel photo seen later [Jenkins *et al.*, 2011; Ritchie and Burton, 2017]. The datasets generated in these contexts, therefore, include portraits under uncontrolled lighting conditions, high variability among the different pictures and, often, time difference between captures [Jenkins *et al.*, 2011; Burton, 2013; Fysh and Bindemann, 2018]. High variability has the undoubted advantage to emulate more realistic scenarios, where images from two different sources are to be compared (e.g. CCTV vs a police mugshot), and it has been used in the past for experiment aiming at answering face learning processes [Ritchie and Burton, 2017]. However, the presence of too many variables might limit the conclusions drawn from an experiment, if the aim is to identify specific image characteristics (e.g. light, lenses or pose).

Considering previous research on the sensitivity of unfamiliar face matching to variability in image characteristics, the dataset created in this thesis aimed at limiting as much as possible other image manipulations, such as variations in time or expressions, so that the conclusions drawn on the image comparison could be ascribed solely to the effect of Dynamic Range and lighting.

Previous literature has highlighted the negative impact that changes in lighting has on face recognition [Braje, 2003; Longmore *et al.*, 2008; Johnston and Bindemann, 2013]. Johnston *et al.* [1991] investigated the effect of lighting

direction, inversion and brightness reversal (i.e. negative images) on recognition of familiar faces. Even if the faces were known to the participants, changes in lighting produced high error rates (up to 30%).

Several studies have already shown that, while scanning generic images, darker low-frequency regions (i.e. shadows) are most often avoided. Faces are typically gazed at using stereotypical fixation patterns [Althoff and Cohen, 1999; Peterson and Eckstein, 2012]. Hermens and Zdravković [2015], conducted an eye movement study on facial stimuli with shadowed regions. In one experiment, they created artificial shadows (i.e. using Photoshop) across half of the face on images extracted from the FACES database [Ebner *et al.*, 2010] and the Radboud Face Database [Langner *et al.*, 2010]. The shadow divided the face in two halves: one half with relatively high contrast (i.e. average luminance of $31.6cd/m^2$), while the other half with low contrast (i.e. average luminance of $7.56cd/m^2$). In a second experiment, they used uncontrolled stimuli (i.e. images publicly available online) that presented a side illumination. Their results showed a deviation from normal fixation patterns of the participants' eyes movement: the participants were redirecting the gaze so to avoid the area with the lower contrast.

Hermens and Zdravković's experiment, together with several other studies (e.g. [Braje, 2003; Longmore *et al.*, 2008; Favelle *et al.*, 2017]) on the influence of light suggests how the availability of facial stimuli presenting variable lighting conditions could be a useful resource to further investigate face perception processes. For this reason, multiple lighting conditions - five - were included in the Warwick HDR Face Dataset. Moreover, the creation of harsh shadows represents an exemplar case where HDR imaging can be used to its full potential. In fact, as opposed to LDR, HDR can recover details in both well and poorly exposed areas of the image.

5.2 Design

The Warwick HDR Face Dataset serves as a basis for conducting the two face matching experiments described in Chapter 6 and Chapter 7. The aim of the dataset is to have a number of faces captured in HDR under diverse lighting conditions which are radiometrically calibrated and encode absolute luminance

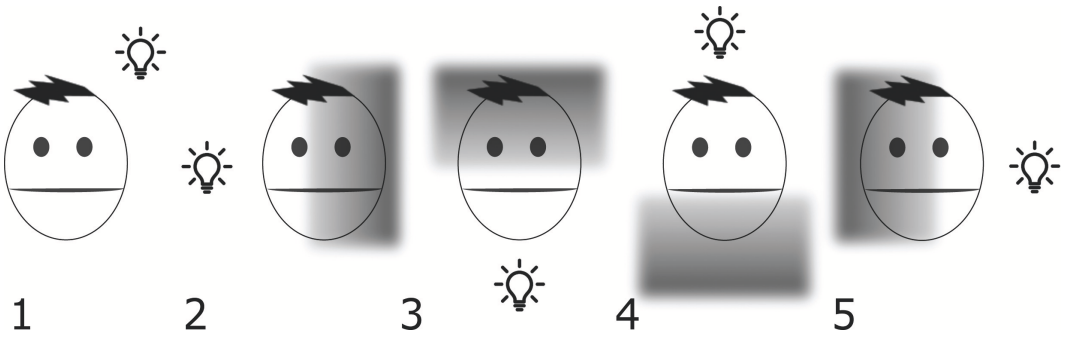


Figure 5.1: Warwick HDR Face dataset. Schematic representation of the different lighting conditions created for the Warwick HDR Face Dataset. 1. Full lit face - passport-like shot; 2. Light source located on the left side; 3. Light source located underneath the participant's face; 4. Light source located on top of the participant's head; 5. Light source located on the right side.

values.

In order to be able to assess human perceptual performance at matching faces under challenging lighting conditions five different illumination patterns were chosen, schematically represented in Figure 5.1. Lighting during the capturing has been arranged in order to obtain harsh lighting conditions. This has a dual purpose: exploiting the intrinsic characteristics of HDR imaging when capturing a wide range of light as well as disrupting, in certain cases, normal illumination patterns, making the stimuli more challenging from a perceptual point of view.

The data collection has been performed so that light sources distances and intensity were standardised. For the five lighting conditions in the dataset, all the distances between the participant and the light were carefully chosen in order to meet limitations in the geometry of the laboratory and, more importantly, to respect international safety standards and good practice [European Commission, 2012] so that there were absolutely no short or long term negative effects for the participants. The lighting conditions *2* and *5* only differ in terms of direction to account for the fact that complete symmetry is very rare in human faces [Majumdar and Sinha, 1992]. The parameters selected for each lighting setup (i.e. distances and directions) are specified in Figure 5.2.

The distances and camera-subject relative positions were selected to minimise deformations and lens flare as well as maximising the sensor's area

where the useful information (i.e. the face) was located. In order to ensure accuracy and reproducibility of this study, the camera to capture the stimuli was radiometrically calibrated following the procedure described in Section 5.5.

The HDR images were created with the exposure bracketing technique (see Section 2.6.1). This method requires the subject being photographed to remain still for a few seconds (max 6s). Therefore, it was decided to record only a neutral pose, as it is quite challenging to maintain any other facial expression for a very long time.

In total 170 facial HDR images of seventeen males photographed in neutral pose over five lighting conditions were captured. For each lighting condition, two HDR images were captured.

5.3 Participants

The participants were recruited on a voluntary basis through advertising on the Department's notice-boards at the University of Warwick. Seventeen participants consisting of employees and students were recruited. The participants were aged between 20 and 52, eleven *White* (*British* and *other White background*), six Asian (*Indian* and *Pakistani*) - ethnicity classification according to the UK governmental specifications (see [UK Ethinicity, 2011]).

5.4 Procedure

The data capture was performed in a dedicated room within the International Digital Laboratory at the University of Warwick. This room has completely black walls, so as to avoid any unwanted light interference. Five different photographic stations were arranged, each one with a seat and specific lighting. After explaining the procedure and collecting the participants' consent, they were asked to sit comfortably, in turn, in each of the station's chair, assuming a neutral expression. Using a tripod an operator performed two captures of the sequence of five images at different exposures (see Figure 5.3).

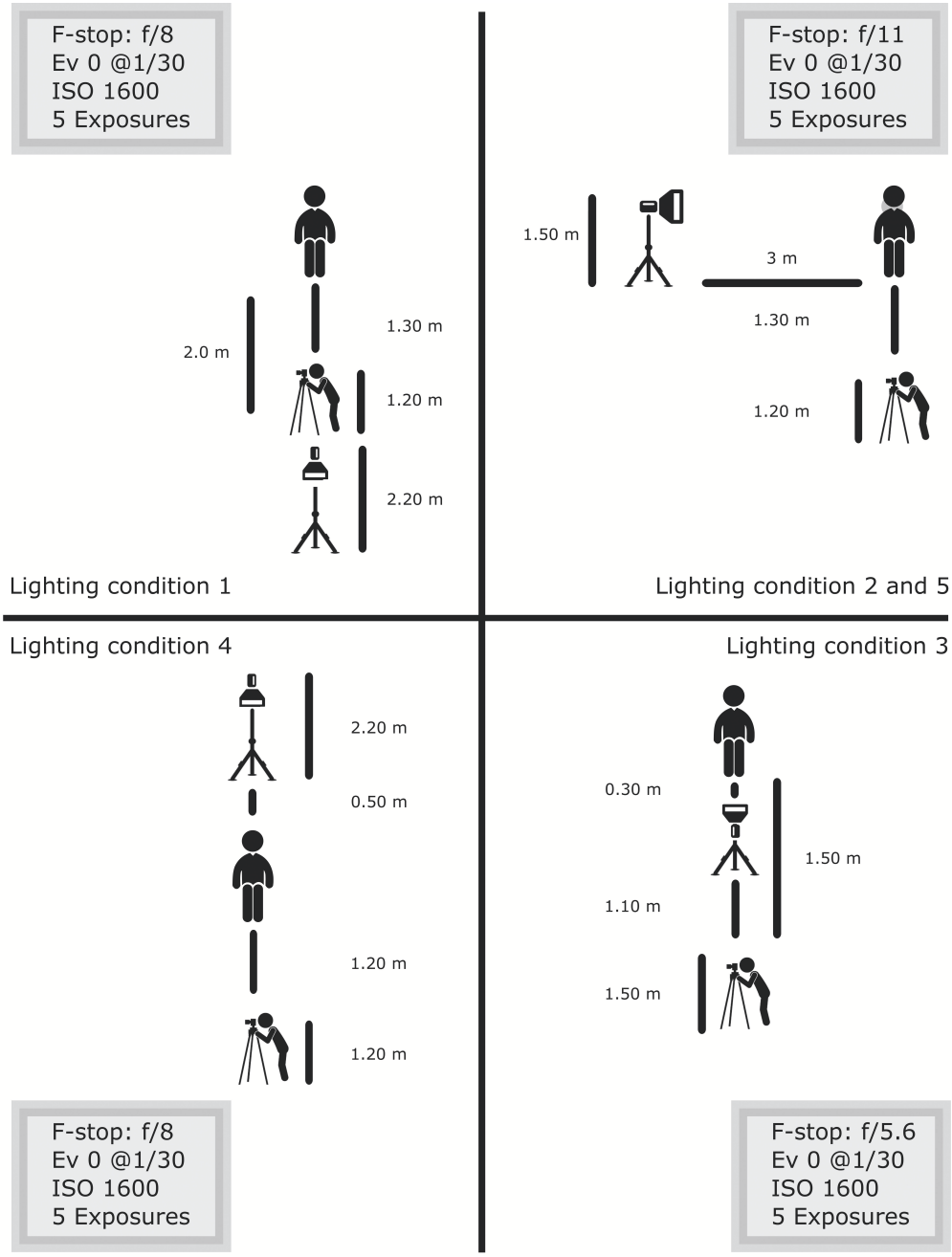


Figure 5.2: Diagram with details of distances and camera settings adopted for the Warwick HDR Face Dataset capture. Clockwise, from the top left corner: Light source in front of the participant - full lit face; Light source on the left/right of the participant; Light source underneath the participant's face; Light source on the top of the head of the participant.

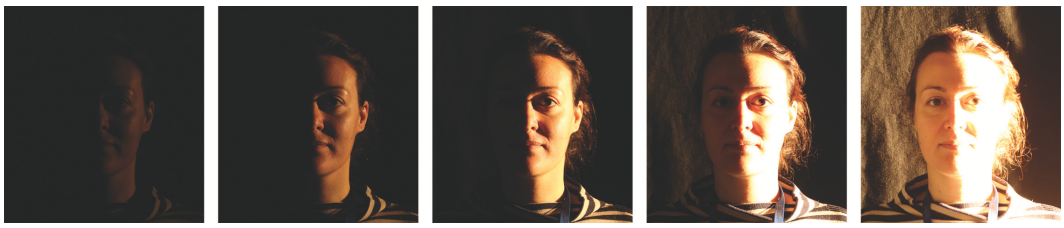


Figure 5.3: Example of the five different exposures used for image bracketing. It is not possible to show the actual images used in the experiment in order to protect the participants' privacy.

5.5 Materials

Lighting The lamps adopted for the lighting conditions *2* and *5* are ARRI Lite plus 2000W (the minimum suggested safety distance is 2m). For the light position *4* and *1*, the standard studio illumination image, a light with smaller power has been employed, in order to minimise distress to the participants. Specifically, an Interfit INT184 Stellar X Tungsten 500 Watt was used. This is a lamp normally used in photographic studios, set on a tripod and raised at 2.20m, the maximum height allowed by the ceiling. For light position *3*, considering the closer distance to the participant's face, a halogen desk lamp (50W) was used. The diagram in Figure 5.2 shows the details on all the specific distances.

Camera Setup The camera adopted for the shooting is a Canon EOS 5D Mark III with Canon EF 24-105mm f/4L IS USM Lens fixed at 50mm with a HOYA PRO1D UV Filter. The images were recorded in RAW format (14 bits).

Camera Calibration The camera was calibrated in order to be able to use the camera sensor as a light-meter and therefore correlate real-world lighting to each HDR image pixel value. This calibration was conducted following a procedure similar to the one proposed by Kim and Kautz [2008]. In their experiment, they created an *ad hoc* transparent target by photographically enlarging the IT8.7/1 colour chart [ANSI IT8.7/1-1993 (R1999), 1999] onto Kodak Ektachrome professional film (8-by-10 inch).

The difference, in this case, is that only the luminance channel was

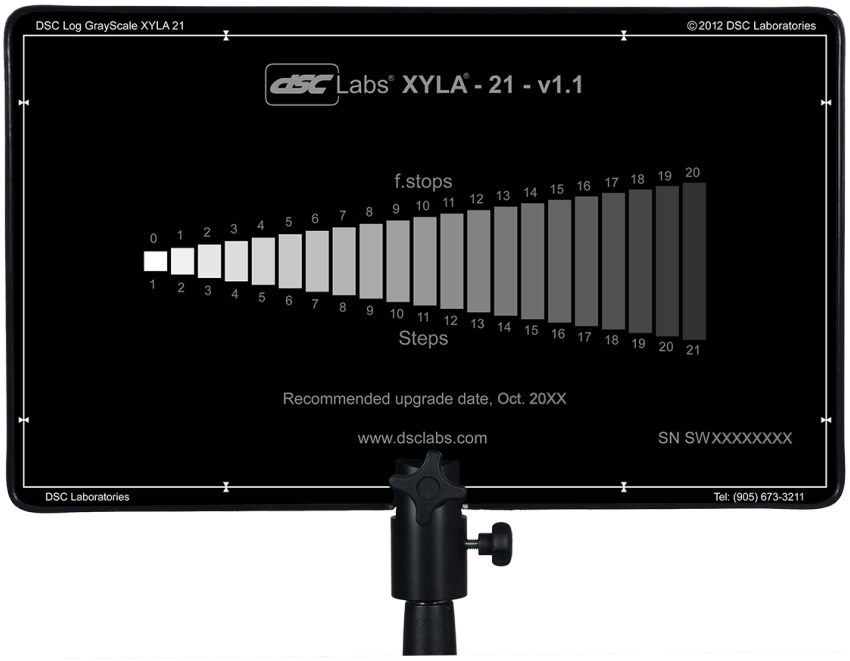


Figure 5.4: Xyla-21 Chart, [DSC-Laboratories]. 20 f-stops of dynamic range (21 steps) with regulated backlight source.

considered and replaced their photographic target with a Xyla chart (Figure 5.4). This is a 21 stepped xylophone shaped back-illuminated chart offering a DR of 20 f-stops (i.e. the brightest bar is 2^{20} times brighter than the darkest bar) [DSC-Laboratories]. This is commonly used to evaluate cameras' dynamic range.

The procedure followed for the camera calibration was:

- Capture a sequence of seven exposures of a Xyla Dynamic Range Test Chart;
- For each of the single exposure bars collect accurate measurements of the luminance values using a Minolta LS-150 luminance meter [Konica Minolta];
- Reconstruct the CRF of this image with the Robertson's method [Robertson *et al.*, 1999] (see Section 2.6.2) using the pfstools 2.10 software library [Mantiuk *et al.*, 2007];
- Reconstruct an HDR image from single exposures of the Xyla chart using

pfstools 2.10;

- Map the real-world luminance values collected with the luminance meter to HDR pixels' values;

- The luminance in the generated HDR images corresponds to real-world values in cd/m^2. The camera response function calculated in this way is then applied to the processing of the entire stimuli set.

Image Processing After the image capturing phase, the participants' portraits were processed using pfstools 2.10. The bracketing sequence merging was implemented so as to minimise alteration of the raw data. Linear merging of the RAW 14 bits images using the previously mentioned CRF was performed to produce HDR images.

The resulting images were then filtered to remove singularities with the CleanWell function from the MATLAB HDR Toolbox [Banterle *et al.*, 2017] and then cropped in a passport-style fashion to 4,320 × 3,370 pixels - height × width ratio: 0.78.

5.6 Dataset Properties

The uniqueness of this dataset resides in:

- Camera calibration;

- Representation of absolute luminance values within the HDR files;

- Controlled lighting positions;

- Availability of two different images of each person for five lighting positions.

The average Dynamic Range in the images depends on the lighting position. The average values are provided in Table 5.1. The light intensity has been adjusted to meet limitations in the geometry of the laboratory and to be within the recommended intervals to avoid any harm to the participants' eyesight.

Table 5.1: Average Dynamic Range for the Warwick HDR Face Dataset

Lighting	Dynamic Range	F-stops
(1)	150	7.09
(2)	1224	10.14
(3)	482	8.70
(4)	213	7.39
(5)	1838	10.78

5.7 Discussion

This chapter described the creation of an HDR facial stimulus dataset. A thorough search of the relevant literature revealed this to be the first available dataset in its genre. Its creation was in response to one of the main issues with HDR imaging at the moment: apart from CGI, very little content is currently available. Although exposure bracketing is used by photographers, a limitation is presented by the end devices (print, standard screens), so the content is tone-mapped in most cases, rather than stored and displayed in HDR format. The process of camera calibration and the accuracy towards technical aspects of the image capture phase makes this dataset ideal when studies require accurate luminance reproduction.

Although the dataset presents a limited size and only one sex is represented, further effort is currently being made to expand it. Two female participants were also recruited, but it was decided to remove female faces from the dataset in subsequent experiments and use the female faces only for the practice trials, as experiments in the field have shown how sex categorisation is completed generally more quickly (between 5 to 7 milliseconds [Bruce *et al.*, 1993]) than familiarity judgement Bruce and Young [1986]; Wild *et al.* [2000]. Recognition requires recruiting mechanisms to extract and encode details that make the specific face unique, rather than simply categorise it (e.g., as male or female), by extracting and encoding the information shared with most same-sex faces. While not fully comprehensive, the dataset has been prepared in a fully controlled environment both in terms of lighting and image capture.

5.8 Summary

This chapter has described the design and procedures followed during the creation of the Warwick HDR Face Dataset. This was a necessary step towards answering the research question of whether HDR can help improve face matching accuracy. The next chapter will describe the use of the dataset in a subjective experiment where the HDR and LDR imaging pipelines were compared.

Chapter 6

An Evaluation of High vs Low Dynamic Range Imaging for Face Matching

The previous chapter has introduced the procedures followed for the creation of the Warwick HDR Face Dataset. The current chapter will describe the motivation and the methodology adopted to perform a controlled subjective experiment (N =39) where HDR stimuli are compared to LDR in an unfamiliar face matching task. The aim of this experiment is to establish whether HDR imaging is beneficial to improve participants' performance.

6.1 Introduction

Unfamiliar face matching is a critical task especially in contexts like surveillance and forensics. As emphasised by the detailed review on the topic of forensic face matching by Fysh and Bindemann [2017], possible impostors or mismatches constitute an increasing security concern. As already introduced in Chapter 3, a consistent amount of research on perceptual psychology [Megreya *et al.*, 2011; Burton, 2013; Kemp *et al.*, 2016; Young and Burton, 2017] has shown that unfamiliar face recognition is very error-prone, and even experienced professionals are subject to false identification. This is further compounded when the quality of the stimuli is poor [Burton *et al.*, 1999; White *et al.*, 2014;

Norell *et al.*, 2015; Young and Burton, 2017]. The only exception is represented by the so-called super-recognisers: a very small group of untrained individuals whose face recognition abilities are far superior to the norm [Russell *et al.*, 2009].

