Subjective Evaluation of High Dynamic Range Imaging for Face Matching

Rossella Suma, Kurt Debattista, Derrick Watson, Elisabeth Blagrove, and Alan Chalmers

***Abstract*—Human facial recognition in the context of surveillance, forensics and photo-ID verification is a task for which accuracy is critical. Quite often limitations in the overall quality of facial images reduces individuals' ability in taking decisions regarding a person's identity. To verify the suitability of advanced imaging techniques to improve individuals' performance in face matching we investigate how High Dynamic Range (HDR) imaging compares with traditional low (or standard) dynamic range (LDR) imaging in a facial recognition task. An HDR face dataset with five different lighting conditions is created. Subsequently, this dataset is used in a controlled experiment (N=40) to measure performance and accuracy of human participants when identifying faces in HDR vs LDR. Results demonstrate that face matching accuracy and reaction time are improved significantly by HDR imaging. This work demonstrates scope for realistic image reproduction and delivery in face matching tasks and suggests that security systems could benefit from the adoption of HDR imaging techniques.**

***Index Terms*—imaging, High Dynamic Range, subjective face matching, face dataset.**

I. Introduction

FACIAL recognition is an activity routinely carried out by billions of 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 [1] with significant security implications. Although increasing technological support is provided, the ultimate decision making relies on the human observer [2]. Further detail and information in the stimuli presented can increase accuracy in the process.

As emphasised by the detailed review on the topic of forensic face matching by Fysh and Bindemann [3], possible impostors or mismatches constitute an increasing security concern. A consistent amount of research on perceptual psychology [4], [5] 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 [5], [6]. 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 [7].

The goal of the present research is to develop and assess an advancement in the technology adopted for such a highly error-prone task. Jenkins *et al.* [8] have underlined the impact of lighting, camera, and lens characteristics. The work of Bindemann *et al.* [9] on the effect of resolution/pixelation, the analysis written by Norrel *et al.* [6] on the impact of image quality and the work of Liu *et al.* [10] on the impact of several post-processing algorithms to weaken shading effects have tried to tackle the impact that specific image attributes have on recognition. However, no study has yet, to the best of our knowledge, explored the impact of dynamic range. In particular, the lack or presence of illumination across both images in a face pair is one of the key elements affecting performance [8], [11] and it is a characteristic strongly related to the dynamic range of a scene.

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. In the current work, we investigate whether an alternative, High Dynamic Range (HDR) imaging, is capable of increasing facial matching performance when faces are portrayed under different lighting conditions. HDR's potential lies in its capability of capturing, storing and visualising all the colours and brightness levels visible to the human eye [12]. In this work, the potential of HDR to outperform LDR in facial matching is investigated.

The main contributions are:

- The creation of the Warwick HDR Face Dataset: a collection of calibrated HDR images of faces with five controlled lighting conditions;
- The evaluation of performance (reaction time) and accuracy in the context of facial image comparison when using HDR vs LDR stimuli;
- The demonstration of the intrinsic advantage of HDR for face matching tasks, when using digital images only.

The remainder of the paper provides a brief description of HDR imaging technology and an overview of the face recognition problem in general and within the context of surveillance and person identification in Section II. Section III frames the problem in more detail and highlights the existing gaps in the literature. Section IV presents the procedures for the creation of the Warwick HDR Face Dataset. Section V describes the HDR vs LDR perception experiment and a

Manuscript submitted September 2, 2019. Rossella Suma was partially supported by the Rabin Ezra Scholarship Trust.

R. Suma, K.Debattista and A. Chalmers are with Warwick Manufacturing Group (WMG), University of Warwick, CV4 7AL, Coventry, UK (email: r.suma@warwick.ac.uk; k.debattista@warwick.ac.uk; alan.chalmers@warwick.ac.uk).

E. Blagrove and D. Watson are with the Psychology Department, University of Warwick, CV4 7AL, Coventry, UK (email: e.l.blagrove@warwick.ac.uk; d.g.watson@warwick.ac.uk).

detailed review of the results is given in Section VI, followed by Discussion and Conclusion.

II. Background and related work

This section discusses the background and related work in HDR imaging and face recognition.

