

RESEARCH

Open Access



Screen reflections impact on HDR video tone mapping for mobile devices: an evaluation study

Miguel Melo^{1,2*} , Maximino Bessa^{1,2}, Luís Barbosa^{1,2}, Kurt Debattista³ and Alan Chalmers^{3,4}

Abstract

This paper presents an evaluation of high-dynamic-range (HDR) video tone mapping on a small screen device (SSD) under reflections. Reflections are common on mobile devices as these devices are predominantly used on the go. With this evaluation, we study the impact of reflections on the screen and how different HDR video tone mapping operators (TMOs) perform under reflective conditions as well as understand if there is a need to develop a new or hybrid TMO that can deal with reflections better. Two well-known HDR video TMOs were evaluated in order to test their performance with and without on-screen reflections. Ninety participants were asked to rank the TMOs for a number of tone-mapped HDR video sequences on an SSD against a reference HDR display. The results show that the greater the area exposed to reflections, the larger the negative impact on a TMO's perceptual accuracy. The results also show that under observed conditions, when reflections are present, the hybrid TMOs do not perform better than the standard TMOs.

Keywords: High-dynamic-range video, Tone mapping operator evaluation, Mobile devices, Screen reflections

1 Introduction

Current imaging techniques, also known as standard dynamic range (SDR) or low dynamic range (LDR), are not capable of representing all the real-world color gamut and contrast in a way that matches the human visual system (HVS)'s dynamic range. To overcome this limitation, high-dynamic-range (HDR) imaging was developed. Ensuring HDR is maintained along the entire imaging pipeline from capture to display allows the full range of captured scene data to be used in a number of applications, including security, broadcasting in difficult lighting conditions, etc.

When an HDR display is available, it is possible to deliver HDR content in a relatively straightforward manner [1]; however, the majority of displays currently available are still LDR. This is particularly true to mobile devices where there is, as yet, no HDR display. It is thus necessary to map any content's dynamic range to match that of the targeted display. The dynamic range

reduction can be achieved by employing tone mapping operators (TMOs). These take into account scene characteristics and/or the HVS properties in order to provide the best viewing experience on the LDR display from the available HDR data. A large variety of TMOs have been proposed, with only a few dedicated to HDR video and none designed specifically for HDR video on small screen devices (SSDs). TMOs for SSDs may need to take into account their portability which can result in situations in which there is a sudden exposure to widely differing luminance levels and reflections that could impact on the viewing experience.

Mobile devices have become widespread, and their penetration rate is reaching nearly 100 %, that is, one mobile device per person in the world [2]. Furthermore, mobile devices are already being widely used to consume multimedia, and indeed, it is estimated that around 51 % of the traffic on mobile devices is now video [3]. In fact, a recent report [4] showed that the online video requests are more and more made by mobile devices and, while in 2002 only 6 % of all online video was requested by mobile devices, in 2014, the number has increased to 30 % and it is projected to reach over 50 % by the end of 2015.

*Correspondence: emekapa@sapo.pt

¹ Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal

² INESC-TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

Full list of author information is available at the end of the article

With this growing popularity, it becomes important to consider possible challenges posed by mobile device displays compared to the traditional desktop devices. These include limitations in terms of dynamic range, viewing angle, and distance as well as size. In addition, to help ensure an optimal viewing experience, it is important to take into account external factors such as luminance levels and screen reflections. Furthermore, the current diversity of screens for mobile devices may require the re-targeting of content to play properly on any delivery screen. This can be a major challenge [5].

While previous work has addressed comparisons of diverse TMOs on mobile devices, none of this has addressed the impact of reflections on the screen when viewing HDR video on SSDs. As an SSD is very likely to be subjected to reflections, this paper presents a novel investigation on the visualization of HDR video on mobile devices under conditions where the display is exposed to reflections in order to understand their impact on the visualization quality. As reflections are more likely to happen outdoors, this work was concerned with simulating an outdoor environment with bright luminance levels. The insights gained from this study should suggest if there are advantages in developing a hybrid TMO to account for reflections and minimize any negative impact on the viewing experience. One possible future application arising from this study is the creation of automated methods that detect reflections on a screen and applies the best TMO according to the usage scenario. The evaluation of the HDR video tone mappers in this paper is carried out with three different scenarios:

- With a reflection across the entire screen
- With no reflections
- With a reflection on half of the screen (this can be either the left or right side)

The division of the screen was chosen so we would be capable of defining precise areas of the display where we could employ a specific TMO. This study considered one of the larger size SSDs, an iPad 4 which has a 9.7-in. screen. This is representative of current SSDs including mobile phones, whose screen size is increasing.

For the experiments, six HDR videos were used (Fig. 1). Two TMOs were considered: the model of visual adaptation [6] and the display adaptive technique [7], both successful TMOs in a number of previous experiments, for example, [8].

2 Related work

A wide variety of TMOs have been proposed to date. These have all been developed with different concerns or goals in mind. TMOs can be based on simple mathematical operations such as exponential or logarithmic



Fig. 1 Frames from each of the six HDR videos that were used for the evaluations

functions performing linear transformations as well as be more elaborate and inspired by, for example, features of the HVS [9]. They can be broadly divided on two categories: global and local. As the name suggest, global TMOs process the image as a whole, applying the same computation to every pixel while local TMOs process the image pixel by pixel taking into account the adjacent pixels. Standard global and local approaches work well with images but when the target is video, there is another important feature to account for: temporal coherence. A new category of TMOs has thus emerged: time-dependent TMOs.

Due to the wide variety of TMOs proposed and the advantages that HDR can bring, it has been important to perform evaluations in order to understand and identify which are the best TMOs for certain scenarios. Despite the efforts on TMO evaluation for HDR video, few papers have explored TMO evaluation for mobile devices and none has addressed the problem of reflections on a screen as often happens when using mobile devices, especially outdoors. The next section presents a brief survey of previous methodologies to evaluate TMOs.

2.1 TMO evaluation

As several TMOs have been proposed over the years, several TMO evaluation studies were also conducted adopting two main methodologies: error metrics and psychophysical experiments.

The error metrics methodology consists of objective measures based on theoretical models using computers to compare images/videos. This methodology can be based

on simple approaches such as measuring and comparing individually pixel values as well as something more complex as, for example, simulating the HVS and identifying the differences between the original content and the produced content based on visual perceptible differences. One popular example of such a technique is the HDR-VDP (visual difference predictor) [10] and HDR-VDP2 [11].