Traditional low (or standard) dynamic range (LDR) imaging does not preserve reliably the totality of real-world lighting resulting in images which lack the contrast and luminance present in the real world. HDR imaging is, on the other hand, capable of dealing with the acquisition, processing, storage and display of higher quality images. The experiment introduced in this chapter, the *HDRvsLDR* experiment, aims at tackling one of the main problems often pointed out by the facial recognition literature: the accuracy in reproducing reality, in terms of luminance and image and display quality [Hole and Bourne, 2010; Norell *et al.*, 2015; Fysh and Bindemann, 2017]. Furthermore, it evaluates whether the improved accuracy in reproducing reality given by the HDR pipeline, equates to better performance in unfamiliar face matching tasks. The results of this experiment could result in more dependable data and decision-making with greater precision.

6.2 Methodology

This section presents the methodology adopted in the *HDRvsLDR* experiment. The main objective of this work is to form an understanding of perceptual sensitivity to differences in luminance while performing a same/different identity judgment. More specifically it aims to:

- Evaluate face matching performance and accuracy when the dynamic range reflects reality more accurately like in the case of HDR;

- Evaluate performance and accuracy with faces exposed to same/different lighting.

Results will show whether HDR can help improve the accuracy and speed when performing unfamiliar face matching tasks where high precision is required.

6.3 Design

The goal of this experiment is to verify and quantify the benefits brought by the adoption of HDR face stimuli for facial image comparison. In order to achieve this, a forced choice experiment was designed, where participants were presented with two concurrent faces and had to decide if it was the same person or not by pressing a button on a keyboard.

The test aimed at verifying performance or reaction time (termed *time*) and accuracy (termed as *accuracy*) while varying the dynamic range (labelled *DR*) of the stimuli and the lighting condition represented in the stimuli (*position*). *time*, measured in seconds, and *accuracy*, measured by the percentage of correct answers, are the dependent variables (DV) while *DR* and *position* are the independent variables (IV).

For the independent variable of the dynamic range (*DR*), three scenarios were tested:

- LDR vs HDR (*mix*): one of the two faces is an LDR while the other is an HDR stimulus;
- HDR vs HDR (*HDR*): both faces are encoded as HDR stimuli;
- LDR vs LDR (*LDR*): both faces are encoded as LDR stimuli.

DR is a within-participants variable. In order to appreciate the impact of the lighting on the matching task a second within-participants independent variable (*position*) was evaluated:

- Same lighting (*same*): both faces are lit by the same light source both in terms of intensity and direction;
- Different lighting (*different*): the two faces are lit by different light sources in terms of intensity and direction.

There were three general hypotheses:

- H_1^1: *HDR* will outperform other modalities (i.e. *LDR* and *mix*) as it provides an intrinsic advantage due to the higher fidelity in which real-world lighting is reproduced;

- H_1^2: reaction time (*time*) is less affected by variation of light position when stimuli are in HDR;

- H_1^3: accuracy (*accuracy*) is less affected by variation of light position when stimuli are in HDR.

Each participant was presented with a total of 60 trials, each trial contained a permutation of the three *DR* scenarios. Each block contained twenty images presenting two concurrent faces under *same* or *different* lighting equally split between same-person pairs and different-person pairs. No feedback was provided to the participants regarding their accuracy while performing the task. The use of the two keys to indicate a match/mismatch were assigned randomly to participants and counterbalanced across the experiment.

To avoid the familiarity effect, particular care was taken so that a specific face was not shown more than 10 times [Longmore *et al.*, 2008] throughout the trials. Also, to avoid picture comparison, even when the *same* lighting same-person pair was presented the two different image capture available in the dataset were used (see Chapter 5 for details on the dataset).

The faces depicted were cut out with an elliptic shape similarly to the approach taken by Megreya *et al.* [2011], so that only the internal features were visible. Previous research has shown that external features (hair, head outline, neck and shoulders) disrupt face recognition ability [Longmore *et al.*, 2015]. Internal features, on the other hand, are less susceptible to change throughout time and they constitute a high number features used by experts to perform an identification task [FISWG] (e.g. Scars/Blemish, Eyes, Nose, Mouth, Mouth Area, Forehead, Cheek Area). Care was taken to ensure the elliptic shape had the same identical dimension for every face portrayed in order to avoid participants subconsciously to be able to identify faces based on the external shape and not by looking at the facial features [Young and Burton, 2017]. The faces portrayed on the screen occupied 7 degrees of vertical visual angle, which is within the intervals adopted in the literature for experiments on face perception (see [Peterson and Eckstein, 2012; Ghuman *et al.*, 2014; Goffaux and Greenwood, 2016]).

Following the literature [Burton *et al.*, 2010; Bobak *et al.*, 2016; Megreya *et al.*, 2011], only the luminance channel was used for the images, although

the images, in this case, were HDR grayscale (i.e. they have a far higher number of shades of grey than 256). This was done in order to limit additional variables given by the two colour channels and also considering that chroma components seem to be less relevant in the recognition process [Hancock *et al.*, 2000; Longmore *et al.*, 2008]. From the HDR stimuli only the luminance channel was used. No further manipulation was necessary due to the nature of the Warwick HDR Face Dataset.

To obtain LDR stimuli[1] from the HDR images, a TMO was needed. The Optimal Exposure algorithm described in Section 2.9 was selected. This algorithm operates in a similar fashion to what most camera's embedded software would do when taking a shot: adjusting the exposure to maximise the information contained in the image histogram. The target display for the LDR images was characterised by luminance in the range between 0 and 300 cd/m^2. This is the typical range provided by most off-the-shelf LED screens.

Lighting in the room was controlled and enough time was allowed to ensure that the participant's eyes were properly adapted to the room luminance levels. The participants were seated at 185 cm distance from the screen. This is the best viewing distance given the size of the screen (see Recommendation [ITU-R BT.2022, 2012]).

A black screen with a central white fixation cross was alternated to the stimulus for 1500ms. This is standard practice in perceptual experiment: the use of visual masks to manipulate the time course of processing allows the participant retinal image to be refreshed (i.e. backwards masking [Cowan, 2008]).Although the choice of inter-stimuli background colour is a relevant question for perceptual psychology experiments, very little work has been done in this direction. One of the few related investigations was conducted by Freeman *et al.* [2015]. They suggest that a grey (i.e. average output luminance) blank screen should be used to limit 'extraneous factors'. However, their result suggests that for stimuli duration over 24ms there is no statistical difference between a black and a blank screen. A black background was chosen for the *HDRvsLDR* experiment considering the need to limit the HDR display thermal drift [Lasance and Poppe, 2013] and therefore black is the 'colour' (i.e.

[1] As previously illustrated in Chapter 2, LDR images are characterised by an 8 bit per colour channel encoding with screen referred luminance.

the backlight is switched off) that minimises this effect. Moreover, the inter-stimulus duration was set to 1500ms, which is well above the 24ms identified by Freeman *et al.* [2015].

6.4 Participants

The study involved 40 participants (15 males and 25 females) aged between 18 and 38 (mean age = 22.7, $SD = 4.8$) recruited amongst the University of Warwick students and employees through the *University of Warwick Sona System* [SONA]. The participants' ethnicity varied[2]: nine *Other-White background*, eight *British-White*, seven *Indian*, six *Chinese*, five *Other-Asian background*, two *Pakistani*. All had normal or corrected-to-normal vision and no colour deficiency or colour blindness. They were selected so that the people portrayed in the stimuli set were unknown to the participants. Participants were asked to complete a demographic questionnaire and also to answer the PI20 test by Shah *et al.* [2015] (see Section 3.6 for further details). The answers to the latter were used to identify outliers (i.e. people with face recognition impairment). In particular, one of the participants scored a value above threshold (i.e. over 65), and therefore removed from the analysis.

6.5 Procedure

Each participant was tested individually. They were brought into a dedicated room within the International Digital Laboratory at the University of Warwick. The room has dark walls to exclude any interference from external light sources. After being briefed on the nature and procedure of the experiment, the participants were asked to sign a consent form and fill in a demographic questionnaire.

During the experiment, the lights were switched off, so that the participant could focus on the screen. The participants were asked to indicate whether the two images side-by-side portrayed the same person or not by pressing one of two available keys on a keyboard (see Figure 6.1). They were invited to

[2]The participant's ethnicity was classified in accordance to what recommended by the UK Government website [UK Ethinicity, 2011].

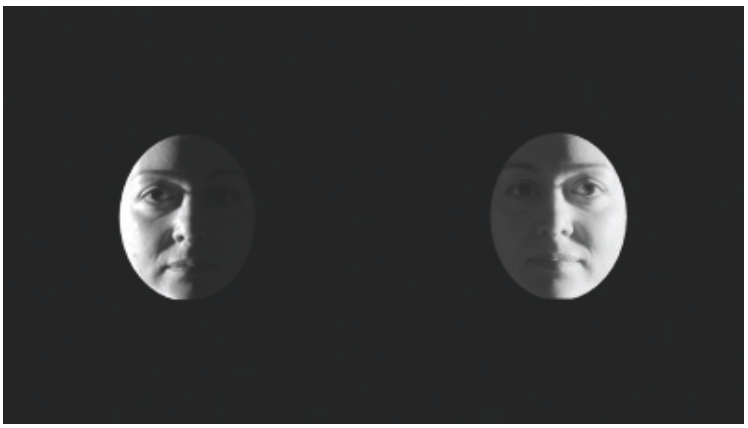


Figure 6.1: Example of the stimuli proposed in the *HDR vs LDR* experiment. Stimulus: *different* lighting, LDR vs HDR (mix). The HDR (right) has been tone-mapped in order to be visualised on this thesis.

perform the task as accurately as possible and to be as quick as they could. The stimulus was visible until the participant made a choice.

The participants were given a few trial images - the faces shown were not included in the stimuli test. This allowed them to experience the stimuli before their timing was recorded and ensured that their eyes were properly adapted to the room luminance levels. Halfway through the test participants were allowed to take a short break.

6.6 Materials

The HDR face dataset described in Chapter 5 was used as stimuli for this experiment. All images were displayed at a full-HD resolution (1080×1920 pixels) on an HDR SIM2 47 inch display. This device has a peak luminance of 4000 cd/m^2, a black level of 0.005 cd/m^2 and 12-bit colour depth. The screen output luminance was calibrated so the output luminance was linearly related to luminance values recorded in the HDR files and therefore reflected the actual luminance perceivable at the moment of the data capture.

The visualisation of the images and recording of the response was done through an *ad hoc* win32 application. The program was written in C++ using the OpenGL library in order to display the images in the native format required by the SIM2 display. The input was recorded through a keyboard connected to

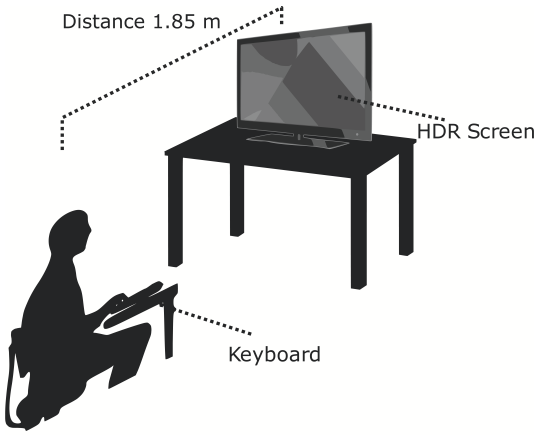


Figure 6.2: Schematic representation of the *HDRvsLDR* experiment setup.

a USB3 port acquiring high-resolution timestamps using the native Windows API QueryPerformanceCounter. The choice of this type of input device is considered acceptable according to Damian [2010]. The machine used for the experiment was equipped with an Intel Xeon E5-2620 @ 2.00GHz CPU and an Nvidia GeForce GTX 750 graphics card. See Figure 6.2 for an illustration of the setup.

6.7 Results for the HDR vs LDR Experiment

Descriptive statistics corresponding to IVs *DR* and *position* for the two DVs of *time* and *accuracy* are reported as the time in seconds taken for a correct answer, and the percentage of correct choices. Table 6.1 illustrates mean (μ) and SD (σ) of *accuracy* and *time*. Each of the first three column shows the results for a specific level of *DR* (respectively *HDR, LDR, mix*). The first and second row contain the results of each level of *position*: *same* and *different*, and the third row shows the average across *position*. The fourth column of the table, named *position*, reports values of *accuracy* and *time* for *same* and *different* lighting averaged across *DR*.

As can be seen in Table 6.1 for each case *HDR* has more correct answers and is faster than the other two conditions of *LDR* and *mix*. In general, it is faster and more accurate when judging *same* rather than *different* (see also Figure 6.3 for the line plots). In the following, the results of statistical tests on

Table 6.1: Correct Answers and Reaction Time Descriptive Statistics

accuracy (%)								
Lighting	*HDR*		*LDR*		*mix*		*position*	
	μ	σ	μ	σ	μ	σ	μ	σ
same	95.28	9.34	88.32	14.14	83.36	20.71	88.99	9.38
different	84.88	9.03	78.44	12.88	80.95	14.12	81.42	8.72
average	90.08	6.17	83.38	9.52	82.16	12.95		
time (s)								
Lighting	*HDR*		*LDR*		*mix*		*position*	
	μ	σ	μ	σ	μ	σ	μ	σ
same	2.76	0.63	3.10	1.06	2.91	0.87	2.93	0.74
different	3.35	1.19	3.52	1.10	3.47	1.06	3.45	0.17
average	3.06	0.85	3.31	1.02	3.19	0.90		

the data are reported.

6.7.1 Multivariate Analysis

A repeated measures 3 (*DR*) $\times$ 2 (*position*) MANOVA was conducted for the DVs of both *time* and *accuracy*. In the overall using Pillai's trace there was a significant effect of *position* on *time* and *accuracy* V = 0.60, F(2, 37) = 28.19, $p < 0.01$. Furthermore, there was a significant effect of *DR* on *time* and *accuracy* using Pillai's trace, V = 0.47, F(4, 35) = 7.82, $p < 0.01$. No interaction effect of *position* $\times$ *DR* was observed, V = 0.125, F(4, 35) = 1.25, $p = 0.308$. Results indicated that *position* and *DR* have an independent effect on the timing and hence we can accept H_1^1.

Due to the overall significance of the analysis, in order to further analyse the results univariate ANOVAs of 3 (*DR*) $\times$ 2 (*position*) were run for both *time* and *accuracy*.

6.7.2 Univariate Analysis: *accuracy*

Univariate analysis on the DV of *accuracy* was significant for the main effect of *DR* with Greenhouse-Geiser corrections (Mauchly's Test of Sphericity was violated, $p < 0.05$) F(1.63, 61.87) = 8.52, $p < 0.01$. Pairwise comparisons with Bonferroni corrections for *DR* showed a significant difference between *HDR* (μ = 90.08) and both *L*DR (μ = 83.38) and *m*ix (μ = 82.16), while no significant

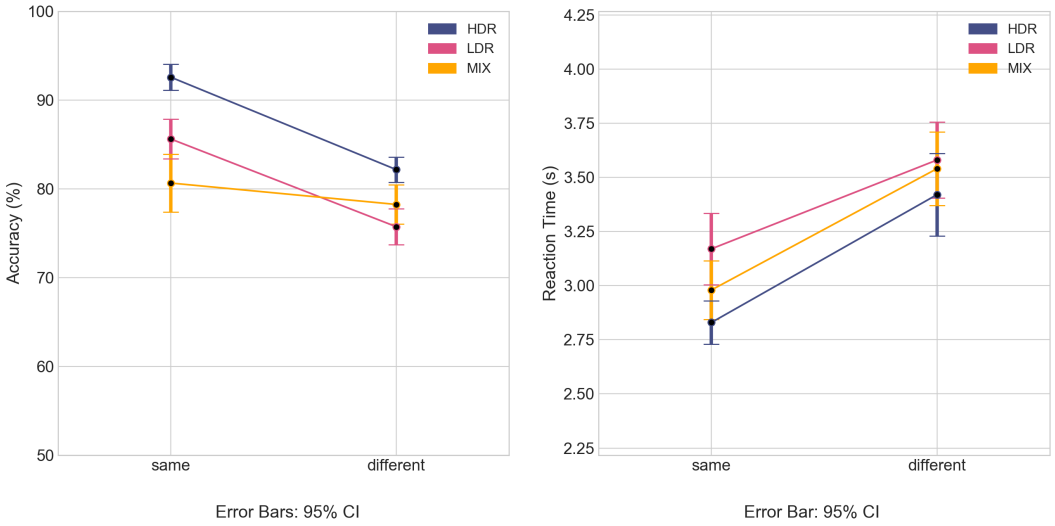


Figure 6.3: The line graph illustrates the results of the subjective experiment *HDRvsLDR* for accuracy (left) and reaction time (right) for each lighting condition (i.e. *same* and *different*). Overall participants show better performance when using *HDR* stimuli rather than *LDR* and *mix* ones, with higher accuracy and lower reaction time. Also, these results, in line with the literature, confirm the detrimental effect on performance due to difference in lighting. Note: Considering these are within-participants variables, error bars were adjusted using Cousineau's method [2005].

difference was found between *LDR* and *mix*. These results demonstrate that *DR* has a significant effect on the accuracy of perceptual matching of unfamiliar faces with HDR images improving matching rates significantly, hence we can accept H_1^3.

The main effect of *position* was also significant for *accuracy*, $\mathrm{F}(1, 38) = 14.41$, $p < 0.05$. Pairwise comparisons showed a significant difference between *same* ($\mu = 88.99$) and *different* ($\mu = 81.42$). These results show that facial matching is affected by the position of the lighting with participants lit with the same lighting on different occasions more easily recognisable than under different lighting conditions.

The interaction of *DR* $\times$ *position* was not found to be significant for *accuracy* with Greenhouse-Geiser corrections. (Mauchly's test of Sphericity violated, $p < 0.05$) $\mathrm{F}(1.67, 74.80) = 2.31$, $p = 0.11$.

6.7.3 Univariate Analysis: *time*

The main effect of *DR* was found to be significant F(2, 76) = 3.71, $p < 0.05$, when analysing the DV of *time*. Pairwise comparisons, with Bonferroni corrections for *DR* ($\alpha_{adjusted} = 0.05/2$), showed a significant difference between *HDR* ($\mu = 3.06$) and *LDR* ($\mu = 3.31$) but not significantly different to *mix*. *LDR* and *mix* were not found to be significantly different. This indicates that dynamic range of the content significantly affects how quickly a face is matched.

The main effect of *position* on *time* was also found to be significant F(1, 38) = 33.31, $p < 0.01$. Pairwise comparisons also showed a significant difference of *same* ($\mu = 2.91$) with *different* ($\mu = 3.47$). Again these results show it is quicker to recognise faces when the lighting is similar than under different lighting conditions.

The interaction of *DR* × *position* was not found to be significant for *time*, F(2, 76) = 0.86, $p = 0.43$. Due to the results presented in this subsection we can also accept H_1^2.

6.7.4 Multivariate Analysis: *position: same*

Due to the significance of the univariate ANOVAs for both *accuracy* and *time*, a further in-depth analysis was conducted for both DVs for the IV of *position* for *same* in this sub-section and for *difference* in the following sub-section.

A repeated measures MANOVA for *position:same* using Pillai's trace showed a significant effect of *DR* on *time* and *accuracy* V = 0.34, F(4, 35) = 4.49, $p < 0.01$.

The univariate results for *accuracy* show a main effect of *DR* with Greenhouse-Geiser corrections (Mauchly's test of Sphericity significant, $p < 0.05$) F(1.69, 64.24) = 6.18, $p < 0.01$. Pairwise comparisons with Bonferroni corrections ($\alpha_{adjusted} = 0.05/2$) show a significant difference between *HDR* ($\mu = 95.28$) and *LDR* ($\mu = 88.32$) and also between *HDR* and *mix* ($\mu = 83.36$). No significant difference was observed between *LDR* and *mix*.

For *time* univariate results showed a main effect of *DR* F(2, 76) = 3.56, $p < 0.05$. Pairwise comparisons, with Bonferroni corrections ($\alpha_{adjusted} = 0.05/2$), showed significant differences between *HDR* ($\mu = 2.76$) and *LDR* ($\mu = 3.1$) but no further significant differences with these two conditions and *mix*

($\mu = 2.91$).

These results show that HDR images elicit quicker and more correct responses for unfamiliar face matching for faces lit under the same lighting positions.

6.7.5 Multivariate Analysis: *position: different*

A repeated measures MANOVA for *position:different* using Pillai's trace showed a significant effect of *DR* on *time* and *accuracy* V = 0.32, F(4, 35) = 4.17, $p < 0.01$.