A. HDR technology

The dynamic range (DR) of an image is the difference between its brightest and darkest part. This is commonly measured as the logarithm (base 10) of the difference in luminance (cd/m^2). The human visual system is capable of adapting to a wide range of lighting (over four orders of magnitude in a single scene without adaptation and significantly more with adaptation [13]), yet existing display technology or cameras commonly in use are able to cope with just two or three orders of magnitude at once. HDR imaging technology provides a way to capture real-world content with higher accuracy than LDR imaging both in terms of luminance levels and colour rendition [12].

a) HDR image capture: The most commonly adopted method to create HDR image content using standard cameras is to apply a bracketing technique [13]. This method consists of having several pictures captured in a temporal sequence with different exposure times. Normally this is conducted using a tripod to stabilise the camera position. The sequential capture allows the camera sensor on each shot to collect, per pixel, a different number of photons each time. The end result is a sequence of images of the same scene ranging from under to overexposed. For each pixel, all the information regarding light and colour coming from each exposure is then merged it into a single HDR image. When merging the different exposures several weighting functions are available from the literature [13]. The majority of them favour the middle exposures as they contain more reliable information. This helps to minimise artefacts (e.g. saturated pixels, thermal noise). The resulting image is a function of the irradiance at each pixel.

b) HDR vs LDR image and tone-mapping: HDR imaging provides a more accurate representation of each colour channel using floating-point values as opposed to the traditional eight bits integers per channel per pixel in the LDR imaging case. It also offers a representation where values are a good approximation of photometric quantities [14] unlike the non-linear relation used in LDR imaging (i.e. gamma-offset-gain).

Tone-mapping operators (TMOs) are functions which reduce the information contained in an HDR image to something that is representable on LDR media (screen or paper). This operation clearly involves a loss when the dynamic range in the HDR image exceeds that which can be directly displayed. Depending on the TMO, often particular attention to preserving specific image features or characteristics will be given, while discarding others. There is a significant number of TMOs described in the literature [13], but no unanimous agreement on which operator is the best for all circumstances.

c) HDR display: The most desirable feature when visualising images, especially for scientific purposes, is to reproduce reality as accurately as possible. Some aspects in modern display technology exceed visual acuity, for example, in terms of refresh rate and spatial resolution [15], but reaching levels of brightness comparable to what is experienced in our daily life still remains a challenge.

Some of the most recent models of HDR screen prototypes can reach a peak luminance of 10,000 cd/m^2 as opposed to the 400 cd/m^2 maximum luminance of standard screens and also offer 12 bit resolution to drive the back panel intensity. At the time of writing some companies have begun producing consumer displays which are labelled as HDR and can reach a peak luminance of about 1,000 cd/m^2.

d) Applications of HDR imaging: Increasingly more attention has been directed towards HDR technology in recent years. Examples of applications exploiting it can be found in several fields. For example, HDR photography has been applied to smart-phones when dealing with unfavourable lighting conditions [16]. Medical imaging has seen an increased interest in HDR for applications such as MRI or ultrasound visualisation where it helps professionals to be more confident in their medical report due to the added image details [17], [18]. Further applications of HDR include feature detection and people tracking [19], as well as capturing rocket launches [20].

B. Face recognition

Although the common assumption is that adults are "face-experts", this is generally true only in the case of familiar faces [5]. 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) [21] and they become progressively less accurate as the degree of other-race contact in their daily experience decreases.

When encountering an unfamiliar face this task might be challenging and often a source of misidentification errors [5]. The difficulty is partially related to the capability of coping with a person's changes in appearance over a range of photos [22]. 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.

a) Factors affecting face recognition: 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 (i.e. from memory) tasks [4].

This class of experiments allows 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 [23] or configuration [4]), but also to the properties of the technology adopted to conduct the experiment.

As Young and Burton [5] have pointed out, sometimes the extent to which image variability influences performance, for example, due to lighting, camera and lens characteristics, is overlooked. Bindemann *et al.* [9] 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. 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.

Other factors impacting recognition are described by Noyes and Jenkins [24] in a study focusing on camera-to-subject distance and how it can hinder perceptual matching of unfamiliar faces and pointing out how actually governmental rules are diverse and seem to regard this element only marginally.

b) Facial stimuli datasets: A typical aspect of studying face matching depends on subject appropriate stimuli used 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 [25] and the recently released Kent Face Matching Test [26]. Unfortunately, frequently these stimuli set present technical limitations especially in terms of image resolution, lack of information regarding camera calibration, and colour accuracy. The literature also offers other datasets, 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 [5], [27].