Psychophysical experiments, on the other hand, are based on studies conducted with human participants and therefore subjective. One key aspect of the experiments is to guarantee the preservation of the variables over the experiments in order to accomplish a well-controlled scenario where the participants take the tests and perform their judgment over the contents being evaluated. Another key aspect is the proper randomization of the variables to be evaluated between participants in order to avoid bias. The evaluations can be made with or without a reference. For HDR, the reference is typically the real-world scene or an image or video shown on an HDR display.

One of the first TMO psychophysical studies conducted was by Drago et al. that evaluated seven different TMOs applied to four different scenes with 11 participants on a pairwise comparison without reference [12]. The first evaluation study made using an HDR display as reference was conducted by Ledda et al. that evaluated six TMOs applied to 24 images that were evaluated by 18 participants [13]. An example of experiments using real-world scenes as reference are the ones conducted by Čadik [14] that evaluated 14 TMOs in three different scenes.

More recently, evaluations started addressing HDR video, for example, the work by Eilertsen et al. [15] that evaluated 11 TMOs on a set of HDR videos that were both camera captured and computer generated. A total of 36 participants took part in the experiments where they were asked to make pairwise comparisons between the tone-mapped footage without reference. Regarding mobile devices, only a few studies have been done. The paper by Urbano et al. was the first one that was aimed specifically at SSDs [16]. The study evaluated TMOs on different sized displays (17" for conventional sized and 2.8" for small sized) through pairwise comparison against the real-world scene and concluded that for mobile devices, content that offered stronger detail reproduction, more saturated colors, and overall brighter image appearance was preferred.

Akyüz et al. also evaluated both TMOs and exposure fusion algorithms on SSDs [17]. The study was divided in two pairwise comparison experiments where some scenes were evaluated with a reference and some without a reference. The displays used were 24" for the conventional sized display and 3" for the SSD. The participants were asked to evaluate color, contrast, and detail.

Other work which considered HDR video tone mapping for mobile devices was conducted by Melo et al. [18]. The evaluation was performed using a 37" LCD display or a 9.7" display with an HDR display as reference. The results demonstrated that there was a statistically significant difference between the choice of TMOs between the SSD and the large screen display although the TMOs accuracy order remained the same across the two displays. Further work by the same authors investigated the impact on visualization of HDR video on mobile devices under different lighting levels. Three scenarios were considered: dark, dim, and bright lighting levels. The study showed that under dark and dim environments, the TMOs' accuracy ranking obtained was different than that for the bright lighting level environments. The paper concluded that participants gave more importance to contrast and naturalness over details and color saturation in bright environments [8].

2.2 Video tone mapping operators

HDR video tone mapping is a growing field and a number of successful video TMOs have been proposed and successful in previous evaluations using mobile devices ([8, 15]) such as the time-dependent visual adaptation TMO [6], the display adaptive TMO [7], the visual adaptation for realistic images TMO [19], or the temporal coherence tone mapping method [20]. Other examples of well-known TMOs that were designed to address HDR video are the encoding of HDR with a model of human cones [21], the temporally coherent local tone-mapping of HDR Video [22], or the real-time noise-aware tone mapping [23].

The time-dependent visual adaptation TMO [6] exploits the fact that the HVS does not adapt instantly to big changes in luminance intensities. In this method, the appearance is modified in order to match the viewers' visual responses so they can perceive the scene as they would in reality. It uses a global adaptation model based on Hunt's static model of color vision and uses the retina response signals for rods for calculating the luminance information and the response vector for color information. In addition, temporal coherency is added. The method is not computationally complex and thus is suitable to use in real-time applications.

Regarding the time-dependent visual adaptation TMO [7], it is a TMO that is capable of adapting to the display features. This TMO offers a set of default, ready to use profiles that are pre-configured, and it is also possible to configure parameters individually such as display reflectivity, the peak luminance, or black level of the display. This TMO uses an HVS model in order to minimize the visible contrast distortions taking into account the characteristics of the given display. The TMO ensures temporal coherency through the limitation of the temporal

variations above 0.5 Hz as this is the peak sensitivity of the HVS for temporal changes.

The TMO proposed by [19] is based on a model of visual adaptation from psychophysical experiments. It considers key aspects of the HVS such as visibility, visual acuity, and color appearance. For modeling photopic and scotopic vision, this operator uses TVI functions. In order to achieve the mesotopic range, the authors use a linear combination of both the photopic and scotopic ranges.

The method proposed by [20] follows a different approach as it post-processes the HDR content as it is not capable of doing real-time processing. This is because initially, the method analyzes the whole video sequence in order to preserve the temporal stability of the video sequence. It is used in combination with static TMOs, and the focus was on optimizing it to be used with the Reinhard's photographic tone reproduction method. This operator first processes each frame of the video individually with a static TMO, and then, it considers the luminance of each frame taking into account the features of the whole HDR video sequence. The encoding of HDR with a model of human cones TMO developed by [21] is based on the dynamical response characteristics of primate cones and deals with the temporal coherency by employing temporal filters to handle noise through the absorption of the retinal illuminance by visual pigment. This is achieved by using two low-pass filters where the first is responsible for reducing the dynamic range of the content in order to fit the displays' dynamic range by applying a combination of dynamic non-linearities. The second low-pass filter is based on a non-linear differential equation that reduces noise and automatically adapts to the prevailing scene luminance.

The real-time noise-aware TMO developed by [25] offers a video tone-mapping process that controls the visibility of noise as well as it is capable of adapting itself to the display and viewing conditions and minimizes contrast distortions. Authors describe their method based on three main parts: edge-stopping spatial filter, local tone curves, and noise-aware control over image details. The first is responsible for transforming the input into a log domain, a base layer that describes luminance variances over time and a detail layer that describes the local features. Then, the local tone curves block compresses the dynamic range of the base layer using a set of local tone-curves that are distributed spatially through the scene. Each tonecurve is responsible for mapping the luminance range of the input into the range of luminance that is afforded by the target display. Regarding the noise-aware control over image details block, it gets as input a base layer, a detail layer and the tonecurves in order and allows users to preserve or enhance the local contrast and details of the scene and the visibility noise is controlled based on the noise characteristics on the input layers. The final step consists

in applying an inverse model in order to transform the colometric values into pixels.

The temporally coherent local tone-mapping of HDR Video proposed by [23] was designed having as concern the temporal artifacts and the limited local contrast reproduction capability common to TMOs in general. In order to avoid these problems, the authors worked on a temporal domain extension of the common spatial base-detail layer decomposition. The pipeline of this TMO can be divided into 3 main steps: a spatiotemporal filtering that is performed on adjacent frames and uses optical flow estimates to warp each frame's temporal neighbourhood and avoid artifacts; a temporal filtering that reduces temporal artifacts by penalizing flow vectors with high gradients; and the final tone mapping step that is capable of maintaining the average value of brightness over time as well as an high local contrast.