Univariate results for *accuracy* showed a main effect of *DR* using Greenhouse-Geiser corrections (Mauchly's test significant $p < 0.05$), F(1.65, 62.54) = 3.75, $p < 0.05$. Pairwise comparisons, with Bonferroni corrections ($\alpha_{adjusted} = 0.05/2$), for *accuracy* showed a significant difference between *HDR* ($\mu = 84.88$) and *LDR* ($\mu = 78.44$) but no significant difference with *mix* ($\mu = 80.95$).

For *time* univariate results showed no main effect for *DR* F(2, 76) = 1.19. Pairwise comparisons were not conducted due to lack of significance in the main effect but these results appear to indicate that *HDR* ($\mu = 3.35$) may be marginally faster than both *LDR* ($\mu = 3.52$) and *mix* ($\mu = 3.47$).

Results indicate a higher accuracy for HDR images over the other modalities for differing lighting conditions, however, no significant difference in response time was found.

6.8 Discussion

The work presented in this chapter aimed to establish whether HDR imaging is an effective and reliable technology in the context of surveillance and identity verification and forensic imaging.

The HDR imaging pipeline, due to image representation and visualisation closer to what humans perceive in reality, poses itself as a valid alternative to the current technology employed in every context where accuracy is essential in the decision making process. Applications can range from passport checks to witness testimony.

The HDR vs LDR experiment demonstrated added value of HDR compared to LDR imaging. The analysis of reaction time and accuracy of the three different blocks showed a clear advantage in using HDR technology, leading us to accept H_1^1. This indicates that the dynamic range of the presented content significantly affects how accurately and quickly a face is matched. Specifically, the results show that HDR improves both reaction time and accuracy, hence we can accept both H_1^2 and H_1^3.

Moreover, the results align with the literature which examines how lighting interferes with the facial recognition process, especially for unfamiliar faces. In particular when lighting conditions differ in the two presented stimuli, performance and accuracy are worse than when the lighting is the same. Nevertheless, HDR is shown to elicit more accurate response than the other two modalities even when performing a face matching task with different lighting conditions.

HDR was compared side-by-side with LDR face representation for symmetry reasons (labelled as *mix*). The results show that the recognition in most cases is not significantly quicker than with LDR. One possible explanation is that the lack of perceptually relevant information in the LDR image slows down or hinders the recognition process. This does not happen when HDR images are compared side-by-side: they result in superior performance.

The limitations of this work are related to the size of the stimuli dataset, although every precaution has been taken to avoid image memory and repetition of the same face so as to impede the observers from learning them. Questions about the ecological validity and generalizability of the experimental results, due to the nature of the stimuli can arise.

If these preliminary results are to be confirmed a more varied face stimuli set is necessary. In addition, to fully explore the impact of lighting, uncontrolled lighting (e.g. outdoor, harsh shadows, night/dim lighting) should be considered. The merit of this work lies in showing scope for the applicability of HDR imaging in face matching tasks.

6.9 Summary

This chapter has described the design, procedures and results obtained when performing the *HDRvsLDR* experiment. The statistical analysis has highlighted the potential of HDR imaging in improving reaction time and accuracy for unfamiliar face matching task. Given the granularity with which HDR can be adopted, the next chapter will present a new subjective experiment. In this experiment the effects of the integration of HDR elements in the LDR pipeline while performing an unfamiliar face matching task are tested.

Chapter 7

Backwards-Compatible HDR Imaging for Face Matching

The experiment presented in the previous chapter compared the full-LDR and HDR pipelines. The results showed that adopting the HDR pipeline is preferable to an LDR one when performing a face matching task. This chapter aims to answer the question of whether only a partial upgrade of the LDR pipeline would provide satisfactory results. The motivation for the design of a new experiment, the apparatus used, the participant recruitment process and the results obtained, are presented.

7.1 Motivation

The *HDRvsLDR* experiment described in Chapter 6 demonstrated potential for the use of HDR imaging to improve reaction times and accuracy of the unfamiliar face matching task. Although HDR technology is relatively mature, a holistic and immediate update of the LDR pipeline is most probably not realistic, for example, in the uncountable number of security cameras already deployed in a surveillance scenario. A transition phase will be necessary. The question asked with this experiment is whether it is possible to upgrade the LDR pipeline to HDR only partially, and, if so, which part between the capture and the visualisation stage would be more effective if upgraded (see Figure 7.1).

If HDR imaging is used in the capture stage, but the display stage of the

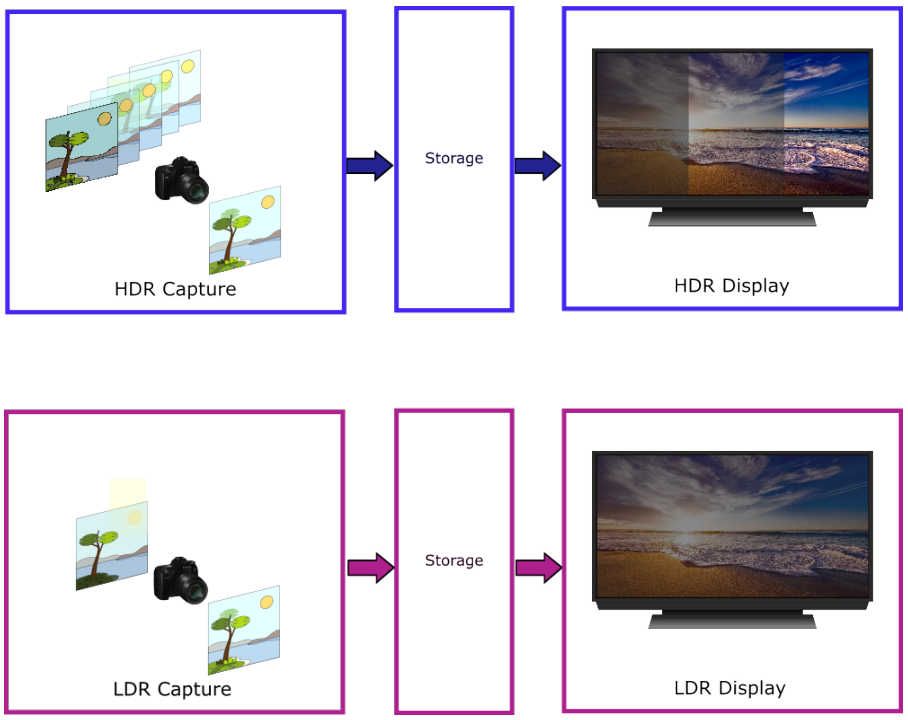


Figure 7.1: Main stages of the HDR and LDR pipelines.

pipeline remains LDR it is necessary to use an algorithm that visualises native HDR images on LDR displays. As illustrated in Chapter 2, this is a problem that has a number of solutions proposed in the literature termed TMOs. The choice of a specific TMO will be discussed in Section 7.4.1.

The symmetrical problem concerns the final stage of the pipeline, assuming only the display part has been upgraded to be HDR, it would be desirable to adapt the visualisation onto such a device of face stimuli previously captured as LDR. The technique used to perform this operation is an Expansion Operator (EO) also called Inverse TMO (introduced in Section 2.10). The choice of the specific EO for the experiment presented in this chapter is described in Section 7.4.2.

This experiment aims at verifying whether it is possible to adopt advanced imaging techniques to improve the pre-existing imaging pipeline (see Figure 7.1). The upgrade of the LDR pipeline can be performed at the capture stage or at the display stage, see Figure 7.2. This would allow:

- to upgrade LDR face datasets through the adoption of EOs for them to

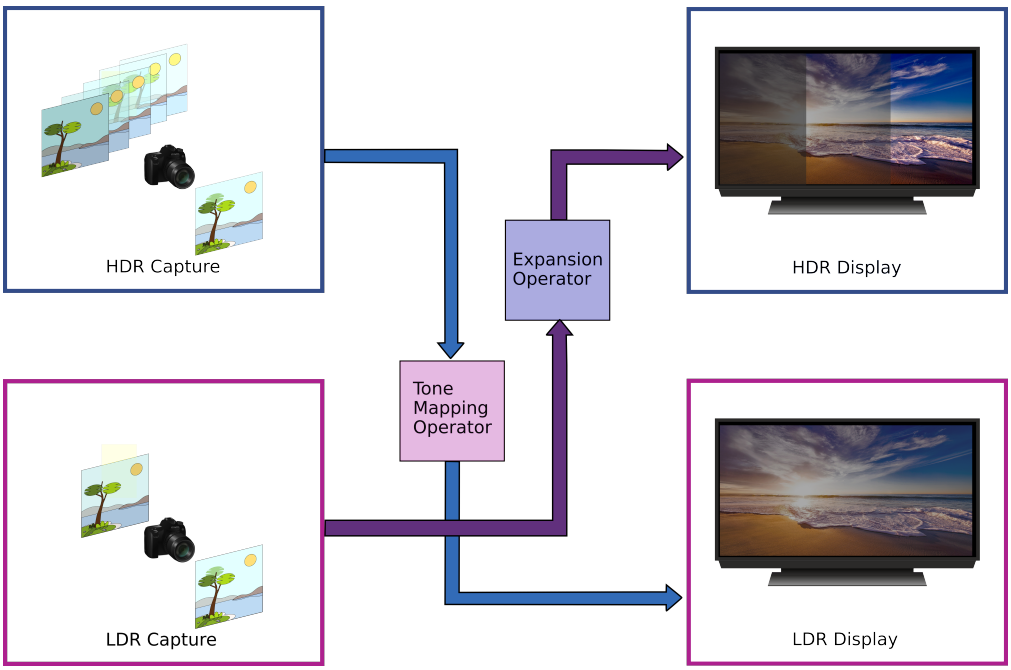


Figure 7.2: This figure represents the mixed pipelines compared in the *Backwards-Compatibility HDR* experiment. In one case an HDR captured stimulus is tone-mapped to be visualised on a traditional LDR display, while, in the other case, an LDR captured stimulus is retargeted through an EO to expand its dynamic range and it is visualised on an HDR screen.

be visualised on an HDR screen;

- to tone-map newly captured HDR imaging dataset so that they can easily be visualised on LDR screens, preserving at least part of the additional information available in the HDR image.

The results of this experiment will indicate whether some form of HDR capture or display is preferable. The overall design of this experiment is similar to the experiment presented in Chapter 6, details follow in the next section.

7.2 Design

For this experiment, named *Backwards-Compatibility HDR*), a two forced choice design was used. Participants were presented with two concurrent faces and were asked to choose whether the faces portrayed were the same person or

not. The test measured reaction time and accuracy while dynamic range and lighting condition of the stimuli were modified.

The *Backwards-Compatibility HDR* experiment used two independent variables (IVs): Dynamic Range (termed *DR*) and Lighting (termed *position*).

For the within-participants independent variable dynamic range (*DR*) two scenarios were tested:

- EO_LDR (*LDR*): LDR images enhanced through an Expansion Operator;
- TMO_HDR (*HDR*): HDR images tone-mapped to become LDR images.

Moreover, in order to test the impact of lighting disparity on the performance in the face matching task, a second within-participants independent variable (*position*) was evaluated:

- Same lighting (*same*): same light direction for both the faces portrayed;
- Different lighting (*different*): the two faces are lit with different intensity and direction light sources.

The dependent variables (DV) were reaction time (termed *time*) and accuracy (termed *accuracy*). *accuracy* is reported as percentage of correct answers and *time* is reported in seconds.

There were three general hypotheses:

- H_1^1: LDR image derived from a tone-mapped HDR image (TMO_HDR) will outperform the other modality (i.e. EO_LDR) as the original image captures more perceptually relevant details and reproduces more accurately real-world lighting;
- H_1^2: reaction time (*time*) is less affected by variation of light position when the original stimuli are HDR tone-mapped (TMO_HDR) as opposed to native LDR stimuli expanded to HDR (EO_LDR);
- H_1^3: accuracy (*accuracy*) is less affected by variation of light position when the original stimuli are HDR tone-mapped (TMO_HDR) as opposed to native LDR stimuli expanded to HDR (EO_LDR).

Each participant was presented with a total of 60 trials, each trial contained a permutation of the two *DR* scenarios. Each block contained thirty images presenting two concurrent faces under *same* or *different* lighting equally split between same-person pairs and different-person pairs. The use of the two keys to indicate a match/mismatch were assigned randomly to participants and counterbalanced across the experiment.

To avoid the familiarity effect, particular care was taken so that a specific face was not shown more than 10 times [Longmore *et al.*, 2008] throughout the trials. Also, to avoid picture comparison, even when the *same* lighting same-person pair was presented the two different image captures available in the dataset were used (for details on the dataset see Chapter 5).

The faces portrayed on the screen occupied 7 degrees of vertical visual angle, which is within the intervals adopted in the literature for experiments on face perception (see [Peterson and Eckstein, 2012; Ghuman *et al.*, 2014; Goffaux and Greenwood, 2016]).

Following the literature [Burton *et al.*, 2010; Megreya *et al.*, 2011; Bobak *et al.*, 2016], only the luminance channel was used for the images to limit additional variability given by the chroma component and also in view of the fact that colour components seem to be less relevant in the recognition process [Hancock *et al.*, 2000; Longmore *et al.*, 2008]. The faces depicted were cut out with an elliptic shape similarly to the approach taken by Megreya *et al.* [2011], so that only the internal features were visible, similarly to the approach adopted for the *HDRvsLDR* experiment. Care was taken to ensure the elliptic shape had the same identical dimension for every face portrayed in order to avoid participants to subconsciously identify faces based on the external shape (i.e. picture comparison) and not by looking at the facial features [Young and Burton, 2017]. Lighting in the room was controlled and enough time was allowed to ensured that the participant's eyes were properly adapted to the room luminance levels. No feedback was provided to the participants regarding their accuracy while performing the task.

Distance of each participant from the screen was controlled (185 cm) in accordance with the Recommendation ITU-R BT.2022 [2012]. In between stimuli, a black screen with a central white fixation cross was alternated for 1500 ms. This is standard practice in perceptual experiment: the use of visual

masks to manipulate the time course of processing allows the participant retinal image to be refreshed (i.e. backwards masking [Cowan, 2008]). The colour black was chosen for the background of the inter-stimuli fixation cross (the justification for this choice is given in Section 6.3).

7.3 Participants

The study involved 42 participants aged between 18 and 37 (mean age = 22.5 *SD* =4.7), 14 males and 28 females, recruited amongst the University of Warwick students and staff through the *University of Warwick Sona System* [SONA]. The participants' ethnicity varied[1]: eleven *English/Welsh/Scottish/Northern Ireland*, eleven *Any Other-White Background*, seven *Indian*, three *Chinese*, two *Arab Background*, two *Any Other/Mixed Ethnic Group*, two *Any Other-Asian Background*, one *White and Asian*, one *White and Black African*, one *White and Black Caribbean*, one *Any Other Ethnic Group*.

All participants were tested ensuring they had no colour deficiency and normal or corrected-to-normal vision. They were selected so that they had not participated in the *HDRvsLDR* experiment (Chapter 6) and also the faces portrayed in the stimuli set were unknown to them. Moreover, all the participants were asked to fill in a demographic questionnaire and to complete the PI20 questionnaire [Shah *et al.*, 2015], introduced in Section 3.6. The PI20 results allowed to identify outliers (i.e. people with face recognition impairment). In particular, three participants were considered outliers and so their results were removed from the analysis.

7.4 Apparatus and Materials

The apparatus and materials used for this experiment were similar to the ones used for the experiment *HDRvsLDR* described in Chapter 6. The stimuli dataset used was the Warwick HDR Face Dataset (Chapter 5), and, similarly to what had been done in the *HDRvsLDR* experiment, the faces depicted were cut out with an elliptic shape, so that only the internal features were visible.

[1]The participant's ethnicity was classified in accordance to what recommended by the UK Government website [UK Ethinicity, 2011].

EO_LDR Creation: The HDR stimuli were generated in two steps using the images in the Warwick HDR Dataset. The details related to the choice of TMO are explained in Section 7.4.1. The guiding principle was, in the first step, to emulate a native LDR capture through the application of the Optimal Exposure algorithm [Debattista *et al.*, 2015](as done in the experiment *HDRvsLDR*). In the second step, the EO by Marnerides *et al.* [2018] was applied in order to reconstruct an HDR stimulus, and expand the LDR original dynamic range.

As presented before (see 2.9), Optimal Exposure emulates the behaviour of the camera software at the moment of the capture, selecting the single exposure that preserves maximum information by analysing the image's histogram [Debattista *et al.*, 2015]. The ExpandNet, described in Section 2.10.3, is a machine learning technique that allows the expansion of generic LDR image by guessing their full dynamic range. The method returns normalised luminance values $L(x, y) \in [0, 1]$. In order to preserve the accurate luminance reproduction and coherence with the previous experiment, every image was rescaled to its original luminance range. The implementation of the ExpandNet algorithm was the one made available on the author's online repository[2].

TMO_HDR Stimuli Creation: The algorithm chosen for the creation of the LDR stimuli was the TMO proposed by Reinhard *et al.* [2002] as described in the introduction of this chapter. For this experiment the implementation of the algorithm available in the MATLAB HDR Toolbox [Banterle *et al.*, 2017] was used.

This algorithm offers a *global* and a *local* version. The local version was applied using the default parameters indicated by the authors in the original paper: $\phi = 8.0$ defines the neighbourhood size and α defines the key of the image and is evaluated as a factor of the image luminance with values between 0.18 and 0.72. The images' maximum luminance range was rescaled at the average LDR screen luminance (i.e. 250 cd/m^2) in order to emulate a common visualisation scenario.

Setup All images were displayed at full-HD resolution (1080×1920 pixels) on an HDR SIM2 47 inch display. This device has a peak luminance of 4000

[2]Repository available at the following url: `https://github.com/dmarnerides`

cd/m^2, a black level of 0.005 cd/m^2 and 12-bit colour depth. The screen output luminance was calibrated so the output luminance was linearly related to luminance values recorded in the HDR files and therefore reflected the actual luminance perceivable at the moment of data capture. The participants were seated at 185 cm from the HDR display.

The visualisation of the images and recording of the response was done through an *ad hoc* win32 application. The program was written in C++ using the OpenGL library in order to display the images in the native format required by the SIM2 display. The input was recorded through a keyboard connected to a USB3 port acquiring high-resolution timestamps using the native Windows API QueryPerformanceCounter. The choice of this type of input device is considered acceptable according to Damian [2010]. The machine used for the experiment was equipped with an Intel Xeon E5-2620 @ 2.00GHz CPU and an Nvidia GeForce GTX 750 graphics card.

7.4.1 Tone Mapping Operator Selection

As already described in Section 2.9, many perceptual experiments have tried to establish the best TMO. The majority of the literature engaging with comparisons amongst different TMOs has established that the choice of TMO is application oriented and it depends on specific characteristics of the images [Čadík *et al.*, 2008; Parraga *et al.*, 2018]. Therefore, considering the literature and the specificity of faces as object, an evaluation study was conducted in order to establish which TMO to select in the current experiment.

Design: A quantitative analysis study was chosen, where six TMOs were evaluated using two metrics: TMQI and FSTIM. These two metrics were introduced introduced in Section 2.11. They are based on perceptual studies results and have been extensively used in the literature (see for example Khan *et al.* [2017]; Krasula *et al.* [2017]; Debattista [2018]) as methods to effectively assess and quickly compare TMOs output quality without the need for a subjective experiment. The question of interest is:

> Q_1*:* Which is the most suitable TMO for reproducing HDR captured faces on LDR display with the highest fidelity?

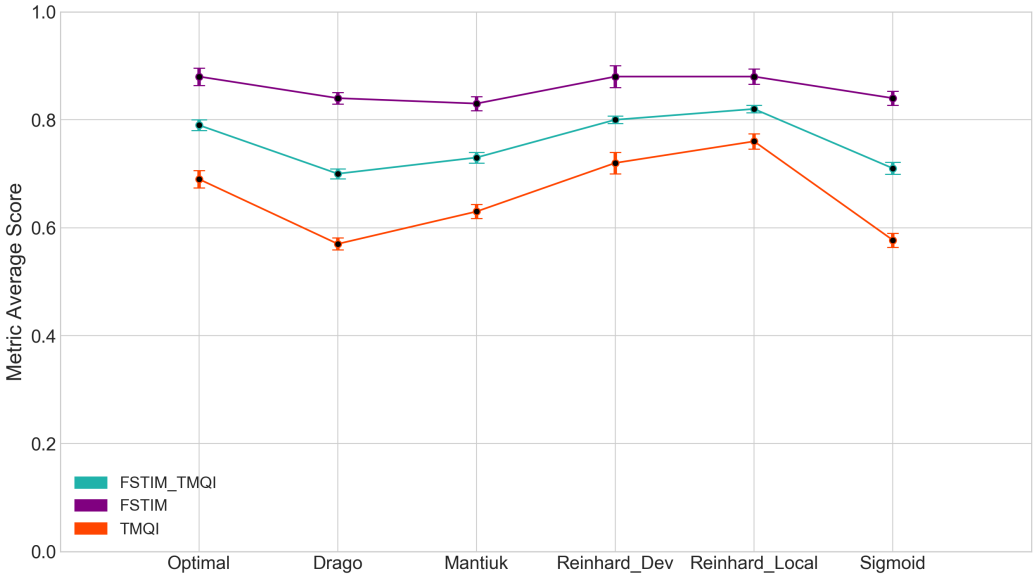


Figure 7.3: The line graph illustrates the average scores for the three metrics, TMQI, FSTIM and FSTIM&TMQI. *Reinhard_Local* is the TMO achieving the highest scores for all the metrics.