Other attempts involve computer graphics generated stimuli, but this might raise the objection of drifting even further away from ecologically valid stimuli [28]. There is also a large availability of faces 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 [4], [29].

c) Facial identification: The need for standards and reliability in the face forensic context is clear as international (e.g. Facial Identification Scientific Working Group [30]) and national organisations (e.g. South Africa [31]) try to coordinate and deliberate on this. 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 [32]. 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 [32] for a detailed analysis). Usually, the only manipulation possible on the picture is related to luminance adjustments [31], showing clearly how a more accurate lighting reproduction could bring an advantage to the forensic practice.

Forensic facial reviewers are a category of experts often enrolled in the law enforcement [33] whose task is to perform quick and less rigorous identification. In these scenarios the risk of misidentification is higher because of the time constraints, therefore technological support can be advantageous especially for tackling data limits (e.g. illumination, viewpoint, image degradation, within-target variation) [3].

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 [6] 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 [6], therefore, the impact brought by advanced imaging could be valuable.

III. Motivation

The likelihood of determining whether distinct instances of images correspond to the same person 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 [34], and the demand for human examiners to perform face identification especially in the legal system is still extremely high [2].

This work aims at tackling one of the main problems often pointed out by the facial recognition literature: the accuracy in reproducing reality, in terms of image and display quality, for colour rendition and luminance [3], [6]. Furthermore, it evaluates whether this improved accuracy equates to better performance in unfamiliar face matching tasks, which could result in more secure and dependable data, and decision making with greater precision.

HDR imaging is perceptually closer to how humans perceive the world [12] and this work investigates whether HDR can outperform traditional imaging in face matching tasks. In order to conduct such an evaluation, the need for the creation of an HDR facial database was identified. Until now, the only two possible candidates for HDR facial databases were presented by Ige *et al.* [35] and Korshunov *et al.* [36]. They were not considered suitable for use in a controlled experiment. Ige *et al.* [35] due to the lack of data regarding camera calibration and Korshunov *et al.* [36] due to the uncontrolled environment and unspecified lighting conditions.

The first part of this paper (Section IV) introduces the creation of an HDR face stimuli dataset, the Warwick HDR Face Dataset, for use in digital forensics, psychological or visual perception studies. While not fully comprehensive, due to its size, it is prepared in a fully controlled environment both in terms of lighting and image capture.

In the second part of this work (Section V and VI), the Warwick HDR Face Dataset is used to conduct the HDR vs LDR experiment. This consists of the evaluation of performance (reaction time and accuracy) in a face matching task when HDR and/or LDR face stimuli are provided. Within the same experiment, the availability of faces stimuli portraits under different lighting conditions allows addressing the question related to variation in performance due to change in the illumination. [37]

IV. Warwick HDR Face Dataset

This section describes the creation and criteria adopted for the HDR imaging face stimuli dataset and the methodology used for its creation.

Fig. 1. 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.

A. Design

The Warwick HDR Face Dataset serves as a basis for conducting the face matching experiment in Section V. The aim of the dataset is to have a number of faces captured in HDR under diverse lighting conditions. In order to be able to assess human perceptual performance at matching faces under challenging lighting conditions a number of distinct lighting positions were chosen.

The five different illumination patterns chosen are schematically represented in Figure 2(a). The data collection has been performed so that light sources distances and intensity were standardised. 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 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

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. When performing a face matching task, the availability of two different images with the same lighting will allow to elicit face-related cognitive processing in the observer, rather than apply low-level image comparison techniques [4].

In future, the dataset will be expanded, in order to have an equal number of female participants to make it generalizable.