3 Experimental setup

This section describes the experiments undertaken. In the experiments, the participants evaluated six HDR video sequences with four TMOs under three different scenarios (with a reflection across the entire screen, with no reflections, and with reflections on half of the screen).

3.1 Method

An experimental framework that makes use of randomization of the videos and TMO combinations was used in order to minimize selection bias. A rank-based evaluation was carried out across the TMO methods over six HDR videos. Since the goal was to evaluate HDR video tone mapping for mobile devices, a method was needed that allowed us to define reflections on the screen with precision under controlled conditions. The solution adopted was to point a photographic softbox directly towards an area of the screen as such a device allows the distribution of light in a well-defined area. The vertical division point from the part of the screen that was under reflections and the other part of the screen that was not under reflections was the middle of the screen. To ensure the reflections were the same for all relevant experiments, we had some markers (that were invisible to participants) that indicated that the reflection was being applied correctly. To avoid bias in the scenarios where there was reflection, the scenario was always randomized between participants so that the reflection could be over the left or right half of the display. Figure 2 shows the mobile device with the reflection applied to half of the screen.

The experiments were conducted in a room in which all the environmental variables could be controlled. There are three independent variables: the set of TMOs used, the reflections on the display, and the scene groups. The scene groups and the reflections on the display were

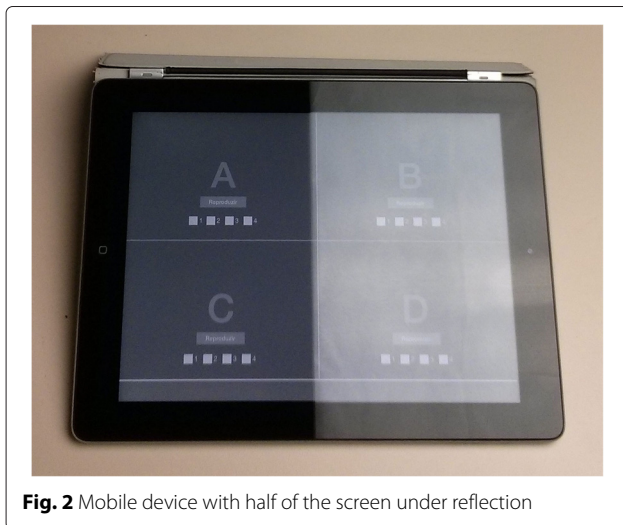


Fig. 2 Mobile device with half of the screen under reflection

in-between participant independent variables; the TMOs was a within-participant variable.

The overall results were analyzed using a 2 (scene groups) \times 3 (reflections on display) \times 4 (TMOs) mixed factorial ANOVA. The main effects calculated across all videos were of the group (each viewed three videos), TMO (the four TMOs used), and the scenario (with reflections, without reflections, and with reflections on half of the screen). Regarding each evaluated scenario, the data gathered is relative to 30 participants and was analyzed using a 2 (scene groups) \times 4 (TMOs) mixed factorial ANOVA.

The set of TMOs used were the display adaptive TMO, the time-dependent visual adaptation for fast realistic image display TMO, and two hybrid approaches of both. The reflections on the screen variable consisted of three conditions: reflections across the whole screen (scenario 1), no reflections on the screen (scenario 2), and reflections on half of the screen (scenario 3). The HDR videos were tone mapped in four different ways: using Pat across the whole frame; using only Man across the whole of the frame and dividing the video vertically in half and applying Man to one side and Pat to the other (referred from now on as ManPat); and dividing the video vertically in half and applying Pat on one side and Man (referred from now on as PatMan) on the other. The scene groups variable consisted of two groups, the participants were divided, where half of them evaluated the first three scenes and the other half the second three scenes.

To avoid bias, one concern was to ensure that the participant was at a correct distance from the HDR display. The participant was placed at a distance of approximately 1.8 m since the suggested viewing distance for high-definition displays is three times the height of the display (which was approximately 60 cm) according to the International Telecommunication Union Recommendation BT.500-13 [24]. A further concern was the luminance

adaptation of the human eye. To avoid maladaptation, we thus gave the participant some time before beginning the experiment to adapt himself/herself to the experiment scenario. Another concern was the auto-brightness feature of mobile devices that increases display brightness in brighter scenarios. To address this, one has set the devices' brightness to the maximum level ensuring that the mobile device was having its best performance on the given conditions.

Bias is always an important factor in subjective studies since this type of study is based on participants' answers that could be influenced by many factors. It is important, therefore, to pay as much attention as possible to account for possible disturbances. In the case of mobile devices, visual angle and viewing distance are important variables. The mobile device was thus placed on a stand (Fig. 3), and we have been careful to place each participant at approximately the same position so the viewing angle and viewing distance were approximately the same during the experiments. It was not possible to use a chin rest to ensure a stable position of the participant but between each video the position of the participant was checked and he/she was instructed to move his/her head only in a vertical axis in order to look at the HDR display and at the mobile device to evaluate the TMO accuracy. Reflection is a complex issue and since there is no previous work in the field, we considered only a larger size SSD. Future work will consider the effect of reflection on mobile devices with different screen sizes.

3.2 Materials

The generic setup for all the experimental scenarios is presented in Fig. 4. This work arises from previous work that evaluated a set of TMOs for mobile devices under different ambient lighting levels [8]. As the main goal is to evaluate the impact of reflections on TMO performance rather than evaluating which of the existent TMOs is the best for the depicted scenarios, the evaluations