TMO Selection: Six TMOs were evaluated including a sigmoid function (called *Sigmoid*), the Optimal Exposure operator [Debattista *et al.*, 2015] (called *Optimal*), and four more operators selected among those that previous experiments had shown to achieve good performance in terms of naturalness and fidelity:

- Drago *et al.* [2003] (referred as *Drago*);
- Reinhard *et al.* [2002] (referred as *Reinhard_Local*) ;
- Reinhard and Devlin [2005] (referred as *Reinhard_Dev*) ;
- Mantiuk *et al.* [2006b] (referred as *Mantiuk*) .

Default parameters were selected as advised by the authors if required.

Method: All the images in the Warwick HDR Face Dataset were tone-mapped with the six chosen TMOs and their individual TMQI [Yeganeh and Wang, 2012] and FSTIM [Nafchi *et al.*, 2014] values were calculated. These

metrics accept as input the original HDR image and its tone-mapped version. Both metrics provide an output value between 0 and 1. The authors of FSTIM also suggest that these two metrics can be used in conjunction, to complement each other and mediate each other similarity estimation mistakes [Nafchi *et al.*, 2014]. FSITM&TMQI is calculated as the average of FSTIM and TMQI. Krasula *et al.* [2017] have shown that FSITM and FSITM&TMQI are the most reliable metrics for generic images (i.e. not faces) and successfully predict which images are perceived to be of better quality by users.

Results: In the case of TMQI a one-way repeated measures ANOVA (see Table 7.1 for mean (μ)and SD (σ) values), shows a significant difference amongst the different TMOs. The value of this test, with Greenhouse-Geisser correction (Mauchly's sphericity assumption was violated, $p<0.01$) is F(2.02,360.27)=534.38, $p<0.01$. Post hoc tests using Bonferroni correction revealed that *Reinhard_Local* scores significantly higher than others. In particular its difference from *Reinhard_Devlin* - the other TMO with the second highest score - is significant, t(170)=-8.54, $p<0.01$.

Table 7.1: TMO comparison using the TMQI, the FSTIM and the FSTIM&TMQI metrics. Mean scores across all Warwick HDR Face dataset[1].

TMO	TMQI		FSTIM		FSTIM&TMQI	
	μ	σ	μ	σ	μ	σ
Optimal	0.67	0.16	0.882	0.04	0.79	0.07
Drago	0.55	0.13	0.842	0.04	0.70	0.06
Mantiuk	0.61	0.14	0.833	0.06	0.73	0.07
Reinhard_Dev	0.70	0.15	0.877	0.04	0.80	0.05
Reinhard_Local	0.74	0.16	0.884	0.04	0.82	0.05
Sigmoid	0.56	0.14	0.842	0.06	0.71	0.07

[1]For FSTIM, given the closeness of the scores, three significant digits have been provided.

For FSTIM average scores a one-way repeated measures ANOVA shows a significant difference amongst the different TMOs. The value of this test with Greenhouse-Geisser correction - due to Mauchly's sphericity violation ($p<0.01$)- is F(3.11,554.08)=111.29, $p<0.01$. Post hoc tests using Bonferroni correction revealed that *Reinhard_Local* scores significantly higher ($\mu = 0.88$) than others, except for *Optimal* and *Reinhard_Devlin*. However, there is no statistical

difference between *Reinhard_Local* and *Optimal* t(170)=-0.77, p=0.50. While, on the other hand, there is a statistical difference between *Reinhard_Local* and *Reinhard_Devlin*, t(170)=-3.80, p<0.01.

For the FSTIM&TMQI metric average scores a one-way repeated measures ANOVA shows a significant difference amongst the different TMOs. The value of this test, with Greenhouse-Geisser correction - due to Mauchly's sphericity violation (p<0.01) - is F(2.55,454.02)=462.64, p<0.01. Post hoc tests using Bonferroni correction revealed that *Reinhard_Local* scores significantly higher than others ($\mu = 0.82$). In particular its difference from *Reinhard_Devlin* ($\mu = 0.80$)- the other TMO with the second highest score - is significant, t(170)=-13.71,p<0.01.

In view of the results presented here Reinhard's local TMO [Reinhard *et al.*, 2002] was chosen to process the Warwick HDR Face Dataset creating the tone-mapped LDR stimuli for the *Backwards-Compatibility HDR* experiment.

7.4.2 Expansion Operator Selection

The idea behind the adoption of such manipulation is to test whether it is possible to use advanced imaging techniques to enable LDR datasets available in the literature to be displayed effectively on HDR displays.

The literature offers a number of EOs, however none of them have been applied in the context of facial stimuli, therefore the only option is to search amongst those created for generic LDR scenes. Some of the algorithms proposed in the literature require user input, however, this is not desirable for the purpose of this experiment as the end aim is to have an automated pipeline. In recent years some algorithms which automate the process using CNNs and offer visually pleasing results have been proposed. Amongst these the choice for this experiment was directed towards ExpandNet, proposed by Marnerides *et al.* [2018] (see Section 2.10.3). The reasons behind this choice are the following:

- in the original version it outperformed all other methods - including non deep learning ones;
- it is first algorithm to adopt a dedicated CNN solution for end-to-end inverse tone-mapping;

- its code is open source and freely available;

- it is an algorithm that has been used several times in many publications as a benchmark for new expansion operators (e.g. [Wang *et al.*, 2019; Fan *et al.*, 2019]).

7.5 Procedure

Each participant was tested individually. They were brought into a dedicated room within the International Digital Laboratory at the University of Warwick. The room has dark walls to exclude any interference from external light sources. After being briefed on the nature and procedure of the experiment, the participants were asked to sign a consent form and fill in a demographic questionnaire.

During the experiment, the lights were switched off, so that the participant could focus on the screen. The participants were asked to indicate whether the two images shown side-by-side portrayed the same person or not by pressing one of two available keys on a keyboard (see Figure 6.1). The participants were invited to perform the task as accurately as possible and to be as quick as they could. The stimulus was visible until the participant made a choice. A black screen with a central white fixation cross was alternated to the stimulus for 1500ms.

Participants carried out 5 practice trials. The practice trial faces were not seen again in the main experiment. This allowed participants to experience the stimuli before their timing was recorded and ensured that their eyes were properly adapted to the room luminance levels. No feedback was provided to the participants regarding their accuracy while performing the task. Participants were allowed to take a break halfway through the experiment. The experiment would take, on average, about 10 minutes to be completed.

Table 7.2: Correct Answers and Reaction Time Descriptive Statistics for the *Backwards-Compatibility HDR*

accuracy (%)						
position	*TMO_HDR*		*EO_LDR*		*average*	
	μ	σ	μ	σ	μ	σ
same	90.04	11.25	86.40	14.82	88.22	11.02
different	80.19	11.76	75.02	11.32	77.60	8.72
average	85.11	8.61	80.70	9.36		
time (s)						
position	*TMO_HDR*		*EO_LDR*		*average*	
	μ	σ	μ	σ	μ	σ
same	2.87	0.70	3.06	1.02	2.96	0.87
different	3.25	0.90	3.43	1.27	3.34	1.06
average	3.06	0.80	3.25	1.12		

7.6 Results: Backwards-Compatibility HDR experiment

Descriptive statistics corresponding to independent variables *DR* and *position* for the two DVs of *time* and *accuracy* reported as the time in seconds taken for a correct answer and the percentage of correct choices are reported in Table 7.2. For each case *TMO_HDR* has more correct answers and is faster than the other condition of *EO_LDR*. In general, participants were faster and more accurate when judging *same* rather than *different* (see Figure 7.4). In the following, the results of inferential statistics on the data are reported.

7.6.1 Multivariate Analysis

A repeated measures 2 (*DR*) $\times$ 2 (*position*) MANOVA was conducted for the DVs of both *time* and *accuracy*. Overall using Pillai's trace there was a significant effect of *position* on *time* and *accuracy* V = 0.56, F(2, 37) = 24.00, $p< 0.01$. Furthermore, there was a significant effect of *DR* on *time* and *accuracy*: Pillai's trace, V = 0.26, F(2, 37) = 6.50, $p< 0.01$. There was no significant interaction effect of *position* $\times$ *DR*, V = 0.005, F(2, 37) = 0.19, P = 0.91.

Due to the overall significance of the analysis, in order to further analyse

the results a univariate analysis of 2 (*DR*) × 2 (*position*) was performed for both *time* and *accuracy*.

7.6.2 Univariate Analysis: *accuracy*

Univariate repeated measure ANOVA analysis on the DV *accuracy* was significant for the main effect of *DR*: F(1, 38) = 7.77, $p< 0.01$. There was a significant main effect of *position* with F(1, 38) = 26.15, $p< 0.01$. The interaction of *DR* × *position* was found not to be significant with F(1, 38) = 0.29, $p= 0.67$. Post hoc tests using Bonferroni correction on *DR* showed a significant difference between *TMO_HDR* ($\mu = 85.11$) and *EO_LDR* ($\mu = 80.70$). This difference (4.40), CI[1.20,7.60], was significant t(39) = 2.79, $p< 0.01$. Post hoc tests using Bonferroni correction on *position* showed a significant difference between *different* ($\mu = 77.60$) and *same* ($\mu = 88.22$). This difference (-10.61), CI[-14.81,-6.41] was significant, t(39) = -5.11, $p< 0.01$. No other post hoc tests were conducted as the interaction *DR* × *position* was not significant. These results demonstrate that *DR* and *position* have a significant effect on the accuracy of perceptual matching of unfamiliar faces with HDR images improving matching rates significantly.

7.6.3 Univariate Analysis: *time*

Univariate repeated measure ANOVA analysis on the DV of *time* was significant for the main effect of *DR*: F(1, 38) = 6.23, $p< 0.05$. There was also a significant main effect of *position* with F(1, 38) = 31.23, $p< 0.01$. The interaction of *DR* × *position* was found not to be significant for *time* with F(1, 39) = 0.005, p=0.94. Post hoc tests using Bonferroni correction on *DR* showed a significant difference between *TMO_HDR* ($\mu = 3.06$) and *EO_LDR* ($\mu = 3.25$). This difference (-0.19), CI[-0.34,-0.04], was significant t(39) = -2.50, $p< 0.05$. Post hoc tests using Bonferroni correction on *position* showed a significant difference between *different* ($\mu = 3.34$) and *same* ($\mu = 2.96$). This difference (0.38), CI[0.24,0.51], was significant t(39) = 5.59, $p< 0.01$. No other post hoc tests were conducted as the interaction *DR* × *position* was not significant. These results demonstrate that *DR* and *position* have a significant effect on the reaction time of perceptual matching of unfamiliar faces with HDR images

improving matching rates significantly.

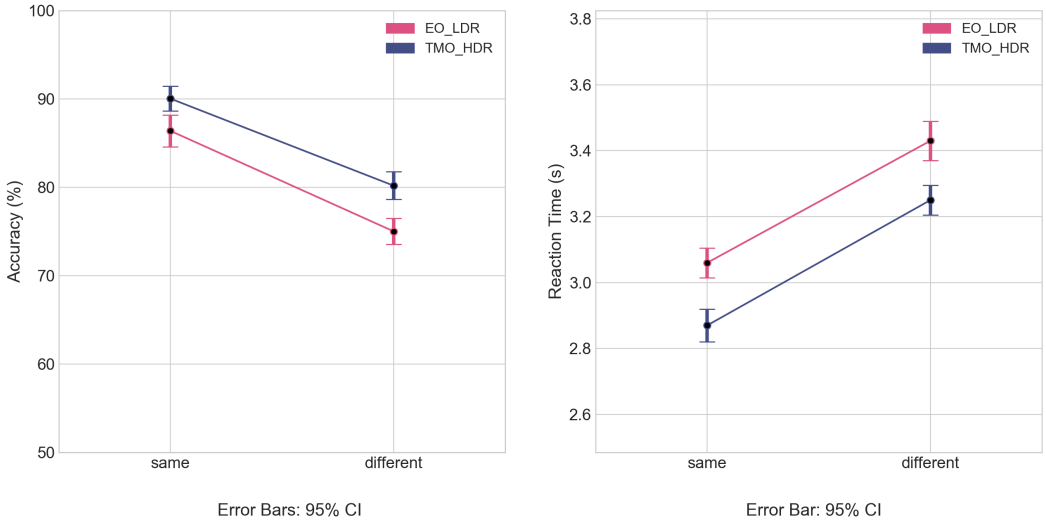


Figure 7.4: These graphs illustrate the results of the subjective experiment *Backwards-Compatibility HDR* for Accuracy (left) and Reaction Time (right). The *TMO_HDR* allows to achieve higher accuracy than the *EO_LDR* when *position:different*. The Reaction Time is better (i.e. participants were quicker) for the *TMO_HDR* stimuli, for both lighting conditions. Note: Considering these are within-participants variables, error bars were adjusted using Cousineau's method [2005].

7.6.4 Multivariate Analysis: *position: same*

Considering the significance of the univariate analysis, a further in-depth analysis was conducted for both DVs for the IV of *position* for *same* in this sub-section and for *different* in the following sub-section.

A repeated measures MANOVA for *position:same* using Pillai's trace showed a significant effect of *DR* on *time* and *accuracy*: V = 0.21, F(2, 37) = 4.95, $p < 0.05$. When *time* is considered as the sole dependent variable a significant effect of *DR* was found: t(39)= - 2.39 (one-tailed) with Bonferroni correction, $p < 0.05$. *TMO_HDR* stimuli response ($\mu = 2.87$) was significantly faster than *EO_LDR* ($\mu = 3.06$). When *accuracy* is considered as the sole

dependent variable, there was no significant effect of *DR*: t(39) = 1.58 (one-tailed), p= 0.06.

These results show that, for the unfamiliar face matching task with faces lit with the same lighting, *TMO_HDR* images elicit quicker responses than *EO_LDR*. No significant difference was found on *accuracy*.

7.6.5 Multivariate Analysis: *position: different*

The results of repeated measures MANOVA for *position:different* showed a significant main effect of *DR*, Pillai's trace: V = 0.71, F(2, 37) = 3.83, $p<$ 0.05. When *accuracy* is considered as the sole dependent variable, a significant effect was found, t(39)= 2.13 (one-tailed) with Bonferroni correction, $p<0.05$, correct answers for *TMO_HDR* stimuli (μ =80.19) where higher than *EO_LDR* (μ =75.02).

When *time* is considered as the sole dependent variable, results showed a significant effect of *DR*: t(39)= 2.05 (one-tailed) with Bonferroni correction, $p<$ 0.05, with quicker response to *TMO_HDR* (μ = 3.25) stimuli than *EO_LDR* (μ =3.43).

Results indicate a better response for *TMO_HDR* images over the other modality for differing lighting conditions, for both DV *time* and *accuracy*.

7.7 Discussion

This chapter has described a subjective experiment on the unfamiliar face matching task, where the impact of HDR and LDR stimuli were compared using a forced-choice design. The experiment aimed at identifying whether it is possible to upgrade partially the current LDR pipeline, to become HDR, in either the display or the data capture stage for the purpose of improving performance.

The results of the multivariate analysis have shown that there is a significant advantage of capturing images in HDR and subsequently displaying them on an LDR display compared to expanding the dynamic range of LDR images and displaying them on an HDR screen, even if adopting advanced machine learning techniques, therefore we can accept H_1^1.

Moreover, we can also accept H_2^1, as the results have shown a significant impact of *DR* on the DV *time*. Participants were quicker in the face matching task both for *position:same* and *position:different*. We cannot accept H_3^1, as the impact of *DR* on *accuracy* was significant only for the case of *position:different*. However, the significance of *TMO_HDR* lighting when *position:different* is not a banal result as previous experimental results - both from the literature and from the experiment described in Chapter 6 - have shown how difference in lighting conditions affects negatively face matching performance. We interpret the lack of significance for the case of *position:same* as the likelihood that task is easy enough not to be significantly affected by *DR*.

The limitations of this work are related, similarly to the experiment in Chapter 6, to the size of the Warwick HDR Dataset. This can raise questions about ecological validity and generalisability. However, this work shows scope for the application of HDR technology in specific stages of the LDR pipeline for the unfamiliar matching tasks. In particular, HDR imaging should be preferred over the ordinary LDR especially in the capture stage, as the results presented in this chapter indicate that HDR capture allows to preserve more perceptually relevant features than LDR capture.

7.8 Summary

This chapter has described the design and the results of the *Backwards-Compatibility HDR* experiment aimed to address the question of which specific element of the HDR imaging technology is more advantageous to upgrade for backwards-compatibility. The next chapter will introduce a new experiment where the results of the *Backwards-Compatibility HDR* experiment are compared with the result obtained in the *HDRvsLDR* experiment (presented in Chapter 6) in order to establish whether a partial upgrade of the pipeline would suffice to improve the unfamiliar face matching task.

Chapter 8

Pipeline Comparison

In Chapter 6 the full HDR pipeline was compared to the full LDR pipeline indicating that HDR supports superior performance in the unfamiliar face matching task. In Chapter 7 the partial upgrade of the LDR pipeline with elements taken from the HDR pipeline was explored. The upgrade of HDR imaging in the capture stage offers an advantage over the visualisation stage. This chapter is dedicated to a new analysis where the results obtained in the experiments presented in Chapter 6 and Chapter 7 are compared.

8.1 Introduction

The experiment described in Chapter 6 (i.e. the *HDRvsLDR* experiment) has established a superiority of matching performance when participants are shown stimuli using the HDR imaging pipeline, as opposed to the LDR ones. The *Backwards-Compatibility HDR* experiment (Chapter 7) results indicated that there is an advantage in capturing the facial stimuli in HDR, even if these are subsequently displayed on LDR screens, as opposed to retargeting LDR captured images to be displayed on HDR screens. HDR face stimuli seem to retain perceptual details that elicit quicker reaction times and higher accuracy - under different lighting conditions.

In view of these results, a logical question is whether a complete upgrade of the imaging pipeline to HDR is necessary or a partial upgrade would be acceptable for the purpose of unfamiliar face matching task. In this chapter the results obtained from these two previous experiments will be compared.

8.2 Design

Considering the symmetrical nature of the *HDRvsLDR* experiment and the *Backwards-Compatibility HDR* experiment in terms of number of participants (N=39) and stimuli type (HDR vs LDR) it is possible to compare the effect of HDR and LDR capture stimuli across the two experiments using a mixed design approach. Figure 8.1 illustrates the general design of this experiment. There are two within-participants factors: *DR* and *position* and one between-participants factor identifying the type of *pipeline*.

The IV *Dynamic Range* (*DR*) has two levels, as seen in previous experiments, identifying respectively the captured stimuli based on their dynamic range:

- *HDR*;
- *LDR*.

The IV *position* has two levels, as seen before, identifying the lighting conditions in the stimuli:

- *same*;
- *different*.

The IV *pipeline* has two levels, identifying the pipeline adopted respectively in the experiment *HDRvsLDR* (Chapter 6) and the experiment *Backwards-Compatibility HDR* (Chapter 7):

- *native*;
- *backwards-compatible*.

The dependent variables (DV) in the case of this analysis are the same as in both the *HDRvsLDR* and *Backwards-Compatibility HDR* experiments:

- reaction time (named *time*);
- accuracy (named *accuracy*). *accuracy* is reported as percentage of correct answers and *time* is reported is seconds.

There were four general hypothesis:

- H_1^1 : the *native:HDR* pipeline helps obtain better reaction times than the *backwards-compatible:HDR* on the unfamiliar face matching task;
- H_1^2 : the *native:HDR* pipeline helps obtain better accuracy than the *backwards-compatible:HDR* on the unfamiliar face matching task;
- H_1^3 : the *native:LDR* pipeline helps obtain better reaction times than the *backwards-compatible:LDR* on the unfamiliar face matching task;
- H_1^4 : the *native:LDR* pipeline helps obtain better accuracy than the *backwards-compatible:LDR* on the unfamiliar face matching task.