B. 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).

C. 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.

D. Materials

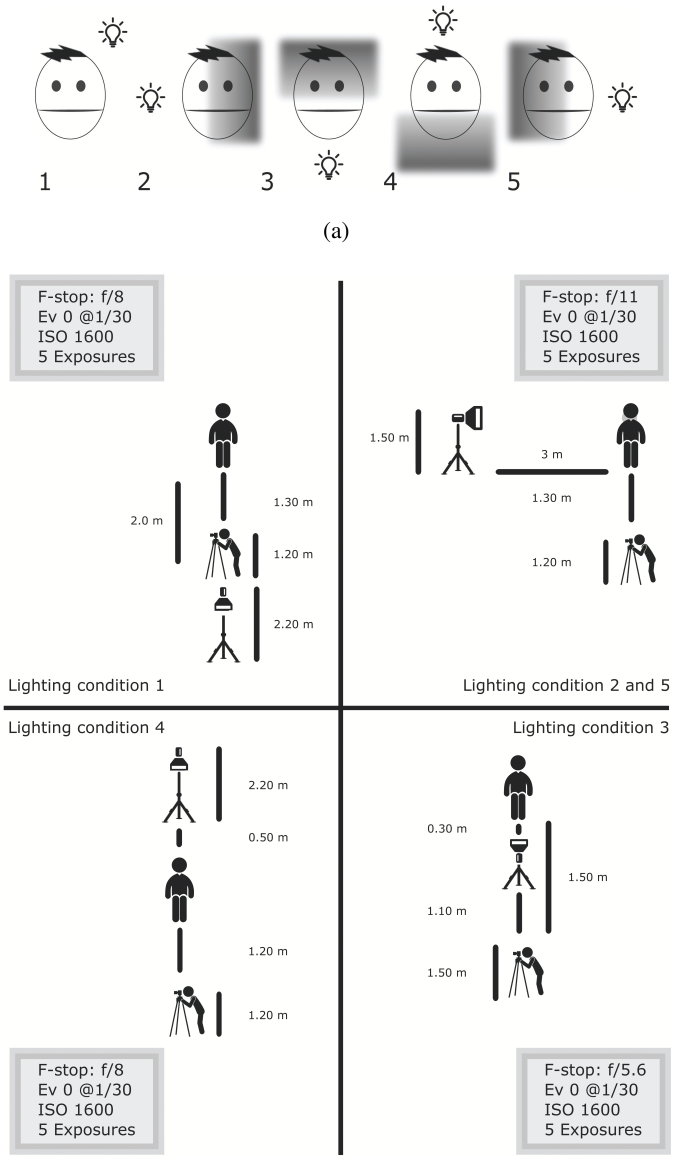
a) Lighting: For the five lighting conditions in the dataset all the distances between the participant and the light were carefully chosen in order to respect international safety standards and good practice so that there were absolutely no short or long term negative effects for the participants.

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 2(b) shows the details on all the specific distances.

b) Camera Setup: Obtaining an HDR image through a sequence of pictures with exposure bracketing (see Section II) requires the subject being photographed to remain still for a few seconds (max 6 sec.). 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. The parameters selected for each lighting setup are specified in Figure 2 (b).

The standardisation of distances and camera-subject relative positions were selected in order to minimise deformations and lens flare as well as maximising the sensor area where useful information (i.e the face) was located. 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).

c) Camera Calibration: The camera was fully 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 [38].



(b)

Fig. 2. Warwick HDR Face dataset. (a) 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. (b) Diagram with details of distances and camera settings adopted for the Warwick HDR Face Dataset capture. Left to right, top to bottom: Light source on the left/right of the participant; Light source on the top of the head of the participant; Light source underneath the participant's face; Light source in front of the participant - full lit face.

The difference, in this case, is that we have only focused on the luminance channel. The procedure was:

- Capture a sequence of seven exposures of a Xyla Dynamic Range Test Chart. This is a 21 stepped xylophone shaped back illuminated chart offering a DR of 20 fstops (i.e. the brightest bar is 2^{20} times brighter than the darkest bar);
- For each of the single exposure bars collect accurate measurements of the luminance values using a Minolta LS-150;
- Reconstruct the camera response function of this image using the pfstools 2.10 software library [39];
- Reconstruct an HDR image of the Xyla chart with the Robertson method [13] using pfstools 2.10;
- 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.

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

TABLE I
AVERAGE DYNAMIC RANGE FOR THE WARWICK HDR FACE DATASET

d) 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 camera response function was performed to produce HDR images. The resulting images were then filtered to remove singularities with the CleanWell function from the MATLAB HDR Toolbox [13] and then cropped in a passport-style fashion to 4,320 $\times$ 3,370 pixels - height $\times$ width ratio: 0.78.

E. 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 I. 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.

V. HDR VS LDR EXPERIMENT

This section presents the HDR vs LDR face matching methodology.

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.

A. 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. Participants were presented with two concurrent faces and had to decide if it was the same person or not by pressing a button.

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.

Each block contained twenty images presenting two concurrent faces under *same* or *different* lighting equally split between same-person pairs and different-person pairs.