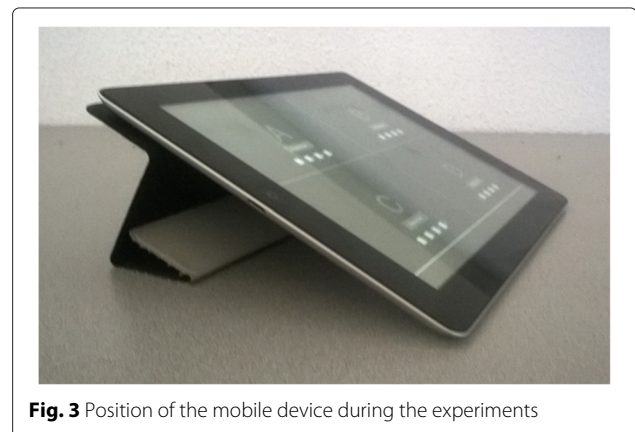
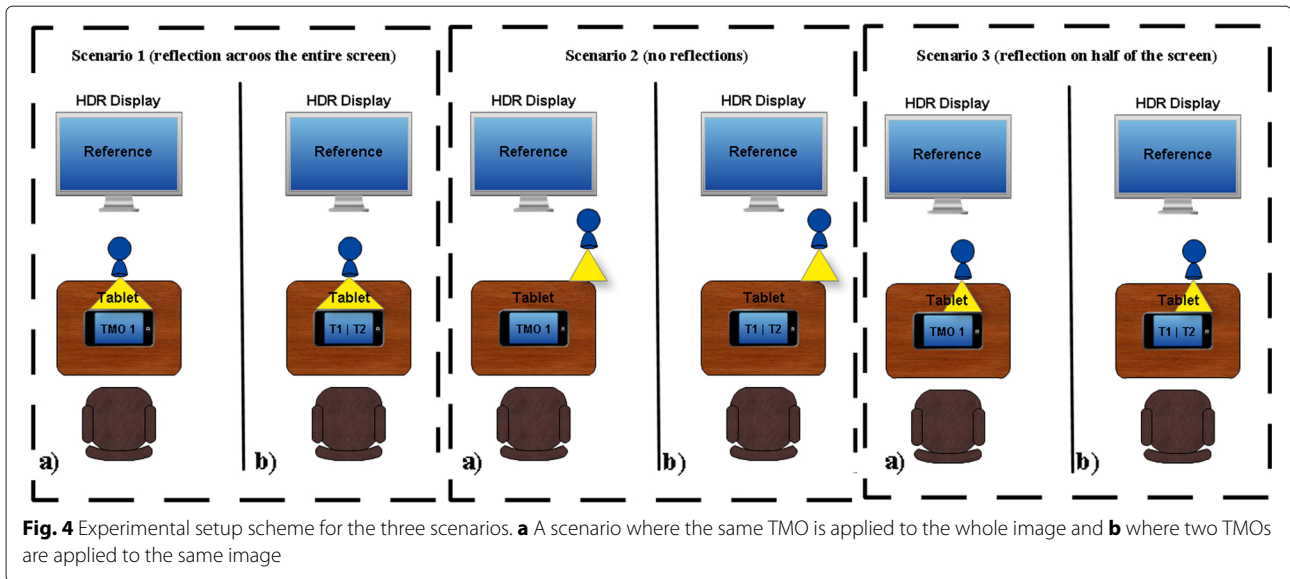


Fig. 3 Position of the mobile device during the experiments



considered two TMOs that performed well on different ambient lighting conditions. Therefore, we selected the best ranked TMO for dark and medium scenarios as well as the best ranked TMO for the bright scenario: the display adaptive TMO ([7]) (that will be referred in this paper as Man) and the time-dependent visual adaptation TMO ([6]) (that will be referred to as Pat). This choice was made since the two TMOs can give insights on reflection impact due to their differences such as, for example, Pat intends to be a visual system simulator that includes temporal coherence by simulating the HVS, while Man is considered a best subjective quality TMO that was designed to produce the best looking images based on subjective studies. In general, looking at the tone-mapped footage produced for the experiments, Pat seems to produce more natural and higher contrast images while Man deliver more detailed and saturated images.

The “join” between the two TMOs was attenuated by a fade between the TMOs so there was no visible division. To achieve it, one filled 60 % of right side with one TMO and 60 % of the left side with the other TMO. Where both TMOs are present, there was applied a sort of gradient with the alpha being that at the middle of the join, there was 50 % of each. An example of processed frames is shown in Fig. 5. The photographic softbox was always placed so it was not between the participant and the reference and in the experimental scenario with no reflections the photographic softbox was also on to ensure the same lighting level across all the scenarios.

For the experiments, the HDR reference display was a properly calibrated 47” SIM2 display and the tablet used was an iPad 4 from Apple. The technical specifications of the displays are presented in Table 1.

Six different HDR videos were considered, labeled “CGRoom”, “Explosion”, “Jaguar”, “Kalabsha”, “Medical” and “Morgan Lovers”. Table 2 shows the features of these videos where the average dynamic range (avg. DR) is expressed log units and the length in seconds. The measurement of the avg. DR was obtained by disregarding the top 1 % and bottom 1 % of the values in each frame and averaging them; this was performed to avoid possible error introduced by noise in the frames.

The experimental room had a controlled luminance level of 1450 cd/m² which is equivalent to the average local outdoor luminance level recorded at the time of the experiments and corresponds to a typical partially cloudy day. This luminance level was obtained by using a strong ceiling illumination as well as four photographic spotlights.

Specific experimental software was devised for ranking the TMOs (Fig. 6). For each video, the participants were able to see the tone-mapped content on the iPad and the corresponding HDR reference on the HDR display simultaneously. They could watch the videos as many times they wanted before ranking the TMOs according to how well the tone-mapped video matched the HDR reference. After ranking the four combinations for a video sequence, a button appeared that allowed participants to proceed to the next video. When all videos were evaluated, a message appeared to inform the participant that he/she had finished.

The optimal TMO settings were determined by a group of three experts (who each had at least 2 years of experience in HDR imaging). The global version of Pat was used since the local version does not support the time-dependent effects. The configurable settings for this TMO are the adaptation levels for cones, and the adaptation