8.3 Procedure

In this experiment the data collected in previous experiments were used. Full details of the data collection procedure can be found in the respective chapters (i.e. Chapter 6 for the *HDRvsLDR* and Chapter 7 for the *Backwards-Compatibility HDR*).

Participants Details of participants are reported here for completeness. The data labelled as *pipeline:native* is the result obtained in the experiment *HDRvsLDR*. The 40 participants involved in the study (15 males and 25 females) were aged between 18 and 38 (mean age = 22.7, $SD = 4.8$) and recruited among the University of Warwick students and employees through the *University of Warwick Sona System* [SONA].

The data labelled as *pipeline:backwards-compatible* is the result obtained in the experiment *Backwards-Compatibility HDR*. The 42 participants taking part in the study (14 males and 28 females) were aged between 18 and 37 (mean age = 22.4 SD =4.7) and were recruited among the University of Warwick students and staff through the *University of Warwick Sona System* [SONA].

Both studies' participants where tested ensuring they had no colour deficiency and normal or corrected-to-normal vision, similarly. They were selected so that they had not participated in any of the previous experiments and also the faces portrayed in the stimuli set were unknown to them. In

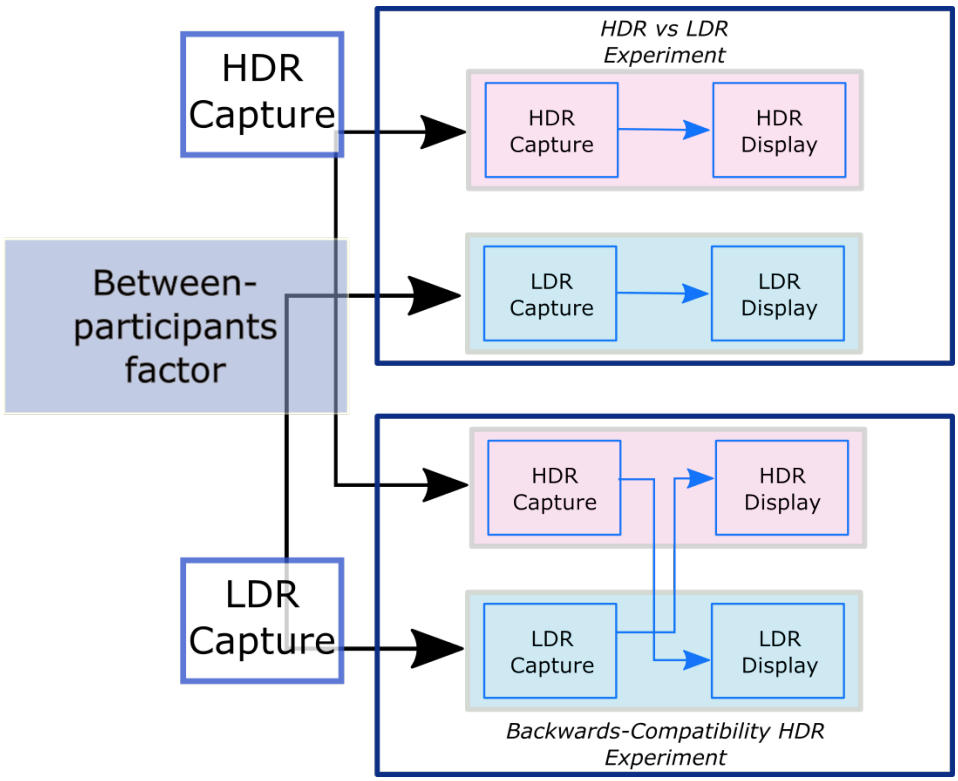


Figure 8.1: This diagram illustrates the structure of the experiment described in this chapter. The results of the experiment *HDRvsLDR* and *Backwards-compatibility HDR* are compared across to answer the question related to the use of a native vs backwards-compatible HDR pipeline.

addition, all the participants had to complete the PI20 test [Shah *et al.*, 2015] (introduced in Section 3.6), which would allow to identify outliers (i.e. people with face recognition impairment).

8.4 Results

The descriptive statistics is available in Table 8.1. The table indicates that generally *accuracy* for HDR stimuli was higher than LDR, although results are superior for the *pipeline:native* than for the *pipeline:backwards-compatible*. Similarly, in case of *time*, performance are superior (i.e. reaction time is shorter) for HDR than for LDR stimuli, however there seems to be a similarity between the results of the two *pipelines*. As for the effect of *position*, the detrimental impact of *different* lighting in comparison to *same* lighting, regardless of *DR*,

Table 8.1: This table contains the descriptive statistics for the *Pipeline Comparison* experiment. Accuracy (*accuracy*) is expressed in percentage of correct answers, while reaction time (*time*) is expressed in seconds.

accuracy(%)													
pipeline:native							*pipeline:backwards-compatibile*						
position	*HDR*		*LDR*		*average*		*position*	*HDR*		*LDR*		*average*	
	μ	σ	μ	σ	μ	σ		μ	σ	μ	σ	μ	σ
same	95.28	9.34	88.32	14.12	91.80	9.99	*same*	90.04	12.25	86.40	14.82	88.22	11.02
different	84.88	9.03	78.44	12.88	81.66	9.14	*different*	80.19	11.76	75.02	11.32	77.60	8.72
average	90.08	6.17	83.38	9.52			*average*	85.11	8.61	80.70	9.36		

time(s)													
pipeline:native							*pipeline:backwards-compatibile*						
position	*HDR*		*LDR*		*average*		*position*	*HDR*		*LDR*		*average*	
	μ	σ	μ	σ	μ	σ		μ	σ	μ	σ	μ	σ
same	2.76	0.63	3.10	1.06	2.93	0.82	*same*	2.87	0.70	3.06	1.02	2.96	0.87
different	3.35	1.19	3.52	1.10	3.44	1.08	*different*	3.25	0.90	3.43	1.27	3.34	1.06
average	3.06	0.85	3.31	1.02			*average*	3.06	0.80	3.25	1.12		

is shown in the results. In order to compare the results of the experiment *HDRvsLDR* (Chapter 6) and the experiment *Backwards-Compatibility HDR* (Chapter 7), a three-way mixed MANOVA was conducted. The results show no significant main effect of *pipeline*: Pillai's trace shows V=0.07, F(2,75)=2.89, p =0.06. This is not surprising, because it represents an average across the *DR* and *position* and simply shows that the two groups' performance were similar across experiments overall.

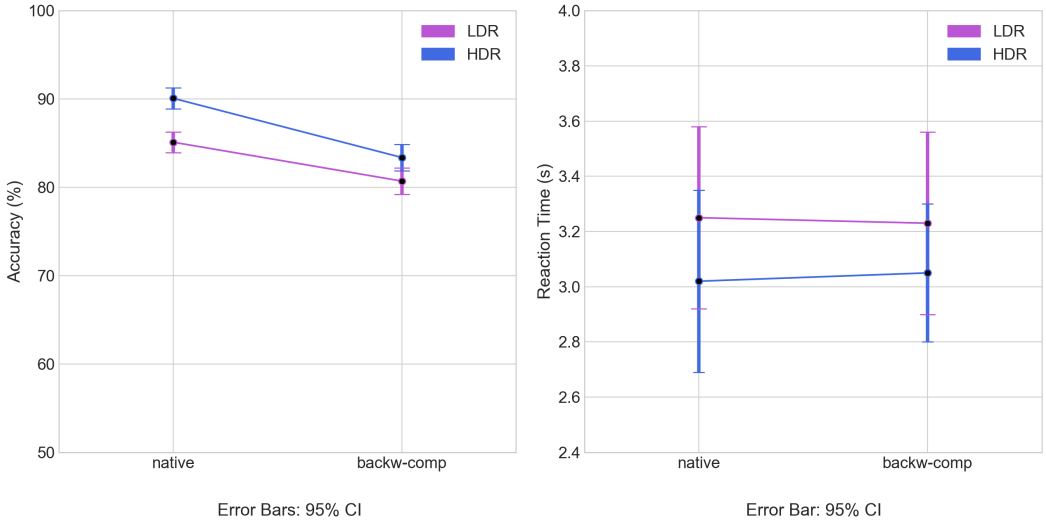


Figure 8.2: Comparison of the effect of HDR and LDR stimuli across two experiments of unfamiliar face matching task for Accuracy and Reaction Time. *native* indicates the *HDRvsLDR* experiment results, while *backw-comp* indicates the *Backwards-Compatibility HDR* experiment results. Accuracy is significantly better for *native:HDR* than for *backwards-compatible:HDR*, while there is no significant difference between the two pipelines for Reaction Time.

There was a significant main effect of *DR*, with Pillai's trace V=0.35, F(2,75)=20.24, p <0.01, and also a significant main effect of *position*, with Pillai's trace V=0.58, F(2,75)=50.89, p <0.01. There was no other two-way significant interaction: *pipeline* ×*DR* with Pillai's trace V=0.02, F(2,75)=0.72, p =0.49. There was no other two-way significant interaction for *pipeline* ×*position* with Pillai's trace V=0.01, F(2,75)=0.54, p =0.59. There was no

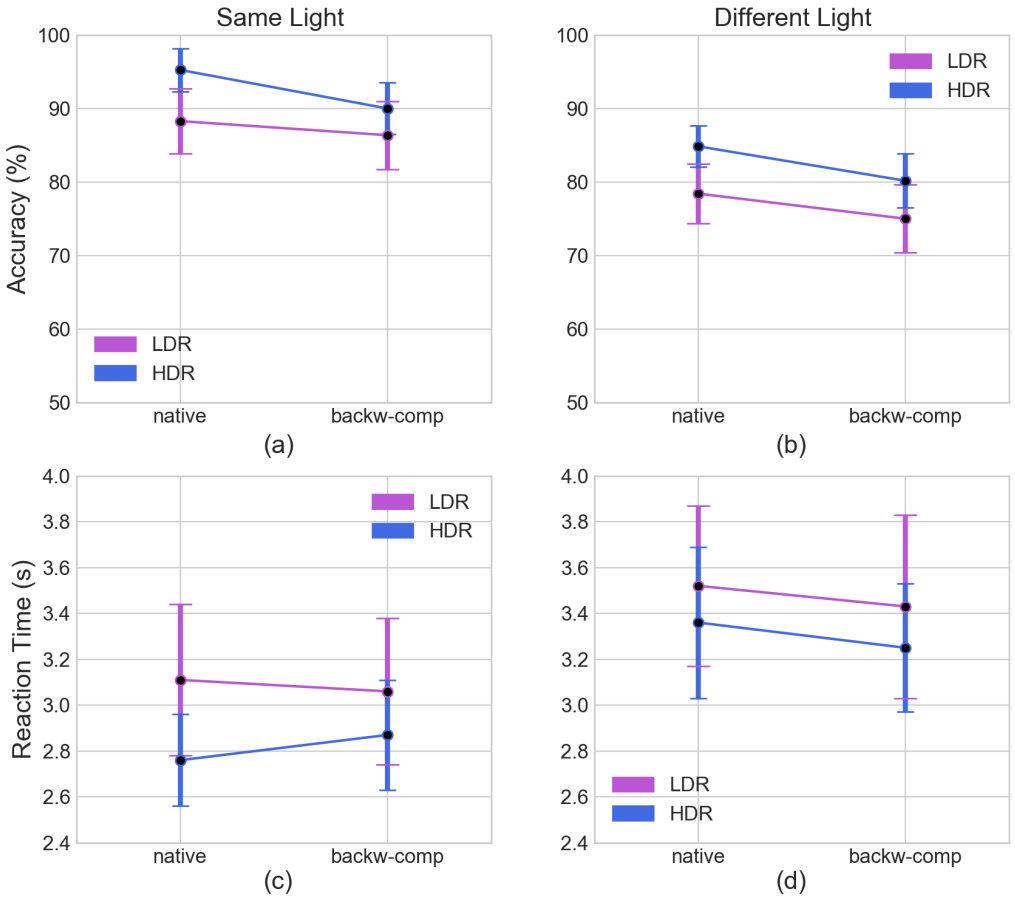


Figure 8.3: Four graphs representing the results of *DR* across *pipeline* (i.e. *native* and *backward-compatible*) for each lighting condition (i.e. *same* and *different*). Error bars indicate 95% CI. The two graphs on top (a) and (b) show *accuracy* as percentage of correct answers, while the two graphs on the bottom (c) and (d) show Reaction Time (*time*) expressed in seconds. For *accuracy*, when stimuli are HDR, results are clearly higher than when stimuli are LDR. However, in the case of *backwards-compatibility HDR*, provides less advantage when compared with the *native* pipeline. In regard to Reaction Time (*time*) there is no significant difference across *pipeline*, diagram (c) and(d). Moreover, when *position:same* observers are significantly better in performing a face recognition task than when the lighting is *different*, diagrams (b) and (d).

other two-way significant interaction for *position* $\times$*DR* with Pillai's trace V=0.02, F(2,75)=0.58, p =0.56. Also, no significant main effect for three way interaction *position*$\times$*DR* $\times$ *pipeline* with Pillai's trace V=0.02, F(2,75)=0.57,

p =0.57 was found. Figure 8.2 illustrates the overall results for *pipeline* × *DR*, while Figure 8.3 has a more detailed representation of the results for each lighting condition.

8.4.1 Univariate Analysis: *accuracy*

The univariate analysis for *accuracy* was significant for the between-participants factor *pipeline*, F(1,76 =5.73, p <0.05. A main effect of *DR* was found, F(1,76)=26.08, p <0.01. Moreover a main effect of *position* was found, F(1,76)=52.63, p <0.01. No interaction was found for: *DR* × *pipeline*, F(1,76)=1.11, p =0.30; *position* × *pipeline*, F(1,76)=0.54, p =0.46; *DR* × *position*, F(1,76)=0.11, p =0.74; *DR* × *position* × *pipeline*, F(1,76)=0.52, p =0.48.

Post hoc tests with Bonferroni adjustment showed that *pipeline:native* led to significantly higher accuracy (μ=86.73) than using *pipeline:backwards* (μ =82.91), with a mean difference of 3.82, CI[0.64, 7.00], t(78)= 2.39, p <0.05. Post hoc tests on *DR* show that response to *HDR* stimuli (μ = 87.60) is significantly more accurate than *LDR* stimuli (μ = 82.05). This difference amounting to 5.55, was significant, CI[3.38, 7.72], t(78)=5.10, p <0.01. Moreover, post hoc tests on *position* show that response to *same* stimuli (μ = 90.01) is significantly more accurate than *different* stimuli (μ = 79.63), this difference (10.38) was significant, CI[7.45, 13.30], t(78)=7.12, p <0.01.

These results demonstrate that *pipeline* has a significant effect on the accuracy of perceptual matching of unfamiliar faces with *native:HDR* images significantly improving accuracy rates.

8.4.2 Univariate Analysis: *time*

The univariate analysis for *time* was not significant for the between-subject factor *pipeline*, F(1,76) = 0.02, p =0.87. A main effect of *DR* was found, F(1,76)=13.59, p <0.01. Moreover, a main effect of *position* was found, F(1,76)=52.63, p <0.01. No interaction was found for: *DR* × *pipeline*, F(1,76)=0.30, p =0.59; *position* × *pipeline*, F(1,76)=1.07, p =0.31; *DR* × *position*, F(1,76)=1.12, p =0.29; *DR* × *position* × *pipeline*, F(1,76)=0.99, p =0.32.

Post hoc tests on *DR* show that response to *HDR* stimuli ($\mu = 3.06$) is significantly more accurate than *LDR* stimuli ($\mu = 3.28$), this difference (0.22) was significant CI[0.10, 0.33], t(78)=3.66, p <0.01. Moreover, post hoc tests on *position* show that response to *same* lighting stimuli ($\mu = 2.95$) is significantly more accurate than *different* lighting stimuli ($\mu = 3.39$). This difference (0.44) was significant, CI[0.32, 0.56], t(78)=7.21, p <0.01.

These results show no significant effect of *pipeline* on the speed of perceptual matching of unfamiliar faces. However, for both pipelines, *HDR* elicits quicker response than *LDR*. Due to the nature of our research question, a further analysis on the impact of *DR* was performed.

8.4.3 Multivariate Analysis: *DR*: *LDR*

A two-factor mixed MANOVA for *DR*:*LDR* using Pillai's trace showed no significant effect of *pipeline*, with Pillai's trace V=0.02, F(2,75)=0.77, p =0.47. However, there was a significant main effect of *position*, with Pillai's trace V=0.41, F(2,75)=26.05, p <0.01. The analysis shows no significant interaction of *pipeline* ×*position*, with Pillai's trace V = 0.01, F(2,75)=0.10, p =0.90.

In the univariate test for *time*, the results show the main effect of *position*, F(1,76)=30.55, p <0.01. *position:same* allows faster recognition ($\mu = 3.08$) than *position:different* ($\mu = 3.47$). Similarly, there is a main effect of *position* in the univariate results for *accuracy*, F(1,76)=24.68, p <0.01. *position:same* has higher accuracy rates ($\mu = 87.35$) than *position:different* ($\mu = 76.73$).

These results show that overall there is no effect of *DR* between the full-LDR imaging pipeline and the backward compatible pipeline, where the LDR stimuli are processed to be displayed on an HDR screen using an EO based on CNNs. Lighting, on the other hand, has a major impact on the accuracy and recognition speed. Hence we can reject H_1^3 and H_1^4.

8.4.4 Multivariate Analysis: *DR*: *HDR*

A two-factor mixed MANOVA for *DR*:*HDR* using Pillai's trace showed a significant effect of *pipeline*, with Pillai's trace V=0.10, F(2,75) =4.26, p <0.05; and also a significant main effect of *position* was observed, with Pillai's trace V=0.51, F(2,75)=38.94, p <0.01. No significant interaction *pipeline* ×*position*,

with Pillai's trace V = 0.02 F(2,75)=0.91, p =0.40.

The univariate test for *DV time* showed no main effect of *pipeline* F(1,76)= 0.8×10^{-5}, p =0.99: participants had similar *RT* for *native* (μ =3.05) and *backwards* (μ = 3.05). There was a main effect of *position*, F(1,76)= 37.29, p <0.01, with participants choosing quickly when *position:same* (μ = 2.81) than when *position:different* (μ = 3.30).

The univariate test for *DV accuracy* showed a significant effect of *pipeline*, F(1,76)=8.56, p <0.01, with *experiment:native* scoring a higher percentage of correct answers (μ=90.08) than *experiment:backward* (μ=85.11). There was also a significant main effect of *position*, F(1,76)=38.29, p <0.01, with *position:same* eliciting more accurate response (μ = 92.66) than *position:different* (μ = 82.53).

These results show that overall there is a significant effect of *DR* between the full-HDR imaging pipeline and the backward compatible pipeline, where the HDR stimuli are tone-mapped to be displayed on an LDR display. Lighting has also a major impact on the accuracy and recognition speed. These results lead us to accept H_1^2, but reject H_1^1, as no significant effect of *pipeline* was shown on *time*.

8.5 Discussion

This chapter analysis aimed at answering the question of whether, for the unfamiliar face matching task, there is any difference in having stimuli presented with a native or a backward-compatible HDR pipeline. The results shown in the previous sections demonstrate that there is a significant difference between the full-HDR imaging pipeline and the backward-compatible one, where HDR images have been tone-mapped to be displayed on an LDR screen. Generally, the adoption of the HDR modality is always preferable to the LDR, as it improves accuracy and shortens reaction times. More specifically, HDR stimuli created using the full-HDR pipeline provide higher *accuracy*, hence we can accept H_1^2. However, the analysis showed no significant difference on reaction times, leading to the rejection of H_1^1. These results suggest that to obtain an improvement of the overall accuracy it is recommended the adoption of the full-HDR pipeline, while the adoption of the backward-compatible HDR pipeline (i.e. HDR capture and LDR display) allows to obtain only comparable

reaction times.

On the other hand there was no significant difference between the *native:LDR* pipeline and *backward-compatible:LDR* pipeline, where an EO was used to retarget LDR captured images to HDR display. This leads to reject both H_1^3 and H_1^4 and suggests that the loss of perceptual information due to an LDR capture cannot - to this day - be recovered. Hence, the increased dynamic range of the image operated by the EO does not suffice to compensate for the loss of perceptually relevant information.

Lighting position has been shown to have a considerable impact on the accuracy and reaction times. The results obtained are in line with the literature [Johnston *et al.*, 1991; Braje *et al.*, 1998] confirming that unfamiliar face recognition is illumination-dependant. Notably, the accurate light reproduction in case of the *DR:HDR* allows for better accuracy and speed if compared to the *DR:LDR* case. The accuracy obtained in case of *DR:HDR* and *position:different* is higher (μ =82.53%) than the accuracy generally obtained in other experiments testing for similar effects (e.g. Braje 2003 reports 70% accuracy for different-illumination trials).