Each participant was presented with a total of 60 trials, each trial contained a permutation of the three *DR* scenarios.

To avoid the familiarity effect, particular care was taken so that a specific face was not shown more than 10 times [40] throughout the trials. Also, to avoid picture comparison, even when the *same* lighting same-person pair was presented the two different images available in the dataset were used (see section IV).

There were two 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: performance (*time*) and accuracy (*accuracy*) are less affected by variation of light position when stimuli are in HDR.

B. Participants

The study involved 40 participants (15 males and 25 females) aged between 18 and 38 (mean age = 22.7, *SD* = 4.8) recruited among the University of Warwick students and employees through the *University of Warwick Sona System*. The participants' ethnicity varied: 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.

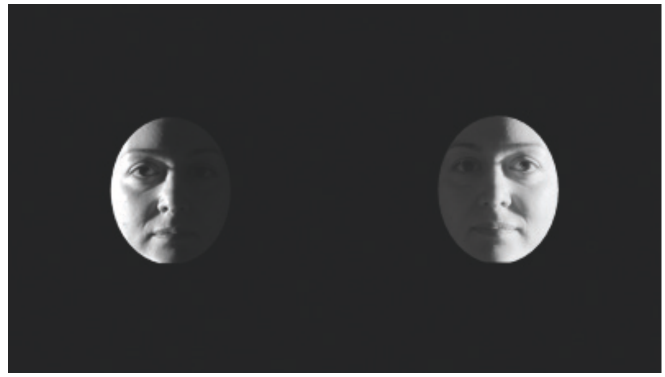


Fig. 3. Example of the stimuli proposed in the *HDR vs LDR* experiment. Stimulus: *different* lighting, LDR vs HDR (mix). The HDR (right) has been tone-mapped in order to be visualised on this paper.

C. 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 3). 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.

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. No feedback was provided to the participants regarding their accuracy while performing the task. Halfway through the test participants were allowed to take a short break.

D. Materials

The HDR face dataset described in Section IV was used as stimuli for this experiment. The faces depicted were cut out with an elliptic shape similarly to the approach taken by Megreya, Bindemann and Havard [41], 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 [42]. Internal features, on the other hand, are less susceptible to changing throughout time and they constitute a high number features used by experts to perform an identification task [30] (e.g. Scars/Blemish, Eyes, Nose, Mouth, Mouth Area, Forehead, Cheek Area).

Following the literature [25], [41], 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 chroma component and also considering that colour components seem to be less relevant in the recognition process [40]. 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 tone-mapping algorithm was needed. For further information on tone-mapping see Banterle *et al.* [13]. The optimal exposure algorithm proposed by Debattista *et al.* [43] 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.

All images were displayed at a full-HD resolution (1080 x 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 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 [44]. 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.

The participants were seated at 185 cm distance from the screen. This is the best viewing distance given the size of the screen (Recommendation ITU-R BT.2022). 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 [45]). See Figure 4 for an illustration of the setup.

VI. 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 II 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 II 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*. In the following, we report the results of statistical tests on the data.

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

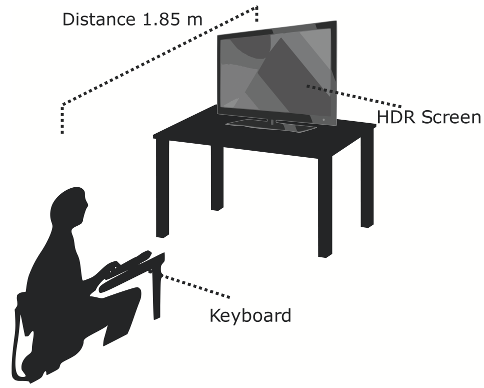


Fig. 4. Schematic representation of the HDR vs LDR experiment setup.

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		

TABLE II
Correct Answers and Reaction Time Descriptive Statistics

A. 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*.

B. 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 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.

The main effect of *position* was also significant for *accuracy*, F(1, 38) = 14.41, P $<$ 0.05. Pairwise comparisons showed a significant difference between *same* (μ = 88.99) and *different* (μ = 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) F(1.67, 74.80) = 2.31, P = 0.11.

C. 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* (μ = 3.06) and *LDR* (μ = 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* (μ = 2.91) with *different* (μ = 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* $\times$ *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 and in subsection VI-B we can also accept H_1^2.