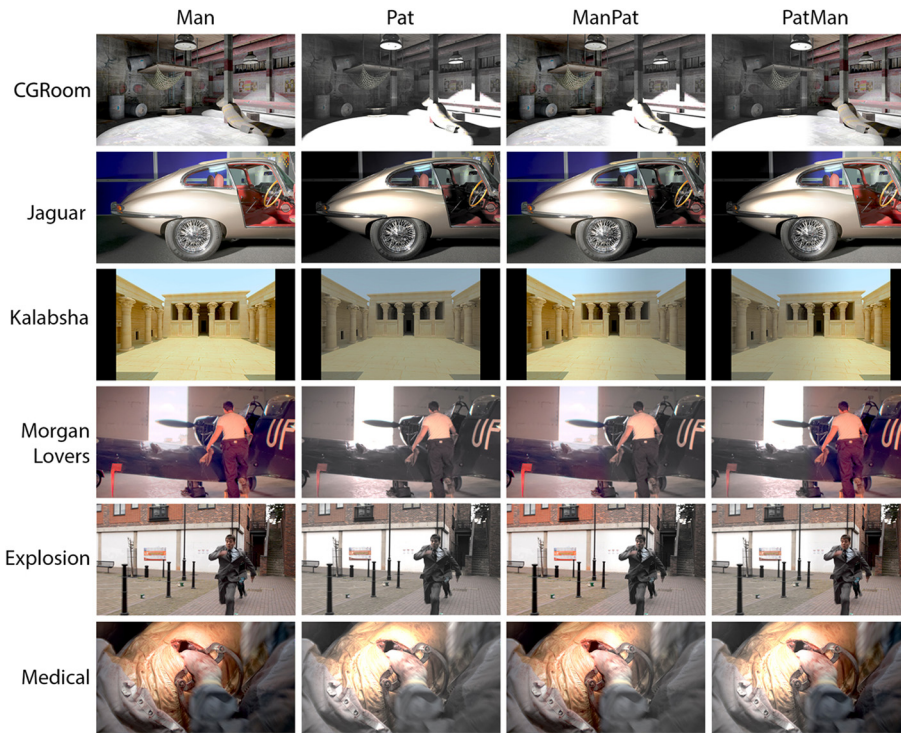


Fig. 5 Example of processed frames used in the experiments

levels for rods that were calculated for each case using the average luminance. For Man, when under reflection-affected areas, the following settings were applied: gamma correction $\gamma = 2.2$, maximum display luminance $L_{max} = 200$, the black levels of the display $L_{black} = 0.8$, reflectivity of the display $k = 0.01$, and the ambient illumination $E_{amb} = 400$. When there were no reflections, the following parameters were used: gamma correction $\gamma = 2.2$, maximum display luminance $L_{max} = 200$, black levels

of the display $L_{black} = 0.8$, the reflectivity of the display $k = 0.01$, and the ambient illumination $E_{amb} = 1450$.

3.3 Participants

A total of 90 participants, 51 men and 39 women aged between 19 and 28 years, were randomly assigned between the different experimental scenarios and between the six scenes (half of them evaluated a set of three videos—CGRoom, Jaguar, and Kalabsha—and the other half the set containing the remaining three videos—Explosion, Medical, and Morgan Lovers). For each experimental scenario, there were a total of 15 evaluations for each video as it was the minimum participants required to obtain significant results. All the participants reported normal or corrected to normal vision. The grouping of the scenes was random.

3.4 Procedure

The participant stood at approximately 1.8 m from the HDR display at a table on which the tablet was placed. Each experimental scenario (reflection conditions and scene group) was randomly assigned between the participants. Before each experiment, the experimental room was prepared accordingly. The participants had a brief explanation of how they would participate in the experiments. As quality is a key factor, the participants were

Table 1 Technical specifications of the displays used in the experiments

	HDR display	Mobile device
Brand	SIM2	Apple
Model	HDR47ES4MB	iPad 4
Size	47"	9.7"
Resolution	1920×1080	2048×1536
Contrast ratio	> 1,000,000:1	877:1
Max luminance	> 4000 cd/m ²	476 cd/m ²
Min luminance	0 cd/m ²	0.48 cd/m ²
View angle (horizontal)	40°	175°
View angle (vertical)	15°	175°

Table 2 Features of the HDR videos used

Video	Length (s)	Avg. DR (log units)	Capture	Device max F-stops
CGRoom	7	16.6	CG	20
Jaguar	13	13.4	Canon 1D Mark II	14
Kalabsha	11	18.5	CG	20
Morgan Lovers	15	19.5	Spheron HDRv	20
Explosion	8	12	Canon 5D	12
Medical	4	15.2	Spheron HDRv	20

asked to take into account color, contrast, naturalness, and details as a whole as these parameters are well known for characterizing an image [25]. The evaluation software was shown to the participants so they could familiarize themselves with it. The average time for each experiment was around 10 to 15 min.

4 Results

A large amount of data was collected and analyzed. To facilitate the presentation of the results, they are divided into subsections with a discussion following in the next section.

As mentioned in Section 3.1, the overall results were analyzed using a 2 (scene groups) \times 3 (reflections on display) \times 4 (TMOs) mixed factorial ANOVA. The main effects calculated across all videos were of the group (each viewed three videos), TMO (the four TMOs used—two standard TMOs and the two hybrid combinations of the standard TMOs used), and the scenario (with reflections, without reflections, and with reflections on half of the screen). Regarding each evaluated scenario, the data gathered is relative to 30 participants and was analyzed using a 2 (scene groups) \times 4 (TMOs) mixed factorial ANOVA. In scenario 3 (when there is reflection only on half of the display) and for ManPat and PatMan, the results were treated

so the TMO under reflection is the first (Man in ManPat and Pat in PatMan). The results also present Kendall's coefficient of concordance W that serves as an estimate of agreement among participants where $W = 1$ signifies perfect agreement among participants and $W = 0$ completes disagreement. Based on Kendall's coefficient of concordance W , we test the null hypothesis as to whether there is no agreement among the participants at $p < 0.05$. This provides an indication of the agreement among the participants. Table 3 summarizes the obtained results. We highlight that, despite the fact that Kendall's coefficient of concordance does not range high in values (it is between 0.14 and 0.24).

4.1 Overall results

For the different scenarios and TMOs that were evaluated, the statistically significant difference is given by their main effect that was $F(5.86, 245.95) = 2.13$ ($p < 0.05$) and $F(2.93, 245.95) = 20.98$ ($p > 0.05$), respectively. Regarding the group main effect, the value reported was $F(2.93, 245.95) = 6.59$ ($p > 0.05$). On all cases, sphericity was violated; therefore, Greenhouse-Geisser correction was applied.

These overall results show that differences are statistically significant between the evaluated TMOs as well as between the different scenarios that were considered. The consistency of the rankings was also significant since the computed Kendall's coefficient of concordance was $W = 0.17$. The rank order obtained shows that Pat was the best ranked TMO. ManPat was second grouped with PatMan and Man while PatMan was third grouped with Man. Also, although there were two groups ranking the scenes, there was no significant difference between them which is a good result indicating coherence between choices.

As for the TMOs, looking closely to the groupings, it is noticeable that there is a TMO that clearly outperforms the others. Pat has been consistently ranked as the most accurate TMO, and the difference between this TMO and the remaining ones is statistically significant. The second ranked TMO was ManPat grouped together (meaning that their difference is not statistically significant) with PatMan and Man. There is a third grouping differentiating PatMan and Man from the others.