The limitations of this study resides in the inability to compare respectively *native:HDR* with *backwards-compatible:LDR* and *native:LDR* with *backwards-compatible:HDR*, due to the specific characteristics of the imaging in the *Backwards-Compatibility HDR* experiment.

In conclusion, a complete upgrade of the pipeline to HDR is always desirable, rather than a partial upgrade of LDR with HDR capture as this further improves accuracy (5.84% increase on average). It is not recommendable to upgrade the *native:LDR* to *backwards-compatible:LDR* (LDR capture and HDR display), because no performance improvement has been found.

8.6 Summary

This chapter has introduced an experiment that evaluated native vs backwards-compatible pipelines, establishing the superiority of a native HDR imaging when performing an unfamiliar face matching task over a pipeline where HDR images are subsequently tone-mapped to be displayed on an LDR screen. Moreover, it has been shown that the upgrade of LDR face stimuli to HDR to be visualised

on HDR screens does not provide any benefit. In the next chapter the main contributions, the limitations and the possible avenue of research of this work will be outlined.

Chapter 9

Conclusions

This thesis has presented several experiments aimed to verify whether the adoption of the HDR imaging pipeline can improve performance in the unfamiliar face matching task. This chapter will summarise the results obtained, highlight the contributions and describe future work.

9.1 Overview

Unfamiliar face matching is a problem of crucial importance in many contexts, from security to forensics. As widely shown by the literature on this topic (see Chapter 3), there are still several open questions related to how humans perceive faces and what the main factors affecting performance in the unfamiliar face matching task are.

When performing a face recognition mediated by imaging systems, such as screens and cameras, it is important to limit the impact of poor imaging quality on the viewer performance so as to minimise (possibly avoid) erroneous identification [Norell *et al.*, 2015; Noyes and Jenkins, 2017]. The work presented in this thesis aimed at investigating whether the adoption of the HDR imaging technology could be beneficial to improve the performance of such a highly error-prone task. The research question asked was:

Does HDR imaging improve unfamiliar face matching performance?

The conclusion from this thesis is that the adoption of HDR imaging is beneficial to observers while performing an unfamiliar face matching task. HDR stimuli

allow to achieve higher overall accuracy rates (90.08%), than LDR (83.38%), (p <0.01), and provide an average reduction of reaction time of about 0.25s (p <0.05).

The results obtained have also confirmed the negative impact that disparity in the illumination has on face recognition accuracy, in line with the literature [Braje, 2003; Favelle *et al.*, 2017]. When lighting conditions are favourable (i.e. same lighting) the HDR imaging allows for levels of accuracy of over 95% and reaction times of 2.76s, while, with unfavourable lighting conditions the accuracy levels are still around 85% and reaction times are around 3.35s. Accuracy using LDR, on the other hand, falls to 78.44% in case of unfavourable lighting with reaction times of 3.52s.

Moreover, the results presented in this thesis indicate that the adoption of an enhanced LDR pipeline by utilising HDR imaging in the capture stage is beneficial in reducing reaction time, although full-HDR provides an increase in accuracy of 5.84% compared to the backwards-compatible HDR (p <0.01).

In summary, this thesis has shown the efficacy of the adoption of HDR technology in improving face matching tasks' performance, and, at the same time, has highlighted the limitations of the currently adopted imaging standards (i.e. LDR) in contexts such as forensic and surveillance. The aforementioned results could have great implications for the improvement of the currently adopted technology.

The answer to the research question was unravelled in several stages. Five main objectives were identified and listed in Section 1.3. In the following section, an overview of the actions undertaken to meet each objective is given, highlighting the specific contributions.

9.2 Objectives and Contributions

The main objective of this thesis was to investigate whether the process of unfamiliar face recognition could be supported by the adoption of advanced imaging techniques that exhibit higher fidelity in the reproduction of reality. Several studies have previously investigated the impact of variations in the characteristics of imaging have on performance (see [Braje, 2003; Longmore *et al.*, 2008; Burton *et al.*, 2010; Bindemann *et al.*, 2013; Fysh and Bindemann, 2018]).

However, to the best of our knowledge, this is the first study investigating the impact of dynamic range when HDR images are used as stimuli.

Previous research has shown unfamiliar face matching to be a highly error-prone task, due to the fact that the observer lacks "experience" of a specific face (i.e. lacks in rich memory representations tolerant to variability that can only be created by repeated observation of a face under different lights, angles, etc.), for example, a police officer observing a CCTV footage [Bruce *et al.*, 1999; Burton, 2013]. Many perceptual experiments have shown how sensitive this task is to the imaging characteristics, besides the obvious difficulties due to the recognition task itself [Jenkins *et al.*, 2011; Fysh and Bindemann, 2017]. In particular, disparity in the illumination of the two faces compared can negatively influence the observer's performance [Braje, 2003; Longmore *et al.*, 2008]. The intrinsic characteristics of the HDR imaging to accurately capture the scene's dynamic range and reproduce real-world illumination with higher fidelity, make it an ideal candidate for the investigation of its effect on the unfamiliar face matching task.

Chapter 2 offered a broad review of the digital imaging process and also offered a detailed overview of the main stages of the HDR imaging pipeline. Chapter 3 expounded the general problem related to face perception, recognition and the specific problem of unfamiliar face matching. With the work presented in Chapter 2 and Chapter 3 the first objective was met: *Review of the literature regarding both the HDR technology pipeline and the process underlying face perception.*

The first contribution of this thesis consists of ascertaining the research gap: the lack of studies investigating the impact of dynamic range on face recognition. This work is presented in Chapter 4 together with a detailed methodology to address said research gap.

9.2.1 Warwick HDR Face Dataset

The first step required to address the research question was the creation of an HDR faces dataset. The lack of an appropriate HDR stimuli-set would not permit the investigation of the problem of unfamiliar face recognition further. There are only two HDR faces datasets available in the literature (i.e. Ige

et al. [2016] and Korshunov *et al.* [2015]), but the lack of several characteristics, including information on the scene luminance levels, standardised camera-to-face distance, have made them not appropriate for the studies conducted in this thesis. Accordingly, the second objective of this study was: the *Creation of a radiometrically calibrated HDR faces dataset.* The creation of the Warwick HDR Face Dataset met this objective.

The dataset encompasses 170 radiometrically calibrated HDR images of 17 male subjects, each portrayed under five different lighting conditions with two different captures of the same lighting condition. A detailed description of the steps followed to create the dataset is presented in Chapter 5.

This is the first radiometrically calibrated HDR face stimuli set available in the literature. HDR imaging is capable of accurately capturing and reproducing all the colours and brightness levels visible to the human visual system [Mantiuk *et al.*, 2016].

An HDR stimuli dataset is a useful resource to explore the impact of dynamic range on face recognition. Moreover, the availability of different portraits for each lighting condition (i.e. two different images with the same lighting) allows the dataset to be used in experiments aiming to elicit face-related cognitive processes in the observer, rather than low-level image matching ones [Burton, 2013]. Table 4.1, reported in Chapter 4, shows a detailed list of image characteristics manipulations that are featured in the Warwick HDR Face Dataset and how this compares with datasets adopted in other unfamiliar face matching experiments. In particular its most relevant features are:

- contains HDR images;
- the images have been radiometrically calibrated;
- offers a high image resolution (1000 pixels/face);
- subjects are portrayed under five different controlled lighting conditions (fully lit, and top, below, left and right lighting);
- the availability of two captures for each lighting condition.

9.2.2 Subjective Evaluation of High vs Low Dynamic Range Imaging for Face Matching

The third contribution of this thesis is the execution of a subjective experiment where the Warwick HDR Face Dataset was used in a controlled experiment to measure accuracy and reaction times of human participants when HDR and LDR stimuli were used.

Chapter 6 describes a subjective experiment ($N = 39$), *HDRvsLDR*, conducted with the aim of comparing an individual's performance (i.e. accuracy and reaction time) when performing an unfamiliar face matching task. The full-HDR and LDR pipeline were compared, from the capturing to the display stage.

This experiment allowed to compare the two pipelines and provided a quantitative measure of the effect of using HDR stimuli on performance. This experiment is the first unfamiliar face matching experiment evaluating the impact of dynamic range on the face recognition process. The results obtained show scope for the adoption of HDR imaging in scenarios where performing an unfamiliar face matching task is required, as a more realistic reproduction of light is shown to be conducive to better accuracy and shorter reaction time. As already mentioned in Section 9.1, the adoption of HDR stimulus data allows to achieve higher accuracy rates (90.08%), than LDR (83.38%), (p <0.01), and provides an average reduction of reaction time of about 0.25s (p <0.05). Furthermore, the results of this experiment have confirmed the detrimental impact of differences in lighting conditions when matching unfamiliar faces (see for example [Braje, 2003]). Although accuracy in this case is generally lower than the same-lighting case, the use of HDR stimuli (84.88%) allows to achieve better performance than LDR (78.44%), (p <0.01), showing potential for the adoption of this technology in contexts where accurate individuals' identification is critical.

9.2.3 Backwards-Compatibility HDR

The fourth contribution of this work is the completion of a second subjective experiment where the Warwick HDR Face Dataset was used to measure accuracy and reaction time of human participants when specific elements of the LDR

pipeline are replaced by HDR ones. With this experiment Objective 4 was met: *Adoption of the Warwick HDR Face Dataset stimuli in a face matching experiment aimed at measuring accuracy and reaction times performance of an enhanced LDR pipeline through the application of advanced imaging algorithms - i.e. Expansion Operators and Tone Mapping Operators.* Considering the results described in Chapter 6, the availability of a technology that allows for backwards-compatible HDR imaging (i.e. EOs and TMOs) and also the fact that an upgrade of the LDR pipeline would happen in phases, it is logical to ask which stage of the LDR pipeline leads to maximum benefit when enhanced.

The literature offers several solutions to address the problem of retargeting LDR imaging to an HDR display and vice-versa. Chapter 7 illustrates the problem in detail and describes the procedures followed to perform a new subjective ($N = 39$) experiment (i.e. *Backwards-Compatibility HDR* experiment). Participants were shown HDR and LDR stimuli in a two forced-choice experiment where their performance in unfamiliar face matching was measured. The Warwick HDR Face Dataset was used for the creation of the stimuli. The HDR stimuli were obtained from LDR images by adopting a machine-learning based Expansion Operator (EO_LDR) [Marnerides *et al.*, 2018]. The LDR stimuli were obtained by applying a Tone-Mapping Operator to HDR stimuli (TMO_HDR) [Reinhard *et al.*, 2002]. Results show that using TMO_HDR is more advantageous compared to EO_LDR both for accuracy (TMO_HDR:μ=85.11% and EO_LDR:μ=80.70%, $p < 0.01$) and reaction time (TMO_HDR:μ=3.06s and EO_LDR:μ=3.25s, p <0.05).

This experiment provides quantitative data in support of the adoption of HDR imaging in the capture stage even if the stimuli are subsequently displayed on an LDR device. The preservation of image fidelity achieved though HDR capture seems to be more effective than the artificial reconstruction of light information even if adopting one of the latest and most performing machine-learning based expansion operators (i.e. ExpandNet).

Furthermore, in line with the literature and the *HDRvsLDR* experiment's results (Chapter 6), the *Backwards-Compatibility HDR* experiment has shown that light disparity in the stimuli affects performance. The average performance results across *DR* are worse for *different* lighting (accuracy: μ=77.60% and RT:μ=3.34s) than for *same* lighting (accuracy:μ=88.22, RT:μ=2.96s). However,

the adoption of TMO_HDR over EO_LDR has a significant effect on improving both accuracy (TMO_HDR:μ=80.19%, EO_LDR:μ=75.02%, $p < 0.05$) and reaction times (TMO_HDR:μ=3.25s, EO_LDR:μ=3.43s, $p < 0.05$).

9.2.4 Pipeline Comparison

The fifth contribution of this work is the final experiment described in Chapter 8, where the two pipeline modalities (i.e. *native* and *backward-compatible*) where compared. This allowed to meet the final objective: *Comparison of the results obtained in the Objective 3 and 4 to establish if the adoption of the full HDR pipeline or a partial upgrade of the LDR pipeline is adequate for sufficiently improving performance.*

A mixed design ($N = 78$) was used to analyse the results obtained in the *HDRvsLDR* and the *Backwards-Compatibility HDR* experiment. The analysis in Chapter 8 has confirmed the results obtained previously, showing that it is preferable to adopt the HDR pipeline, as this modality increases accuracy and reduces reaction time. Moreover, a significant increase (5.84%) in the accuracy is registered when the full-HDR is adopted. The experiment showed no advantage in the upgrade of the LDR pipeline at the display stage.

This experiment provides a quantitative assessment of the impact of a full-HDR pipeline versus an LDR pipeline where HDR elements have been integrated. The results indicate that the adoption of the full-HDR pipeline allows for higher accuracy than the full-LDR or a partially upgraded LDR pipeline. These results provide evidence in support of the adoption of HDR technology in scenarios where accuracy and confidence are required. The HDR is a mature technology and deserves the attention from digital surveillance and forensics communities.

9.3 Future Work

The work presented in this thesis is the first attempt to investigate the impact of dynamic range on the process of face recognition for the unfamiliar face matching task. This is a highly error-prone task and it is very susceptible to low performance as a consequence of the characteristics of the images used.

One of the main limitations of this work is related to the size of the dataset created, which exposes it to ecological validity and generalisability critiques. However, precautions were taken to minimise the impact of image memory: face repetitions were limited according to what the literature suggests [Longmore *et al.*, 2008] (i.e. pictures of the same individuals were limited to maximum 10 repetitions). The use of only male faces was due to the fact that participation in the data collection of the Warwick HDR Face Dataset was on a voluntary basis. Although the faces of two female participants were captured, these portraits were not included as it has been shown that sex categorisation is completed generally more quickly (between 5 to 7ms [Bruce *et al.*, 1993]) than familiarity judgement [Bruce and Young, 1986; Wild *et al.*, 2000] (250-300 ms). Therefore, including these female portraits would have biased the results of the following experiments. Future work will address this disparity and also will aim at including more ethnicities, and more challenging pairs such as lookalikes or the same face with different features (e.g. cosmetics or facial hair).

A natural continuation of this thesis is the addition of colour components to the stimuli. Furthermore, having established the impact of HDR imaging on the unfamiliar face matching task, future work could entail the capture of a new set of HDR radiometrically calibrated portraits with uncontrolled lighting conditions (e.g. outdoor). The accurate reproduction of lighting conditions offered by HDR imaging could be especially beneficial in scenarios where the current technology has shown limits [Lee *et al.*, 2009]. Additionally, this thesis has highlighted the lack of TMOs specific for faces. It would be beneficial to create a TMO dedicated to the preservation of perceptually relevant features, like for example internal features (i.e. nose, mouth, eyes), over secondary elements (e.g. background). This would be useful for video messaging applications. Also, considering the substantial storage space needed for HDR images encoding, the creation of an HDR compression algorithm dedicated to faces, would allow for the preservation of perceptually relevant information only, making the transmission and storage of these images more manageable.

Additionally, on a broader scope, future research could look into discovering the perceptual motivation behind the increased accuracy and shortened reaction time using HDR stimuli. This thesis has demonstrated the efficacy of

the HDR stimuli but has not proposed a theory of perception that can explain this phenomenon. The availability of the Warwick HDR Face Dataset can be beneficial to this end.

Additionally, the availability of emerging technology such as light field cameras (see Wu *et al.* [2017] for an overview), could be exploited to tackle humans susceptibility to lighting conditions and viewing angles.

9.4 Final Remarks

High Dynamic Range imaging is a technology that has not been entirely explored in its capability to reproduce reality with higher fidelity. It is a prominent candidate to support problems where accuracy and precision are essential.

This work has shown that providing human observers with higher quality stimuli can greatly improve their performance in the unfamiliar face matching task. This thesis can serve as a starting point for a new research route fostering the upgrade of surveillance cameras and display systems. In contexts such as forensics and surveillance, the adoption of advanced imaging technology can contribute overcoming the many limitations of the one currently in use.

Appendix A

PI20: Self assessment questionnaire on face recognition ability

Table A.1 illustrates the 20-item prosopagnosia index (PI20). This self-assessment test, proposed by Shah *et al.* [2015], allows to detect the presence of prosopagnosic traits. Each respondent has to select, on a scale from 1 to 5, the extent to which each statement describes their face recognition experiences. PI20 scores above 65 are indicative of:

- mild prosopagnosia: 65-74;
- moderate prosopagnosia: 75-84;
- severe prosopagnosia: 85-100.

,

1	My face recognition ability is worse than most people
2	I have always had a bad memory for faces
3	I find it notably easier to recognize people who have distinctive facial features
4	I often mistake people I have met before for strangers
5	When I was at school I struggled to recognize my classmates
6	When people change their hairstyle, or wear hats, I have problems recognizing them
7	sometimes have to warn new people I meet that I am 'bad with faces'
8*	I find it easy to picture individual faces in my mind
9*	I am better than most people at putting a 'name to a face'
10	Without hearing people's voices, I struggle to recognize them
11	Anxiety about face recognition has led me to avoid certain social or professional situations
12	I have to try harder than other people to memorize faces
13*	I am very confident in my ability to recognize myself in photographs
14	I sometimes find movies hard to follow because of difficulties recognizing characters
15	My friends and family think I have bad face recognition or bad face memory
16	I feel like I frequently offend people by not recognizing who they are
17*	It is easy for me to recognize individuals in situations that require people to wear similar clothes (e.g. suits, uniforms and swimwear)
18	At family gatherings, I sometimes confuse individual family members
19*	I find it easy to recognize celebrities in 'before-they-were-famous' photos, even if they have changed considerably
20	It is hard to recognize familiar people when I meet them out of context (e.g. meeting a work colleague unexpectedly while shopping)

Table A.1: PI20 items. Items with * symbol are reversely scored. Scores above 65 might be indicative of mild to severe prosopagnosia.

Appendix B

Ethical Approvals

PRIVATE

Ms R Suma
WMG
University of Warwick
Coventry
CV4 7AL

9 November 2017

Dear Ms Suma

Study Title and BSREC Reference: *High resolution, radiometrically characterised High Dynamic Range face dataset* REGO-2017-2069

Thank you for submitting the revisions to the above-named study to the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee for approval.

I am pleased to confirm that approval is granted and that your study may commence.

In undertaking your study, you are required to comply with the University of Warwick's *Research Data Management Policy*, details of which may be found on the Research and Impact Services' webpages, under "Codes of Practice & Policies" » "Research Code of Practice" » "Data & Records" » "Research Data Management Policy", at:
http://www2.warwick.ac.uk/services/ris/research_integrity/code_of_practice_and_policies/research_code_of_practice/datacollection_retention/research_data_mgt_policy

You are also required to comply with the University of Warwick's *Information Classification and Handling Procedure*, details of which may be found on the University's Governance webpages, under "Governance" » "Information Security" » "Information Classification and Handling Procedure", at:
http://www2.warwick.ac.uk/services/gov/informationsecurity/handling.
Investigators should familiarise themselves with the classifications of information defined therein, and the requirements for the storage and transportation of information within the different classifications:

> *Information Classifications:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/classifications
> *Handling Electronic Information:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/electronic/
> *Handling Paper or other media*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/paper/.

Please also be aware that BSREC grants **ethical approval** for studies. **The seeking and obtaining of *all* other necessary approvals is the responsibility of the investigator.**

These other approvals may include, but are not limited to:

1. Any necessary agreements, approvals, or permissions required in order to comply with the University of Warwick's Financial Regulations and Procedures.
2. Any necessary approval or permission required in order to comply with the University of Warwick's Quality Management System and Standard Operating Procedures for the governance, acquisition, storage, use, and disposal of human samples for research.
3. All relevant University, Faculty, and Divisional/Departmental approvals, if an employee or student of the University of Warwick.
4. Approval from the applicant's academic supervisor and course/module leader (as appropriate), if a student of the University of Warwick.
5. NHS Trust R&D Management Approval, for research studies undertaken in NHS Trusts.
6. NHS Trust Clinical Audit Approval, for clinical audit studies undertaken in NHS Trusts.
7. Approval from Departmental or Divisional Heads, as required under local procedures, within Health and Social Care organisations hosting the study.
8. Local ethical approval for studies undertaken overseas, or in other HE institutions in the UK.
9. Approval from Heads (or delegates thereof) of UK Medical Schools, for studies involving medical students as participants.
10. Permission from Warwick Medical School to access medical students or medical student data for research or evaluation purposes.
11. NHS Trust Caldicott Guardian Approval, for studies where identifiable data is being transferred outside of the direct clinical care team. Individual NHS Trust procedures vary in their implementation of Caldicott guidance, and local guidance must be sought.
12. Any other approval required by the institution hosting the study, or by the applicant's employer.