D. 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* (μ = 95.28) and *LDR* (μ = 88.32) and also between *HDR* and *mix* (μ = 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* (μ = 2.76) and *LDR* (μ = 3.1) but no further significant differences with these two conditions and *mix* (μ = 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.

E. 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* (μ = 84.88) and *LDR* (μ = 78.44) but no significant difference with *mix* (μ = 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* (μ = 3.35) may be marginally faster than both *LDR* (μ = 3.52) and *mix* (μ = 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.

VII. Discussion

The work presented here 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 first part (Section IV) was related to the creation of an HDR faces stimuli 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 generated, 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.

The second part (Section V) demonstrated the perceptually added value of HDR compared to LDR imaging. The analysis of performance and accuracy of the three different blocks showed a clear advantage in using HDR technology, because of its intrinsic nature of light reproduction, leading us to accept both H_1^1 and H_1^2. This indicates that the dynamic range of the presented content significantly affects how accurately and quickly a face is matched. These results align with the literature which examines how lighting interferes with the facial recognition process, especially for unfamiliar faces. The results show also the advantage offered by HDR 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.

A. Limitations

The limitations of this work are related to the small 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 in necessary. This should include both genders, more ethnicities and also challenging pairs such as lookalikes, or the same faces with different features such as cosmetics or facial hair. 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.

VIII. Conclusion

The research presented in this paper shows that HDR is a technology that deserves further attention from digital surveillance and forensic communities. This study has demonstrated its potential applicability for improved performance in face matching tasks. Its accuracy in light reproduction makes HDR the ideal candidate for advancement in applications where reliability and high confidence is required.

Further research is needed if this technology has to be employed in critical situations. Nevertheless, this type of research could pave the way for novel imaging methods in face matching such as CCTV camera systems pipelines that better reproduce lighting in a more perceptually accurate fashion. This work will see future investigations exploiting the fullness of the stimuli including colour components. Additionally, it could be very interesting to explore the potential of HDR imaging in a live face matching context involving both experts and non experts.

Acknowledgment

The authors would like to thank the Rabin Ezra Scholarship Trust for their financial support and Demetris Marnerides and Tom Bashford-Rogers for their invaluable insight and encouragement.

References

[1] R. A. Johnston and M. Bindemann, "Introduction to forensic face matching," *Applied Cognitive Psychology*, vol. 27, no. 6, pp. 697–699, 2013.

[2] N. A. Spaun, "Handbook of Face Recognition," *Handbook of Face Recognition*, pp. 655–670, 2005.

[3] M. C. Fysh and M. Bindemann, "Forensic face matching: A review," in *Face processing: Systems, Disorders and Cultural Differences*, M. Bindemann and A. M. Megreya, Eds. New York: Nova Science Publishing Inc., 2017, pp. 1–20.

[4] A. Mike Burton, "Why has research in face recognition progressed so slowly? The importance of variability," *Quarterly Journal of Experimental Psychology*, vol. 66, no. 8, pp. 1467–1485, 2013.

[5] A. W. Young and A. M. Burton, "Recognizing Faces," *Current Directions in Psychological Science*, vol. 26, no. 3, pp. 212–217, jun 2017.

[6] K. Norell, K. B. Läthén, P. Bergström, A. Rice, V. Natu, and A. O'Toole, "The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons," *Journal of Forensic Sciences*, vol. 60, no. 2, pp. 331–340, mar 2015.

[7] R. Russell, B. Duchaine, and K. Nakayama, "Super-recognizers: People with extraordinary face recognition ability," *Psychonomic bulletin & review*, vol. 16, no. 2, pp. 252–257, 2009.

[8] R. Jenkins, D. White, X. Van Montfort, and A. M. Burton, "Variability in photos of the same face," *Cognition*, vol. 121, no. 3, pp. 313–323, 2011.

[9] M. Bindemann, J. Attard, A. Leach, and R. Johnston, "The Effect of Image Pixelation on Unfamiliar Face Matching," *Applied Cognitive Psychology*, vol. 27, no. November, pp. 707–717, 2013.

[10] C. H. Liu, W. Chen, H. Han, and S. Shan, "Effects of image preprocessing on face matching and recognition in human observers," *Applied Cognitive Psychology*, vol. 27, no. 6, pp. 718–724, 2013.