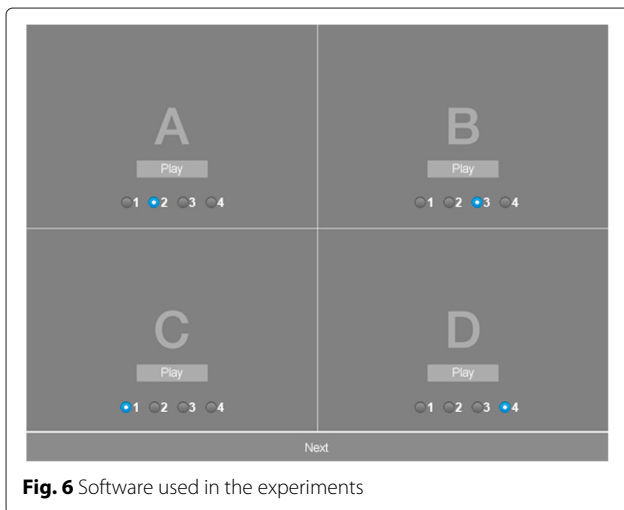
**Fig. 6** Software used in the experiments

Table 3 Overall results obtained for each scenario

	Kendall's coefficient of concordance	1st	2nd	3rd	4th
Across all scenarios	0.17	Pat	ManPat	PatMan	Man
Scenario 1 (reflections across the entire screen)	0.14	Pat	ManPat	Man	PatMan
Scenario 2 (no reflections on the screen)	0.24	Pat	ManPat	PatMan	Man
Scenario 3 (reflections on half of the screen)	0.19	Pat	PatMan	ManPat	Man

Colored groupings represent TMOs that were not found to be significantly different using pairwise comparisons to each other, via Bonferroni adjustment, at $p < 0.05$

4.2 Scenario 1—reflections across the whole screen

Scenario 1 was where the participants were less consistent while ranking the TMOs, and therefore, the groupings were more complex, although the obtained results show that there is significant difference between TMOs. Despite the complex groupings, it is possible to see that Pat and ManPat were the top two TMOs, with no significant difference between them, followed by Man and PatMan.

The reported main effect of TMO was $F(2.68, 74.97) = 5.31$, and the Greenhouse-Geisser correction as sphericity was violated (Maulchy's test for sphericity, $p < 0.01$). As before, the main effect of the group was not

significant $F(2.68, 74.97) (p > 0.05)$. The Kendall's coefficient of concordance was significant at $W=0.14$. Table 4 shows the results obtained for each scene on this scenario.

4.3 Scenario 2—no reflections on the screen

In scenario 2, Pat was once again ranked as the top TMO and there was a significant difference from the other TMOs which were all grouped together (ManPat, PatMan, and Man was the ranking order). In this scenario, the ranking concordance was the highest compared with the other scenarios. There was also no significant difference between the groups' rankings.

Table 4 Results obtained for each video on Scenario 1 (reflections across whole screen). Colored groupings represent TMOs that were not found to be significantly different using pairwise comparisons to each other, via Bonferroni adjustment, at $p < 0.05$.

	Kendall's Co-efficient of Concordance	1 st	2 nd	3 rd	4 th
CGRoom	0.03	Man	Pat	PatMan	ManPat
Jaguar	0.12	Pat	ManPat	Man	PatMan
Kalabsha	0.4	Pat	PatMan	ManPat	Man
Morgan Lovers	0.09	Pat	PatMan	ManPat	Man
Explosion	0.03	ManPat	Man	PatMan	Pat
Medical	0.04	ManPat	Man	PatMan	Pat

Table 5 Results obtained for each video on Scenario 2 (no reflections on the screen). Colored groupings represent TMOs that were not found to be significantly different using pairwise comparisons to each other, via Bonferroni adjustment, at $p < 0.05$.

	Kendall's Co-efficient of Concordance	1 st	2 nd	3 rd	4 th
CGRoom	0.05	PatMan	Pat	ManPat	Man
Jaguar	0.07	Pat	Man	PatMan	ManPat
Kalabsha	0.05	Pat	PatMan	Man	ManPat
Morgan Lovers	0.04	Pat	ManPat	PatMan	Man
Explosion	0.19	PatMan	Pat	ManPat	Man
Medical	0.12	ManPat	Pat	PatMan	Man

The results reported for TMO main effect were $F(2.82, 81.23) = 2.55$ (Greenhouse-Geisser correction as sphericity was violated, $p < 0.01$). The reported group main effect was not significant $F(3, 84) = 3.58$ ($p > 0.05$). The computed Kendall's coefficient of concordance was $W = 0.24$ ($p < 0.05$). For further reference, the results for each scene are shown on Table 5.

4.4 Scenario 3—reflections on half of the screen

The third scenario reported a significant difference between TMOs and, once again, Pat was classified as first. However, in this scenario, it was grouped together with PatMan meaning that there were no significant differences between them. The level of concordance in this scenario was higher than in the scenario in which the screen was entirely under reflection.

The TMOs' main effect was $F(2.91, 81.23) = 7.74$ (Greenhouse-Geisser correction was applied, $p < 0.01$). As before, there was no significant difference between the groups' rankings as the main effect was $F(2.91, 81.23) = 1.57$ ($p > 0.05$). The Kendall's coefficient of concordance on this scenario was $W = 0.19$ ($p < 0.05$). The results for each scene are shown on Table 6 for completeness.

5 Discussion

The main goals of the experiment were to study the impact of reflections on mobile device displays and to provide some insights on whether the perceptual accuracy of the TMOs changed depending on the different scenarios. With this knowledge, it is possible to understand if there is a need to develop a new or hybrid TMO that can deal with reflections better and which features should be taken

Table 6 Results obtained for each video on Scenario 3 (reflections on half of the screen). Colored groupings represent TMOs that were not found to be significantly different using pairwise comparisons to each other, via Bonferroni adjustment, at $p < 0.05$.

	Kendall's Co-efficient of Concordance	1 st	2 nd	3 rd	4 th
CGRoom	0.07	PatMan	Pat	ManPat	Man
Jaguar	0.06	Pat	PatMan	Man	ManPat
Kalabsha	0.12	Pat	PatMan	Man	ManPat
Morgan Lovers	0.16	Pat	Man	PatMan	ManPat
Explosion	0.05	ManPat	Man	PatMan	Pat
Medical	0.04	Pat	PatMan	Man	ManPat

into account when working with TMOs with the purpose of dealing with reflective scenarios.