There is no requirement to supply documentary evidence of any of the above to BSREC, but applicants should hold such evidence in their Study Master File for University of Warwick auditing and monitoring purposes. You may be required to supply evidence of any necessary approvals to other University functions, e.g. The Finance Office, Research & Impact Services (RIS), or your Department/School.

May I take this opportunity to wish you success with your study, and to remind you that any Substantial Amendments to your study require approval from BSREC before they may be implemented.

Yours sincerely

pp. [redacted]

Dr David Ellard
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Sub-Committee
Research & Impact Services
University of Warwick
Coventry, CV4 8UW.
E: BSREC@Warwick.ac.uk

http://www2.warwick.ac.uk/services/ris/research_integrity/researchethicscommittees/biomed

PRIVATE

Miss R Suma
WMG
University of Warwick
Coventry
CV4 7AL

14 June 2017

Dear Miss Suma

Study Title and BSREC Reference: *Investigating faces recognition ability according to variations in luminance* REGO-2017-2024

Thank you for submitting your revisions to the above-named study to the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee for approval.

I am pleased to confirm that approval is granted and that your study may commence.

In undertaking your study, you are required to comply with the University of Warwick's *Research Data Management Policy*, details of which may be found on the Research and Impact Services' webpages, under "Codes of Practice & Policies" » "Research Code of Practice" » "Data & Records" » "Research Data Management Policy", at:
http://www2.warwick.ac.uk/services/ris/research_integrity/code_of_practice_and_policies/research_code_of_practice/datacollection_retention/research_data_mgt_policy

You are also required to comply with the University of Warwick's *Information Classification and Handling Procedure*, details of which may be found on the University's Governance webpages, under "Governance" » "Information Security" » "Information Classification and Handling Procedure", at:
http://www2.warwick.ac.uk/services/gov/informationsecurity/handling.
Investigators should familiarise themselves with the classifications of information defined therein, and the requirements for the storage and transportation of information within the different classifications:

> *Information Classifications:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/classifications
> *Handling Electronic Information:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/electronic/
> *Handling Paper or other media*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/paper/.

Please also be aware that BSREC grants **ethical approval** for studies. **The seeking and obtaining of *all* other necessary approvals is the responsibility of the investigator.**

These other approvals may include, but are not limited to:

1. Any necessary agreements, approvals, or permissions required in order to comply with the University of Warwick's Financial Regulations and Procedures.
2. Any necessary approval or permission required in order to comply with the University of Warwick's Quality Management System and Standard Operating Procedures for the governance, acquisition, storage, use, and disposal of human samples for research.
3. All relevant University, Faculty, and Divisional/Departmental approvals, if an employee or student of the University of Warwick.
4. Approval from the applicant's academic supervisor and course/module leader (as appropriate), if a student of the University of Warwick.
5. NHS Trust R&D Management Approval, for research studies undertaken in NHS Trusts.
6. NHS Trust Clinical Audit Approval, for clinical audit studies undertaken in NHS Trusts.
7. Approval from Departmental or Divisional Heads, as required under local procedures, within Health and Social Care organisations hosting the study.
8. Local ethical approval for studies undertaken overseas, or in other HE institutions in the UK.
9. Approval from Heads (or delegates thereof) of UK Medical Schools, for studies involving medical students as participants.
10. Permission from Warwick Medical School to access medical students or medical student data for research or evaluation purposes.
11. NHS Trust Caldicott Guardian Approval, for studies where identifiable data is being transferred outside of the direct clinical care team. Individual NHS Trust procedures vary in their implementation of Caldicott guidance, and local guidance must be sought.
12. Any other approval required by the institution hosting the study, or by the applicant's employer.

There is no requirement to supply documentary evidence of any of the above to BSREC, but applicants should hold such evidence in their Study Master File for University of Warwick auditing and monitoring purposes. You may be required to supply evidence of any necessary approvals to other University functions, e.g. The Finance Office, Research & Impact Services (RIS), or your Department/School.

May I take this opportunity to wish you success with your study, and to remind you that any Substantial Amendments to your study require approval from BSREC before they may be implemented.

Yours sincerely

pp. [redacted]

Professor John Davey
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Sub-Committee
Research & Impact Services
University of Warwick
Coventry, CV4 8UW.
E: BSREC@Warwick.ac.uk

http://www2.warwick.ac.uk/services/ris/research_integrity/researchethicscommittees/biomed

Bibliography

Adams, A. and Baker, R. *The Camera*. Bulfinch, 1995a. ISBN 9780821221846.

Adams, A. and Baker, R. *The Negative*. Bulfinch, 1995b. ISBN 9780821221860.

Adams, A. and Baker, R. *The Print*. Bulfinch, 1995c. ISBN 9780821221877.

Agrafiotis, P., Stathopoulou, E. K., Georgopoulos, A., and Doulamis, A. D. Hdr imaging for enchancing people detection and tracking in indoor environments. In *VISAPP (2)*, pages 623–630, 2015.

Akyüz, A. O. and Gençtav, A. A reality check for radiometric camera response recovery algorithms. *Computers & Graphics*, 37(7):935–943, nov 2013. ISSN 00978493. doi: 10.1016/j.cag.2013.06.003.

Althoff, R. R. and Cohen, N. J. Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4):997, 1999.

ANSI IT8.7/1-1993 (R1999). Graphic technology - color transmission target for input scanner calibration. tech. rep. ansi it8.7/1-1993 (reaffirmation 1999), ansi, 1999. Technical report, American National Standard Institute, New York, 1999.

Ashikhmin, M. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics Workshop on Rendering*, EGRW '02, pages 145–156, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association. ISBN 1-58113-534-3.

Bachmann, T. Different trends in perceptual pattern microgenesis as a function of the spatial range of local brightness averaging. *Psychological Research*, 49 (2-3):107–111, aug 1987. ISSN 0340-0727. doi: 10.1007/BF00308675.

Bachmann, T. Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, 3(1):87–103, 1991. ISSN 14640635. doi: 10.1080/09541449108406221.

Banterle, F. and Unger, J. Creating hdr video using retargetting. In *High Dynamic Range Video*, pages 45–59. Elsevier, 2017.

Banterle, F., Ledda, P., Debattista, K., and Chalmers, A. Inverse tone mapping. *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia GRAPHITE 06*, 1 (212):349, 2006. doi: 10.1145/1174429.1174489.

Banterle, F., Artusi, A., Debattista, K., and Chalmers, A. *Advanced High Dynamic Range Imaging: Theory and Practice (2nd Edition)*. AK Peters (CRC Press), Natick, MA, USA, July 2017. ISBN 9781498706940.

Bayer, B. E. Color imaging array, U.S. patent 3971065.

Bhatia, S. K., Lakshminarayanan, V., Samal, A., and Welland, G. V. Human face perception in degraded images. *Journal of Visual Communication and Image Representation*, 6(3):280–295, 1995.

Bindemann, M., Attard, J., Leach, A., and Johnston, R. The Effect of Image Pixelation on Unfamiliar Face Matching. *Applied Cognitive Psychology*, 27 (November):707–717, 2013.

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., and Bate, S. An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex*, 82:48–62, 2016. ISSN 19738102. doi: 10.1016/j.cortex.2016.05.003.

Bowmaker, J. K. and Dartnall, H. Visual pigments of rods and cones in a human retina. *The Journal of physiology*, 298(1):501–511, 1980.

Braje, W. L. Illumination encoding in face recognition: effect of position shift. *Journal of Vision*, 3(2):4–4, 03 2003. ISSN 1534-7362. doi: 10.1167/3.2.4.

Braje, W. L., Kersten, D., Tarr, M. J., and Troje, N. F. Illumination effects in face recognition. *Psychobiology*, 26(4):371–380, 1998. ISSN 08896313. doi: 10.3758/BF03330623.

Bruce, V. and Young, A. Understanding face recognition. *British Journal of Psychology*, 77(3):305–327, 1986. ISSN 2044-8295. doi: 10.1111/j.2044-8295.1986.tb02199.x.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., and Linney, A. Sex discrimination: How do we tell the difference between male and female faces? *Perception*, 22(2):131–152, 1993. doi: 10.1068/p220131. PMID: 8474840.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., and Miller, P. Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4):339–360, 1999. ISSN 1939-2192. doi: 10.1037/1076-898X.5.4.339.

Burton, A. M., Bruce, V., and Johnston, R. A. Understanding face recognition with an interactive activation model. *British Journal of Psychology*, 81(3):361–380, 1990. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

Burton, A. M., Wilson, S., Cowan, M., and Bruce, V. Face Recognition in Poor-Quality Video: Evidence From Security Surveillance. *Psychological Science*, 10(3):243–248, may 1999. ISSN 0956-7976. doi: 10.1111/1467-9280.00144.

Burton, A. M., White, D., and McNeill, A. The Glasgow Face Matching Test. *Behavior Research Methods*, 42(1):286–291, 2010. ISSN 1554-351X. doi: 10.3758/BRM.42.1.286.

Burton, M. A. Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, 66(8):1467–1485, 2013. ISSN 17470218. doi: 10.1080/17470218.2013.800125.

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

Calder, A. J. and Young, A. W. Understanding the recognition of facial identity and facial expression. *Nature reviews. Neuroscience*, 6(8):641–51, 2005. ISSN 1471-003X. doi: 10.1038/nrn1724.

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

Carey, S. Becoming a face expert. *Phil. Trans. R. Soc. Lond. B*, 335(1273): 95–103, 1992.

Chalmers, A. and Debattista, K. Hdr video past, present and future: A perspective. *Signal Processing: Image Communication*, 54:49–55, 2017.

Cholkar, K., Dasari, S. R., Pal, D., Mitra, A. K., *et al.* Eye: anatomy, physiology and barriers to drug delivery. *Ocular Transporters and Receptors: Their Role in Drug Delivery*, page 1, 2013.

CIE. Proceedings, International Congress on Illumination. Technical report, Commission Internationale de l'Éclairage, 1931.

Cousineau, D. Confidence intervals in within-subject designs: A simpler solution to loftus and masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1):42–45, 2005. doi: 10.20982/tqmp.01.1.p042.

Cowan, N. 2.03 - sensory memory. In Byrne, J. H., editor, *Learning and Memory: A Comprehensive Reference*, pages 23 – 32. Academic Press, Oxford, 2008. ISBN 978-0-12-370509-9. doi: https://doi.org/10.1016/B978-012370509-9.00172-8.

Damian, M. F. Does variability in human performance outweigh imprecision in response devices such as computer keyboards? *Behavior Research Methods*, 42(1):205–211, 2010.

Debattista, K. Application-specific tone mapping via genetic programming. *Computer Graphics Forum*, 37(1):439–450, 2018. ISSN 14678659. doi: 10.1111/cgf.13307.

Debattista, K., Bashford-Rogers, T., Selmanović, E., Mukherjee, R., and Chalmers, A. Optimal exposure compression for high dynamic range content. *The Visual Computer*, 31(6-8):1089–1099, 2015. ISSN 0178-2789. doi: 10.1007/s00371-015-1121-z.

Debevec, P. *A median cut algorithm for light probe sampling*, pages 1–3. 2008.

Debevec, P. E. and Malik, J. Recovering high dynamic range images. In *proceeding of the SPIE: Image Sensors*, volume 3965, pages 392–401, 1997.

Drago, F., Myszkowski, K., Annen, T., and Chiba, N. *Adaptive logarithmic mapping for displaying high contrast scenes*, volume 22, pages 419–426. Wiley Online Library, 2003.

DSC-Laboratories. Xyla 21, dsc labs. `http://dsclabs.com/product/xyla-21/`. (Date Accessed: 6 January 2020).

Duchaine, B. and Nakayama, K. The cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44 (4):576–585, 2006.

Duchaine, B., Germine, L., and Nakayama, K. Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive neuropsychology*, 24(4):419–430, 2007.

Dufaux, F., Le Callet, P., Mantiuk, R., and Mrak, M. *High Dynamic Range Video: From Acquisition, to Display and Applications*. Academic Press, 2016.

Ebner, N. C., Riediger, M., and Lindenberger, U. FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1):351–362, feb 2010. ISSN 1554-351X. doi: 10.3758/BRM.42.1.351.

Eilertsen, G. *The high dynamic range imaging pipeline*. Linköping Studies in Science and Technology. Dissertations. Linköping University Electronic Press, 2018. ISBN 9789176853023.

Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J. Evaluation of tone mapping operators for HDR-video. *Computer Graphics Forum*, 32(7):275–284, 2013. ISSN 01677055. doi: 10.1111/cgf.12235.

Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R. K., and Unger, J. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36(6):178, 2017.

Eimer, M. Effects of face inversion on the structural encoding and recognition of faces: Evidence from event-related brain potentials. *Cognitive Brain Research*, 10(1-2):145–158, 2000.

Ellis, H. D. and Florence, M. Bodamer's (1947) paper on prosopagnosia. *Cognitive Neuropsychology*, 7(2):81–105, 1990. doi: 10.1080/02643299008253437.

Ellis, H. D., Quayle, A. H., and Young, A. W. The emotional impact of faces (but not names): Face specific changes in skin conductance responses to familiar and unfamiliar people. *Current Psychology*, 1999. ISSN 10461310. doi: 10.4324/9781351300247-6.

Endo, Y., Kanamori, Y., and Mitani, J. Deep reverse tone mapping. *ACM Transactions on Graphics (Proc. of SIGGRAPH ASIA 2017)*, 36(6), Nov. 2017.

Eppig, T. *Luminance, Definition*, pages 1–1. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. ISBN 978-3-642-35951-4. doi: 10.1007/978-3-642-35951-4_641-1.

European Commission. Health effects of artificial light. Technical report, European Commission's Scientific Committee on Emerging and Newly Identified Health Risks, March 2012.

Fairchild, M. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2013. ISBN 9781118653104.

Fairchild, M. D. A color scientist looks at video. In *3rd International Workshop on Video Processing and Quality Metrics (VPQM), Scottsdale, Invited Paper*, volume 1, 2007a.

Fairchild, M. D. The hdr photographic survey. In *Color and imaging conference*, volume 2007, pages 233–238. Society for Imaging Science and Technology, 2007b.

Fan, M., Lee, D.-H., Kim, S.-W., and Ko, S.-J. An optimization framework for inverse tone mapping using a single low dynamic range image. *Signal Processing: Image Communication*, 78:274 – 283, 2019. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2019.07.009.

Fattal, R., Lischinski, D., and Werman, M. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(July 2002):249–256, 2002. ISSN 07300301. doi: 10.1145/566654.566573.

Favelle, S., Hill, H., and Claes, P. About face: Matching unfamiliar faces across rotations of view and lighting. *i-Perception*, 8(6):2041669517744221, 2017. doi: 10.1177/2041669517744221.

Fei, P., Yu, Z., Wang, X., Lu, P. J., Fu, Y., He, Z., Xiong, J., and Huang, Y. High dynamic range optical projection tomography (hdr-opt). *Optics Express*, 20(8):8824–8836, 2012.

Ferwerda, J. A., Pattanaik, S. N., Shirley, P., and Greenberg, D. P. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, pages 249–258, New York, New York, USA, 1996. ACM Press. ISBN 0897917464. doi: 10.1145/237170.237262.

FISWG. Facial Identification Scientific Working Group. `http://fiswg.org/`. (Date Accessed: 20 February 2020).

Forrester, J., Dick, A., McMenamin, P., Roberts, F., and Eric Pearlman, B. *The Eye: Basic Sciences in Practice*. ClinicalKey 2012. Elsevier Health Sciences, 2015. ISBN 9780702055546.

Freeman, T. E., Loschky, L. C., and Hansen, B. C. Scene masking is affected by trial blank-screen luminance. *Signal Processing: Image Communication*, 39(August):319–327, nov 2015. ISSN 09235965. doi: 10.1016/j.image.2015.04.004.

Fussey, P. and Murray, D. Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology. University of essex human rights centre, University of Essex, July 2019.

Fysh, M. C. and Bindemann, M. Forensic face matching: A review. In Bindemann, M. and Megreya, A. M., editors, *Face processing: Systems, Disorders and Cultural Differences*, pages 1–20. New York: Nova Science Publishing Inc., 2017.

Fysh, M. C. and Bindemann, M. The Kent Face Matching Test. *British Journal of Psychology*, 109(2):219–231, 2018. ISSN 20448295. doi: 10.1111/bjop.12260.

Gainotti, G. Not all patients labeled as "prosopagnosia" have a real prosopagnosia. *Journal of Clinical and Experimental Neuropsychology*, 32(7):763–766, 2010.

Gegenfurtner, K. R. and Ennis, R. Fundamentals of color vision ii: high order color processing. In Elliot, A. J., Fairchild, M. D., and Franklin, A., editors, *Handbook of Color Psychology*, pages 70–109. Cambridge University Press, 2015. ISBN 9781107337930. Cambridge Books Online.

Ghuman, A. S., Brunet, N. M., Li, Y., Konecky, R. O., Pyles, J. A., Walls, S. A., Destefino, V., Wang, W., and Richardson, R. M. Dynamic encoding of face information in the human fusiform gyrus. *Nature Communications*, 2014. ISSN 20411723. doi: 10.1038/ncomms6672.

Goffaux, V. and Greenwood, J. A. The orientation selectivity of face identification. *Scientific Reports*, 2016. ISSN 20452322. doi: 10.1038/srep34204.

Gonzalez, R. C. and Woods, R. E. *Digital Image Processing (4th Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2018. ISBN 9780133356724.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

Great Britain, H. O. Surveillance camera code of practice. 2013.

Grossberg, M. D. and Nayar, S. K. What is the space of camera response functions? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. doi: 10.1109/cvpr.2003.1211522.

Grother, P., Ngan, M., and Hanaoka, K. Part 2: : Identification, nistir 8271. In *Face Recognition Vendor Test (FRVT)*, pages 1–185. US Department of Commerce, National Institute of Standards and Technology, sep 2019a. doi: 10.6028/nist.Ir.8271.

Grother, P., Ngan, M., and Hanaoka, K. Part 3: Demographic effects, nistir 8280. In *Face Recognition Vendor Test (FRVT)*, pages 1–82. US Department of Commerce, National Institute of Standards and Technology, dec 2019b. doi: 10.6028/NIST.IR.8280.

Hancock, P. J., Bruce, V., and Burton, A. M. Recognition of unfamiliar faces. *Trends in cognitive sciences*, 4(9):330–337, 2000.

Hatchett, J., Toffoli, D., Melo, M., Bessa, M., Debattista, K., and Chalmers, A. Displaying detail in bright environments: A 10,000 nit display and its evaluation. *Signal Processing: Image Communication*, 76:125–134, 2019.

Hattar, S., Liao, H.-W., Takao, M., Berson, D. M., and Yau, K.-W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science*, 295(5557):1065–1070, 2002.

Haxby, J. V. and Gobbini, M. I. Distributed Neural Systems for Face Perception. In *Oxford Handbook of Face Perception*. Oxford University Press, 2012. ISBN 9780191743672. doi: 10.1093/oxfordhb/9780199559053.013.0006.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. The distributed human neural system for face perception. In *Trends in Cognitive Sciences*, volume 4, pages 223–233, 2000. ISBN 1364-6613 (Print). doi: 10.1016/S1364-6613(00)01482-0.

Hermens, F. and Zdravković, S. Information extraction from shadowed regions in images: An eye movement study. *Vision research*, 113(Pt A):87–96, 2015. ISSN 1878-5646. doi: 10.1016/j.visres.2015.05.019.

Herzmann, G., Schweinberger, S. R., Sommer, W., and Jentzsch, I. What's special about personally familiar faces? a multimodal approach. *Psychophysiology*, 41(5):688–701, 2004.

Hietanen, J., Perrett, D., Oram, M., Benson, P., and Dittrich, W. The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Experimental Brain Research*, 89(1):157–171, 1992.

Hill, H. and Bruce, V. The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4):986, 1996.

Hole, G. and Bourne, V.-J. *Face processing : psychological, neuropsychological, and applied perspectives*. Oxford University Press, May 2010.

Ige, E. O., Debattista, K., and Chalmers, A. Towards hdr based facial expression recognition under complex lighting. In *Proceedings of the 33rd Computer Graphics International*, pages 49–52. ACM, 2016.