[11] W. L. Braje, "Illumination encoding in face recognition: effect of position shift," *Journal of Vision*, vol. 3, no. 2, pp. 4–4, 03 2003.

[12] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "High Dynamic Range Imaging," in *Wiley Encyclopedia of Electrical and Electronics Engineering*. New York, NY: Wiley, 2015, pp. 1–42.

[13] A. Artusi, F. Banterle, K. Debattista, and A. Chalmers, *Advanced high dynamic range imaging: theory and practice*. AK Peters/CRC Press, 2011.

[14] D. Wüller and H. Gabele, "The usage of digital cameras as luminance meters," in *Digital Photography III*, vol. 6502. International Society for Optics and Photonics, 2007, p. 65020U.

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

[16] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 192, 2016.

[17] A. H. Hung, T. Liang, P. A. Sukerkar, and T. J. Meade, "High dynamic range processing for magnetic resonance imaging," *PloS one*, vol. 8, no. 11, p. e77883, 2013.

[18] A. Degirmenci, D. P. Perrin, and R. D. Howe, "High dynamic range ultrasound imaging," *International journal of computer assisted radiology and surgery*, vol. 13, no. 5, pp. 721–729, 2018.

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

[20] B. Karr, A. Chalmers, and K. Debattista, "High dynamic range digital imaging of spacecraft," in *High Dynamic Range Video*. Elsevier, 2016, pp. 519–547.

[21] K. L. Ritchie, F. G. Smith, R. Jenkins, M. Bindemann, D. White, and A. M. Burton, "Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-id paradox," *Cognition*, vol. 141, pp. 161 – 169, 2015.

[22] K. L. Ritchie, R. S. Kramer, and A. M. Burton, "What makes a face photo a good likeness?" *Cognition*, vol. 170, pp. 1–8, 2018.

[23] P. J. Hancock, V. Bruce, and A. Burton, "Recognition of unfamiliar faces," *Trends in Cognitive Sciences*, vol. 4, no. 9, pp. 330–337, sep 2000.

[24] E. Noyes and R. Jenkins, "Camera-to-subject distance affects face configuration and perceived identity," *Cognition*, vol. 165, pp. 97–104, 2017.

[25] A. M. Burton, D. White, and A. McNeill, "The Glasgow Face Matching Test," *Behavior Research Methods*, vol. 42, no. 1, pp. 286–291, 2010.

[26] M. C. Fysh and M. Bindemann, "The Kent Face Matching Test," *British Journal of Psychology*, vol. 109, no. 2, pp. 219–231, 2018.

[27] D. S. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behavior Research Methods*, vol. 47, no. 4, pp. 1122–1135, Dec 2015.

[28] R. H. Logie, A. D. Baddeley, and M. M. Woodhead, "Face recognition, pose and ecological validity," *Applied Cognitive Psychology*, vol. 1, no. 1, pp. 53–69, 1987.

[29] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, "The good, the bad, and the ugly face challenge problem," *Image and Vision Computing*, vol. 30, no. 3, pp. 177 – 185, 2012, best of Automatic Face and Gesture Recognition 2011.

[30] Facial identification scientific working group. [Online]. Available: https://fiswg.org/index.htm

[31] M. Steyn, M. Pretorius, N. Briers, N. Bacci, A. Johnson, and T. Houlton, "Forensic facial comparison in south africa: State of the science," *Forensic Science International*, vol. 287, pp. 190 – 194, 2018.

[32] T. Valentine and J. P. Davis, *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV*. John Wiley & Sons, 2015.

[33] P. J. Phillips, A. N. Yates, Y. Hu, C. A. Hahn, E. Noyes, K. Jackson, J. G. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J.-C. Chen, C. D. Castillo, R. Chellappa, D. White, and A. J. O'Toole, "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *Proceedings of the National Academy of Sciences*, vol. 115, no. 24, pp. 6171–6176, 2018.

[34] P. Fussey and D. Murray. Independent Report on the London Metropolitan Police Services Trial of Live Facial Recognition Technology.

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

[36] P. Korshunov, M. V. Bernardo, A. M. Pinheiro, and T. Ebrahimi, "Impact of Tone-mapping Algorithms on Subjective and Objective Face Recognition in HDR Images," *Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia - CrowdMM '15*, pp. 39–44, 2015.