For the three scenarios, the calculated Kendall's coefficient of concordance was significant for $p < 0.05$ meaning that there was significant agreement between the participants, giving statistical significance to the obtained results. Furthermore, the results showed that there was no significant difference between the two groups that performed the experiment indicating coherence between choices. This gives more strength to the results. A first conclusion that can be extracted from the results is that overall results reported significant differences between the three experimental scenarios and all the scenarios reported significant differences between the TMOs. This indicates that the reflections do indeed have an impact on a TMOs' perceptual accuracy and that their perceptual accuracy can change according to the usage scenario.

Scenario 1 (reflection across the whole display) had a lower Kendall's coefficient of concordance indicating a lower agreement between the participants' rankings. On top of that, the groupings were complex in this case suggesting that it is more difficult to choose which TMO is the best for this condition. This may indicate that when the display is fully exposed to reflections, a TMO's perceptual accuracy can be compromised and the visualization experience negatively affected. Further studies are required to validate this assumption.

Scenario 2 (no reflection) is similar to the experiment described in [8]. As in this paper, Pat was the best ranked TMO. One significant difference between scenario 2 and [8] was that the Kendall's coefficient of concordance reported was $W = 0.24$ for this work compared with $W = 0.89$ in [8]. This can be because in this paper, we selected only the top two TMOs reported in [18] so the choice in quality may not have been that obvious.

Interestingly, in scenario 2, although PatMan and ManPat are a combination of two TMOs on the same frame, they were grouped together with Man and indeed slightly preferred than Man. This result suggests a preference towards the video that includes the most preferred TMO, even if it has not been used for the whole frame. Eye tracking will be used in future work to help confirm whether the participants spent more time looking at the part of the frame computed with the preferred TMO or not.

In scenario 3 (reflection across half of the screen) the participants agreed more on their choices than in scenario 1 (reflection across whole screen) ($W = 0.19$ against $W = 0.14$) which means that reflections have a negative impact on the TMOs' perceptual accuracy and the more the area exposed to reflections the greater this negative impact. The TMOs rank order was Pat, PatMan, ManPat, and Man. An important result is that Pat and PatMan were significantly better than the other two TMOs.

Urbano et al. [16] identified that stronger detail reproduction, more saturated colors, and overall brighter image appearance were preferred. Our findings do not corroborate with those, and we attribute this to visual attention mechanisms as it has also been shown that the mechanisms of visual attention are much more significant for images than for video due to the interframe correlation and shorter viewing time [26]. An additional factor that could have contributed to the differences between Urbano et al.'s [16] work and ours is that while viewing an image, participants focus more on all the details across the whole image whereas when viewing a video, it is most likely that they devote more attention to regions where motion occurs [27].

Another interesting result is that PatMan and ManPat are always grouped together with Man and that they were slightly preferred than Man. This result indicates a preference towards the video that have been tone mapped with the preferred TMO even if this is not on the entire frame. As in scenario 2, this needs to be confirmed with eye tracking.

6 Conclusions

In this paper, we set out to undertake an evaluation of the impact of reflections on a screen by understanding how different HDR video TMOs perform under reflective conditions. In addition, this study intended to clarify if there is a need to develop a new or hybrid TMO that can deal with reflections better than existing TMOs.

Overall, the results have shown that there are significant differences between scenarios where there is no reflection and scenarios with reflections as well as differences between the TMOs' perceptual accuracy. The results further show that reflections have a negative impact on the visualization experience since there was less consistency and coherence in participants' responses in the scenarios where there were reflections on the display. An important result is that Pat was the top TMO for all the scenarios, i.e., with and without reflections. It outperformed the hybrids, PatMan and ManPat. Furthermore, the hybrids were never significantly better than the second preferred TMO, Man. We can conclude therefore that, at least in the scenarios we studied, there is no need for a hybrid TMO. This might also suggest that, similarly to what happens when comparing TMOs' performance on dark environments against TMOs' performance on bright environments, participants seems to prefer contrast and naturalness over color and details, and therefore, these image features should be carefully addressed when developing TMOs to deal with reflective scenarios.

The paper did show that the scenario 3 results were more coherent and consistent than the results of scenario 1. This demonstrates that the reflections' negative impact in this scenario was less than in scenario 1, suggesting that

the more reflections on the display, the more the negative the impact on the viewing experience. The paper did not answer, however, whether having only half of the screen exposed to reflections was more negative for the viewing experience than having the full screen exposed, especially as the visual mechanisms could not adapt properly to the different exposed regions of the screen.

Finally, the results suggest that having two TMOs simultaneously applied to the same frame does not have a negative effect on the perceptual accuracy rankings. Here, the participants demonstrated a preference for the cases where Pat was applied (even only partially). This has raised the question: “Did participants unconsciously rank the videos based on the most accurate region of the frame rather than a whole?” This will need to be investigated further in future work with the help of eye tracking to better understand how and which features participants value more when evaluating videos’ perceptual accuracy in a variety of different scenarios. This study was the starting point of a new research question regarding reflection impact on TMO performance which has shown that reflection does indeed have a negative impact. Future work will consider a more extensive set of state-of-the-art TMOs in order to verify if there are TMOs that can minimize these negative effects. As there is a wide spectrum of mobile devices with different screen features, an important variable that will also need to be taken into account is the impact of the absolute reflection index of the device and how can TMOs take advantage of it.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We would like to thank to the the participants for taking part in the study and to Elmedin Selmanovic for providing the Jaguar HDR footage. This work was partially supported by the Portuguese government, through National Foundation for Science and Technology - FCT (Fundação para a Ciência e a Tecnologia) for supporting this PhD through the grant SFRH/BD/76384/2011. I would also like to acknowledge the European Union (COMPETE, QREN and FEDER) for the partial support through the project REC I/EEI-SII 0360/2012 entitled “MASSIVE - Multimodal Acknowledgeable multiSensual Immersive Virtual Enviroments”. This work was also partially supported by the ICT COST Action IC1005 “HDRi: The digital capture, storage, transmission and display of real-world lighting”. Debattista and Chalmers are partially supported by a Royal Society Industrial Fellowship. This work was also partially supported by the ICT COST Action IC1005 “HDRi: The digital capture, storage, transmission and display of real-world lighting”.

Author details

¹ Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal. ²INESC-TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal. ³WMG, University of Warwick, Gibbet Hill Road, CV4 7AL Coventry, UK. ⁴goHDR Ltd., Gibbet Hill Road, CV4 7AL Coventry, UK.