ISO/IEC 18477-1:2015. Information technology — Scalable compression and coding of continuous-tone still images. Standard, International Organization for Standardization, Geneva, CH, June 2015.

ISO/IEC 29199-2:2009. JPEG XR image coding system. Standard, International Organization for Standardization, Geneva, CH, Aug. 2009.

ITU-R BT.2022. General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays. Recommendation, Geneva, CH, August 2012.

ITU-R BT.2100. Image parameter values for high dynamic range television for use in production and international programme exchange. Recommendation, Geneva, CH, July 2019.

ITU-R BT.709-6. Parameter values for the hdtv standards for production and international programme exchange. Recommendation, International Telecommunication Union, Geneva, CH, June 2015.

Jack, T. and Holly, R. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, 1993. ISSN 02721716. doi: 10.1109/38.252554.

Jain, A. K., Klare, B., and Park, U. Face matching and retrieval in forensics applications. *IEEE Multimedia*, 19(1):20–27, 2012. ISSN 1070986X. doi: 10.1109/MMUL.2012.4.

Jeantet, C., Caharel, S., Schwan, R., Lighezzolo-Alnot, J., and Laprevote, V. Factors influencing spatial frequency extraction in faces: A review. *Neuroscience & Biobehavioral Reviews*, 93(December 2017):123–138, oct 2018. ISSN 01497634. doi: 10.1016/j.neubiorev.2018.03.006.

Jenkins, R., White, D., Van Montfort, X., and Burton, A. M. Variability in photos of the same face. *Cognition*, 121(3):313–323, 2011.

Johnston, A., Hill, H., and Carman, N. Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21:365–75, 1991. reprinted in Perception, 42(11): 1227-1237.

Johnston, R. A. and Bindemann, M. Introduction to forensic face matching. *Applied Cognitive Psychology*, 27(6):697–699, 2013.

Johnston, R. A. and Edmonds, A. J. Familiar and unfamiliar face recognition: A review. *Memory*, 17(5):577–596, 2009.

Kainz, F., Bogart, R., and Stanczyk, P. Technical introduction to openexr. *Industrial light and magic*, page 21, 2009.

Karr, B., Chalmers, A., and Debattista, K. High dynamic range digital imaging of spacecraft. In *High Dynamic Range Video*, pages 519–547. Elsevier, 2016.

Kemp, R. I., Caon, A., Howard, M., and Brooks, K. R. Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology*, 30(4):622–627, 2016.

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., and Grueter, M. First report of prevalence of non-syndromic hereditary

prosopagnosia (hpa). *American Journal of Medical Genetics Part A*, 140 (15):1617–1622, 2006.

Kennerknecht, I., Ho, N. Y., and Wong, V. C. Prevalence of hereditary prosopagnosia (hpa) in hong kong chinese population. *American Journal of Medical Genetics Part A*, 146(22):2863–2870, 2008.

Khan, I. R., Rahardja, S., Khan, M. M., Movania, M. M., and Abed, F. A tone-mapping technique based on histogram using a sensitivity model of the human visual system. *IEEE Transactions on Industrial Electronics*, 65(4): 3469–3479, 2017.

Kim, M. H. and Kautz, J. Characterization for high dynamic range imaging. *Computer Graphics Forum*, 27(2):691–697, 2008. ISSN 01677055. doi: 10.1111/j.1467-8659.2008.01167.x.

Konica Minolta. `https://www.konicaminolta.com/instruments`. (Date Accessed: 27 February 2020).

Korshunov, P., Bernardo, M. V., Pinheiro, A. M., and Ebrahimi, T. Impact of Tone-mapping Algorithms on Subjective and Objective Face Recognition in HDR Images. *Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia - CrowdMM '15*, pages 39–44, 2015. doi: 10.1145/2810188.2810195.

Krasula, L., Narwaria, M., Fliegel, K., and Le Callet, P. Preference of Experience in Image Tone-Mapping: Dataset and Framework for Objective Measures Comparison. *IEEE Journal on Selected Topics in Signal Processing*, 2017. ISSN 19324553. doi: 10.1109/JSTSP.2016.2637168.

Krawczyk, G., Myszkowski, K., and Seidel, H.-P. Computational model of lightness perception in high dynamic range imaging. In *Electronic Imaging 2006*, pages 605708–605708. International Society for Optics and Photonics, 2006.

Kreutzer, J. S., DeLuca, J., and Caplan, B., editors. *Benton Face Recognition Test*, pages 392–392. Springer New York, New York, NY, 2011. ISBN 978-0-387-79948-3. doi: 10.1007/978-0-387-79948-3_4510.

Kronander, J., Gustavson, S., Bonnet, G., Ynnerman, A., and Unger, J. A unified framework for multi-sensor hdr video reconstruction. *Signal Processing: Image Communication*, 29(2):203 – 215, 2014. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2013.08.018. Special Issue on Advances in High Dynamic Range Video Research.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8):1377–1388, 2010. ISSN 0269-9931. doi: 10.1080/02699930903485076.

Larson, G. W., Rushmeier, H., and Piatko, C. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997. ISSN 10772626. doi: 10.1109/2945.646233.

Lasance, C. and Poppe, A. *Thermal Management for LED Applications*. Solid State Lighting Technology and Application Series. Springer New York, 2013. ISBN 9781461450917.

Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H. Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph.*, 24(3):640–648, jul 2005. ISSN 0730-0301. doi: 10.1145/1073204.1073242.

Lee, W.-J., Wilkinson, C., Memon, A., and Houston, K. Matching unfamiliar faces from poor quality closed-circuit television (cctv) footage. *Axis: The Online Journal of CAHId*, 1(1):19–28, 2009.

Logie, R. H., Baddeley, A. D., and Woodhead, M. M. Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, 1(1):53–69, 1987.

Longmore, C. A., Liu, C. H., and Young, A. W. Learning Faces From Photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 2008. ISSN 00961523. doi: 10.1037/0096-1523.34.1.77.

Longmore, C. A., Liu, C. H., and Young, A. W. The importance of internal facial features in learning new faces. *The Quarterly Journal of Experimental Psychology*, 68(2):249–260, 2015.

Luck, S. *An Introduction to the Event-Related Potential Technique*. A Bradford Book. MIT Press, 2014. ISBN 9780262525855.

Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J. Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE international conference on automatic face and gesture recognition*, pages 200–205. IEEE, 1998.

Ma, D. S., Correll, J., and Wittenbrink, B. The chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4): 1122–1135, Dec 2015a. ISSN 1554-3528. doi: 10.3758/s13428-014-0532-5.

Ma, K., Yeganeh, H., Zeng, K., and Wang, Z. High dynamic range image compression by optimizing tone mapped image quality index. *IEEE Transactions on Image Processing*, 24(10):3086–3097, 2015b.

Mai, Z., Doutre, C., Nasiopoulos, P., and Ward, R. K. *Subjective evaluation of tone-mapping methods on 3D images*, pages 1–6. 2011.

Majumdar, T. and Sinha, P. Photographs of the human face and broken projective symmetry. In *Bulletin du Service de Documentation Generale*, volume 79, pages 262–68. Organization Internationale de Police Criminelle (INTERPOL), 1992.

Mann, S. and Picard, R. W. On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures. *Proceedings of IS&T*, pages 442–448, 1995. ISSN 1098-6596. doi: 10.1.1.49.1521.

Mantiuk, R., Efremov, A., Myszkowski, K., and Seidel, H.-P. Backward compatible high dynamic range mpeg video compression. *ACM Transactions on Graphics (TOG)*, 25(3):713–723, 2006a.

Mantiuk, R., Myszkowski, K., and Seidel, H.-P. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP)*, 3(3):286–308, 2006b.

Mantiuk, R., Krawczyk, G., Mantiuk, R., and Seidel, H.-P. High-dynamic range imaging pipeline: perception-motivated representation of visual content. In

Rogowitz, B. E., Pappas, T. N., and Daly, S. J., editors, *Human Vision and Electronic Imaging XII*, volume 6492, pages 382 – 393. International Society for Optics and Photonics, SPIE, 2007. doi: 10.1117/12.713526.

Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):40, 2011.

Mantiuk, R. K., Myszkowski, K., and Seidel, H.-P. High dynamic range imaging. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2016.

Marchessoux, C., de Paepe, L., Vanovermeire, O., and Albani, L. Clinical evaluation of a medical high dynamic range display. *Medical Physics*, 43(7): 4023–4031, 2016. doi: 10.1118/1.4953187.

Marnerides, D., Bashford-Rogers, T., Hatchett, J., and Debattista, K. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Computer Graphics Forum*, 2018. ISSN 14678659. doi: 10.1111/cgf.13340.

Megreya, A. M. and Burton, A. M. Matching faces to photographs: poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, 14(4):364, 2008.

Megreya, A. M., Bindemann, M., and Havard, C. Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica*, 137 (1):83 – 89, 2011. ISSN 0001-6918.

Megreya, A. M., Sandford, A., and Burton, A. M. Matching face images taken on the same day or months apart: the limitations of photo id. *Applied Cognitive Psychology*, 27(6):700–706, 2013. doi: 10.1002/acp.2965.

Mitsunaga, T. and Nayar, S. K. Radiometric Self Calibration. *Proc. of CVPR*, 1:374–380, 1999. ISSN 10636919. doi: 10.1109/CVPR.1999.786966.

Myszkowski, K., Mantiuk, R., and Krawczyk, G. High Dynamic Range Video. page 158, 2008.

Nafchi, H. Z., Shahkolaei, A., Moghaddam, R. F., and Cheriet, M. Fsitm: A feature similarity index for tone-mapped images. *IEEE Signal Processing Letters*, 22(8):1026–1029, 2014.

Norell, K., Läthén, K. B., Bergström, P., Rice, A., Natu, V., and O'Toole, A. The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons. *Journal of Forensic Sciences*, 60(2):331–340, mar 2015. ISSN 00221198. doi: 10.1111/1556-4029.12660.

Noyes, E. and Jenkins, R. Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, 165:97–104, 2017.

Parraga, C. A., Otazu, X., *et al.* Which tone-mapping operator is the best? a comparative study of perceptual quality. *JOSA A*, 35(4):626–638, 2018.

Pattanaik, S. N., Ferwerda, J. A., Fairchild, M. D., and Greenberg, D. P. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, pages 287–298, New York, NY, USA, 1998. ACM. ISBN 0-89791-999-8. doi: 10.1145/280814.280922.

Peli, E. Contrast in complex images. *JOSA A*, 7(10):2032–2040, 1990.

Peterson, M. F. and Eckstein, M. P. Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48):E3314–E3323, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1214269109.

Phillips, P. J. and O'toole, A. J. Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, 32(1):74–85, 2014.

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J., and O'Toole, A. J. An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception (TAP)*, 8(2):1–11, 2011.

Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D., Dunlop, J., Lui, Y. M., Sahibzada, H., and Weimer, S. The good, the bad, and the ugly face challenge problem. *Image and Vision Computing*, 30(3):177–185, 2012.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C. D., Chellappa, R., White, D., and O'Toole, A. J. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115(24):6171–6176, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1721355115.

Poynton, C. A. Rehabilitation of gamma. In *Human Vision and Electronic Imaging III*, volume 3299, pages 232–249. International Society for Optics and Photonics, 1998.

Ramachandran, V. S. Perception of shape from shading. *Nature*, 331(6152): 163–166, 1988.

Reinhard, E. and Devlin, K. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11 (1):13–24, 2005.

Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. Photographic tone reproduction for digital images. *ACM Transactions on Graphics (TOG)*, 21 (3):267–276, 2002.

Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

Remington, L. *Clinical Anatomy of the Visual System*. Elsevier Health Sciences, 2011. ISBN 9781455727773.

Rice, A., Phillips, P. J., Natu, V., An, X., and O'Toole, A. J. Unaware person recognition from the body when face identification fails. *Psychological Science*, 24(11):2235–2243, 2013.

Ritchie, K. L. and Burton, A. M. Learning faces from variability. *Quarterly Journal of Experimental Psychology*, 70(5):897–905, 2017. ISSN 17470226. doi: 10.1080/17470218.2015.1136656.

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., and Burton, A. M. Viewers base estimates of face matching accuracy on their

own familiarity: Explaining the photo-id paradox. *Cognition*, 141:161 – 169, 2015. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2015.05.002.

Ritchie, K. L., Kramer, R. S., and Burton, A. M. What makes a face photo a 'good likeness'? *Cognition*, 170:1–8, 2018.

Robertson, M., Borman, S., and Stevenson, R. Dynamic range improvement through multiple exposures. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, volume 3, pages 159–163, 1999. ISBN 0-7803-5467-2. doi: 10.1109/ICIP.1999.817091.

Robertson, M. A. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219, apr 2003. ISSN 1017-9909. doi: 10.1117/1.1557695.

Robins, M. and Bean, H. Glare reduction system for image capture devices, Oct. 2 2003. US Patent App. 10/112,339.

Russell, R., Sinha, P., Biederman, I., and Nederhouser, M. Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, 35(6):749–759, 2006. ISSN 03010066. doi: 10.1068/p5490.

Russell, R., Duchaine, B., and Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, 16(2): 252–257, 2009.

Salih, Y., Bt. Md-Esa, W., Malik, A. S., and Saad, N. Tone mapping of HDR images: A review. In *ICIAS 2012 - 2012 4th International Conference on Intelligent and Advanced Systems: A Conference of World Engineering, Science and Technology Congress (ESTCON) - Conference Proceedings*, volume 1, pages 368–373, 2012. ISBN 9781457719677. doi: 10.1109/ICIAS.2012.6306220.

Schweinberger, S. R., Huddy, V., and Burton, A. M. N250r: A face-selective brain response to stimulus repetitions. *NeuroReport*, 2004. ISSN 09594965. doi: 10.1097/01.wnr.0000131675.00319.42.

Seetzen, H., Whitehead, L. A., and Ward, G. 54.2: A high dynamic range display using low and high resolution modulators. *SID Symposium Digest of Technical Papers*, 34(1):1450–1453, 2003. ISSN 2168-0159. doi: 10.1889/1.1832558.

Shah, P., Gaule, A., Sowden, S., Bird, G., and Cook, R. The 20-item prosopagnosia index (pi20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, 2(6):140343, 2015. doi: 10.1098/rsos.140343.

Shevell, S. K. *The science of color*. Elsevier, 2003.

SONA. Sona Systems, Ltd. Experiment Management System. `https://warwick.sonasystems.com/`. (Date Accessed: 7 February 2020).

Spaun, N. A. Facial comparisons by subject matter experts: Their role in biometrics and their training. In Tistarelli, M. and Nixon, M. S., editors, *Advances in Biometrics*, pages 161–168, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-01793-3.

Spaun, N. A. Face recognition in forensic science. In Li, S. Z. and Jain, A. K., editors, *Handbook of Face Recognition*, pages 655–670. Springer London, London, 2011. ISBN 978-0-85729-932-1. doi: 10.1007/978-0-85729-932-1_26.

Stephen, I. D. and Perrett, D. I. Color and face perception. In Elliot, A. J., Fairchild, M. D., and Franklin, A., editors, *Handbook of Color Psychology*, pages 585–602. Cambridge University Press, Cambridge, 2016. ISBN 9781107337930. doi: 10.1017/CBO9781107337930.029.

Steyn, M., Pretorius, M., Briers, N., Bacci, N., Johnson, A., and Houlton, T. Forensic facial comparison in south africa: State of the science. *Forensic Science International*, 287:190 – 194, 2018. ISSN 0379-0738. doi: 10.1016/j.forsciint.2018.04.006.

Stockman, A. and Brainard, D. H. Fundamentals of color vision i: color processing in the eye. In Elliot, A. J., Fairchild, M. D., and Franklin, A., editors, *Handbook of Color Psychology*, pages 27–69. Cambridge University Press, 2015. ISBN 9781107337930. Cambridge Books Online.

Stockman, A. and Sharpe, L. T. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., and Heagins, T. B. The mr2: A multi-racial, mega-resolution database of facial stimuli. *Behavior research methods*, pages 1–8, 2015.

Suma, R., Stavropoulou, G., Stathopoulou, E. K., van Gool, L., Georgopoulos, A., and Chalmers, A. Evaluation of the effectiveness of hdr tone-mapping operators for photogrammetric applications. *Virtual Archaeology Review*, 7 (15):54–66, 2016.

Suma, R., Debattista, K., Watson, D. G., Blagrove, E., and Chalmers, A. Subjective evaluation of high dynamic range imaging for face matching. *IEEE Transactions on Emerging Topics in Computing*, December 2019.

Tarr, M. J., Georghiades, A. S., and Jackson, C. D. Identifying faces across variations in lighting: Psychophysics and computation. *ACM Transactions on Applied Perception*, 5(2):1–25, 2008. ISSN 15443558. doi: 10.1145/1279920.1279924.

Towler, A., Kemp, R., and White, D. Unfamiliar face matching systems in applied settings. *Face Processing: Systems, Disorders and Cultural Difference*, 2017a.

Towler, A., White, D., and Kemp, R. I. Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, 23(1):47–58, 2017b. ISSN 1939-2192. doi: 10.1037/xap0000108.

Trentacoste, M., Heidrich, W., Whitehead, L., Seetzen, H., and Ward, G. Photometric image processing for high dynamic range displays. *Journal of Visual Communication and Image Representation*, 18(5):439–451, 2007.

Treuting, P. M. and Dintzis, S. M. *Comparative Anatomy and Histology: A Mouse and Human Atlas (Expert Consult)*. Academic Press, 2011.

UK Ethinicity, 2011. UK Cabinet Office: Ethnicity categories and the 2011 census. `https://www.ethnicity-facts-figures.service.gov.uk/`

ethnicity-in-the-uk/ethnic-groups-and-data-collected, 2011. (Accessed Feb 15th, 2020).

Valentine, T. and Davis, J. P. *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV*. John Wiley & Sons, 2015.

van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion (Washington, D.C.)*, 11(4):907–920, 2011. ISSN 1528-3542. doi: 10.1037/a0023853.

Čadík, M., Wimmer, M., Neumann, L., and Artusi, A. Image attributes and quality for evaluation of tone mapping operators. In *National Taiwan University*, pages 35–44. Press, 2006.

Vinegoni, C., Swisher, C. L., Feruglio, P. F., Giedt, R., Rousso, D., Stapleton, S., and Weissleder, R. Real-time high dynamic range laser scanning microscopy. *Nature communications*, 7(1):1–13, 2016.

Wallace, G. K. The jpeg still picture compression standard. *Communications of the ACM*, pages 30–44, 1991.

Walraven, P. L. and Walraven, P. L. Fundamental chromaticity diagram with physiological axes. *CIE Technical Report*, page 24, 1999.

Wang, C., Zhao, Y., and Wang, R. Deep inverse tone mapping for compressed images. *IEEE Access*, 7:74558–74569, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2920951.

Wang, L., Wei, L.-Y., Zhou, K., Guo, B., and Shum, H.-Y. High dynamic range image hallucination. In *ACM SIGGRAPH 2007 Sketches*, SIGGRAPH '07, New York, NY, USA, 2007. ACM. ISBN 978-1-4503-4726-6. doi: 10.1145/1278780.1278867.

Ward, G. Real pixels. *Graphics Gems II*, pages 80–83, 1991.

Ward, G. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.

Westland, S., Owens, H., Cheung, V., and Paterson-Stephens, I. Model of luminance contrast-sensitivity function for application to image assessment. *Color Research & Application*, 31(4):315–319, 2006.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., and Burton, A. M. Passport officers' errors in face matching. *PLoS ONE*, 9(8), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0103510.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., and O'Toole, A. J. Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814):20151292, 2015.

White, D., Norell, K., Phillips, P. J., and O'Toole, A. J. *Human Factors in Forensic Face Identification*, pages 195–218. Springer International Publishing, Cham, 2017. ISBN 978-3-319-50673-9. doi: 10.1007/978-3-319-50673-9_9.

Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., and Brooke, J. Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, 77(4):269 – 291, 2000. ISSN 0022-0965. doi: https://doi.org/10.1006/jecp.1999.2554.

Wiseman, Y. The still image lossy compression standard-jpeg. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 295–305. IGI Global, 2015.

Wu, G., Masia, B., Jarabo, A., Zhang, Y., Wang, L., Dai, Q., Chai, T., and Liu, Y. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, Oct 2017. ISSN 1941-0484. doi: 10.1109/JSTSP.2017.2747126.

Yeganeh, H. and Wang, Z. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, 2012.

Young, A. W. and Burton, A. M. Recognizing Faces. *Current Directions in Psychological Science*, 26(3):212–217, jun 2017. ISSN 0963-7214. doi: 10.1177/0963721416688114.

Yu, F. and Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.