[37] W. L. Braje, "Illumination encoding in face recognition: effect of position shift," *Journal of Vision*, vol. 3, no. 2, pp. 4–4, 2003.

[38] M. H. Kim and J. Kautz, "Characterization for high dynamic range imaging," *Computer Graphics Forum*, vol. 27, no. 2, pp. 691–697, 2008.

[39] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel, "High dynamic range imaging pipeline: Perception-motivated representation of visual content," in *Human Vision and Electronic Imaging XII*, ser. Proceedings of SPIE, vol. 6492, no. 649212. SPIE, February 2007.

[40] C. A. Longmore, C. H. Liu, and A. W. Young, "Learning faces from photographs." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 1, p. 77, 2008.

[41] A. M. Megreya, M. Bindemann, and C. Havard, "Sex differences in unfamiliar face identification: Evidence from matching tasks," *Acta Psychologica*, vol. 137, no. 1, pp. 83 – 89, 2011.

[42] C. A. Longmore, C. H. Liu, and A. W. Young, "The importance of internal facial features in learning new faces," *The Quarterly Journal of Experimental Psychology*, vol. 68, no. 2, pp. 249–260, 2015.

[43] K. Debattista, T. Bashford-Rogers, E. Selmanović, R. Mukherjee, and A. Chalmers, "Optimal exposure compression for high dynamic range content," *The Visual Computer*, vol. 31, no. 6, pp. 1089–1099, Jun 2015.

[44] M. F. Damian, "Does variability in human performance outweigh imprecision in response devices such as computer keyboards?" *Behavior Research Methods*, vol. 42, no. 1, pp. 205–211, 2010.

[45] M. F. Peterson and M. P. Eckstein, "Looking just below the eyes is optimal across face recognition tasks," *Proceedings of the National Academy of Sciences*, vol. 109, no. 48, pp. E3314–E3323, 2012.

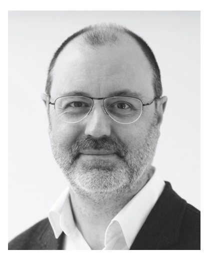
Rossella Suma received a MSc in Computer Engineering from the Polytechnic University of Turin, Italy, and BSc in Computer Engineering from the Polytechnic University of Bari.

She is currently a PhD candidate at the University of Warwick. From 2014 to 2016 she was a Marie Curie Fellow in the Initial Training Networks for Digital Cultural Heritage.

Her research interest include high dynamic range imaging, image processing, cognitive psychology, and virtual reality.

Kurt Debattista is a Reader at WMG, University of Warwick. He holds a BSc in Mathematics and Computing, an MSc and PhD in Computer Science as well as an MSc in Psychology.

His research interests include high-fidelity rendering, perceptual imaging, high dynamic range imaging, and high-performance computing.



Derrick Watson received the BSc in Psychology, MSc in Cognitive Science and PhD in Psychology degrees from the University of Birmingham, UK and the DSc from the University of Warwick, UK.

He was a Research Fellow at the University of Birmingham after which he joined the University of Warwick where he is currently Professor of Psychology.

His research includes: mechanisms of visual attention, capture and selection, face processing, emotion and decision making, and driving simulation and road safety. In 1999 he received the BPS Cognitive Psychology Award jointly with Glyn Humphreys for his work on visual selection.

Prof. Watson is a member of the Experimental Psychology Society, UK.

Elisabeth Blagrove received an LLB, a BSc and PhD in Psychology.

She is currently Senior Teaching Fellow in Psychology at the University of Warwick.

Her research interest include selective attention, processing of emotional faces, social attention, and cognitive ethology.

Alan Chalmers is a Professor of Visualisation at WMG, University of Warwick, UK and a former Royal Society Industrial Fellow with Johnson Tiles.

He has an MSc with distinction from Rhodes University, 1985 and a PhD from University of Bristol, 1991.

He is Honorary President of Afrigraph and a former Vice President of ACM SIGGRAPH (1999-2002). Chalmers has published over 250 papers in journals and international conferences on HDR, high-fidelity virtual experiences, multi-sensory perception, parallel processing and virtual archaeology and successfully supervised 48 PhD students.

He is founder and Scientific Advisor for two spinout companies: Sensomatics Ltd. and trueDR Ltd. In addition, Chalmers recently chaired EU COST Action IC1005 that coordinated HDR research across Europe, and he is a UK representative on IST/37 considering standards within MPEG.