Received: 6 May 2015 Accepted: 8 November 2015

Published online: 15 December 2015

References

1. H Seetzen, W Heidrich, W Stuerzlinger, G Ward, L Whitehead, M Trentacoste, A Ghosh, A Vorozcovs, High dynamic range display systems. *ACM Trans. Graph.* **23**(3), 760–768 (2004). doi:10.1145/1015706.1015797

2. IT Union, The World in 2013: ICT facts and figures (2011). <http://www.itu.int/ITU-D/ict/facts/material/ICTFactsFigures2011.pdf>. Accessed: 21/11/2013
3. CISCO, Cisco visual networking index: global mobile data traffic forecast update, 2012–2017 (2013). http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html. Accessed: 21/11/2013
4. Ooyala, *Global video index q3 2014*. (Report, International Telecommunication Union, Silicon Valley, USA, 2014). <http://go.ooyala.com/rs/OOYALA/images/Ooyala-Global-Video-Index-Q3-2014.pdf>
5. J Ross, R Simpson, B Simpson, Media richness, interactivity and retargeting to mobile devices: a survey. *Int. J. Arts Technol.* **4**(4), 442–459 (2011). doi:10.1504/IJART.2011.043443
6. SN Pattanaik, J Tumblin, H Yee, DP Greenberg, in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '00*. Time-dependent visual adaptation for fast realistic image display (ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000), pp. 47–54. doi:10.1145/344779.344810
7. R Mantiuk, S Daly, L Kerofsky, Display adaptive tone mapping. *ACM Trans. Graph.* **27**(3), 68–16810 (2008). doi:10.1145/1360612.1360667
8. M Melo, M Bessa, K Debattista, A Chalmers, Evaluation of Tone-Mapping Operators for HDR Video Under Different Ambient Luminance Levels. *Computer Graphics Forum.* **34**(8), 38–49 (2015). doi:10.1111/cgf.12606 <http://dx.doi.org/10.1111/cgf.12606>
9. F Banterle, A Artusi, K Debattista, A Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, 1st edn. (AK Peters, Ltd (CRC Press), Natick, MA, USA, 2011). <http://vcg.isti.cnr.it/Publications/2011/BADC11>
10. R Mantiuk, K Myszkowski, H-P Seidel, in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. Visible difference predictor for high dynamic range images (IEEE, New Jersey, USA, 2004), pp. 2763–2769
11. R Mantiuk, KJ Kim, AG Rempel, W Heidrich, Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* **30**(4), 40–14014 (2011). doi:10.1145/2010324.1964935
12. F Drago, W Martens, K Myszkowski, H-P Seidel, *Perceptual evaluation of tone mapping operators with regard to similarity and preference*. (Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Germany, 2002)
13. P Ledda, A Chalmers, T Troscianko, H Seetzen, in *ACM SIGGRAPH 2005 Papers. SIGGRAPH '05*. Evaluation of tone mapping operators using a high dynamic range display (ACM, New York, NY, USA, 2005), pp. 640–648. doi:10.1145/1186822.1073242
14. M Čadik, M Wimmer, L Neumann, A Artusi, Evaluation of HDR tone mapping methods using essential perceptual attributes. *Comput. Graph.* **32**(3), 330–349 (2008)
15. G Eilertsen, J Unger, R Wanat, R Mantiuk, in *ACM SIGGRAPH 2013 Talks*. Survey and evaluation of tone mapping operators for hdr video (ACM, New York, USA, 2013)
16. C Urbano, L Magalhães, J Moura, M Bessa, A Marcos, A Chalmers, Tone mapping operators on small screen devices: an evaluation study. *Comput. Graph. Forum.* **29**(8), 2469–2478 (2010). doi:10.1111/j.1467-8659.2010.01758.x
17. AO Akyüz, ML Eksert, MS Aydin, An evaluation of image reproduction algorithms for high contrast scenes on large and small screen display devices. *Comput. Graph.* **37**(7), 885–895 (2013). doi:10.1016/j.cag.2013.07.004
18. M Melo, M Bessa, K Debattista, A Chalmers, Evaluation of HDR video tone mapping for mobile devices. *Signal Process. Image Commun.* **29**(2), 247–256 (2014). doi:10.1016/j.image.2013.09.010 Special Issue on Advances in High Dynamic Range Video Research
19. JA Ferwerda, SN Pattanaik, P Shirley, DP Greenberg, in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '96*. A model of visual adaptation for realistic image synthesis (ACM, New York, NY, USA, 1996), pp. 249–258. doi:10.1145/237170.237262
20. R Boitard, K Bouatouch, R Cozot, D Thoreau, A Gruson, in *Applications of Digital Image Processing XXXV. Proc. SPIE*. Temporal coherency for video tone mapping, vol. 8499 (SPIE, San Diego, CA, USA, 2012), pp. 84990–8499010. doi:10.1117/12.929600
21. JH Van Hateren, Encoding of high dynamic range video with a model of human cones. *ACM Trans. Graph.* **25**(4), 1380–1399 (2006). doi:10.1145/1183287.1183293

22. TO Aydin, N Stefanoski, S Croci, M Gross, A Smolic, Temporally coherent local tone mapping of hdr video. *ACM Trans. Graph.* **33**(6), 196–119613 (2014). doi:10.1145/2661229.2661268
23. G Eilertsen, RK Mantiuk, J Unger, Real-time noise-aware tone mapping. *ACM Trans. Graph.* **34**(6) (2015). doi:10.1145/2816795.2818092
24. IT Union, *Methodology for the subjective assessment of the quality of television pictures*. (International Telecommunication Union, Geneva, Switzerland, 2012). https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-!!!PDF-E.pdf
25. M Čadik, M Wimmer, L Neumann, A Artusi, in *Proceedings of Pacific Graphics 2006 (14th Pacific Conference on Computer Graphics and Applications)*. Image attributes and quality for evaluation of tone mapping operators (National Taiwan University Press, Taipei, Taiwan, 2006), pp. 35–44. <http://www.cg.tuwien.ac.at/research/publications/2006/CADIK-2006-IAQ/>
26. M Narwaria, MP Da Silva, P Le Callet, R Pepion, in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. Single exposure vs tone mapped high dynamic range images: A study based on quality of experience (Springer International Publishing, Switzerland, 2014), pp. 2140–2144
27. Y Zhai, M Shah, in *Proceedings of the 14th Annual ACM International Conference on Multimedia. MULTIMEDIA '06*. Visual attention detection in video sequences using spatiotemporal cues (ACM, New York, NY, USA, 2006), pp. 815–824. doi:10.1145/1180639.1180824, <http://doi.acm.org/10.1145/1180639.1180824>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
