**RESEARCH**

# 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 Debattista[3] and Alan Chalmers[3,4]

**Abstract**

This paper presents an evaluation of HDR video tone-mapping on a small screen device (SSD) under reflections. Reflections are common on mobile devices as they are predominantly used on-the-go. With this evaluation we intend to study the impact of reflections on screen and how different HDR video TMO's 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 state-of-the-art 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 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 operators evaluation; Mobile Devices; Screen Reflections

*Correspondence:
emekapa@sapo.pt
[1]Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal
Full list of author information is available at the end of the article

## Introduction

The current imaging techniques, also known as Standard Dynamic Range (SDR) or Low Dynamic Range (LDR), are not capable of representing all the real-world colour 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 of 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 techniques known as 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 SSDs. TMOs for SSDs may need to take into account their portability due to situations where there is a sudden exposure to a wide variety of environmental variables such as widely differing luminance levels and reflections that could impact on the viewing experience.

Mobile devices have become very widespread and their penetration rate is reaching nearly 100% suggesting the use of 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 video [3]. In fact, a recent report [4] shows 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 there are projections that support this tendency and point towards 50% by the end of 2015. With this growing popularity it becomes important to consider possible challenges posed by mobile device displays compared to traditional desktop devices, such as limitations in terms of dynamic range, viewing angle and distance as well as size. In addition, to 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 requires the retargeting of content to play properly on any delivery screen. This can be a major challenge [5].

While previous works have addressed comparisons of diverse TMOs on mobile devices, none of them have 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 to understand their impact on the visualization quality. The insights gained from such a study should suggest if there are advantages in developing a hybrid TMO to account for reflections and minimise any negative impact on the viewing experience. The evaluation of the HDR video tone mappers is carried out with 3 different scenarios:

- with a reflection across the entire screen,
- with no reflections, and
- with a reflection on half 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 9.7 inch 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 experiments as, for example, [8].

## 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 developed based on simple mathematical operations such as exponential or logarithmic functions performing linear transformations as well as being more elaborate and inspired by 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 to map them as defined by the TMO. 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 screen as often happens when using mobile devices, especially outdoors. The next section presents a brief survey of previous methodologies to evaluate TMOs.

## 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 judgement over the contents being evaluated. Another key aspect is the proper randomisation 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 HDR display.

One of the first TMO psychophysical studies conducted was by Drago et al. that evaluated 7 different TMOs applied to 4 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 6 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 3 different scenes.

More recently evaluations started addressing HDR video, for example the work done by Eilertsen et al. [15] that evaluated 11 TMOs on a set of HDR videos that were both camera-captured or 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 colours and overall brighter image appearance was preferred.

Akyüz et al. also evaluated both TMOs and exposure fusion algorithms on SDDs [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 and as criteria the participants had to evaluate colour, 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 for bright lighting level environments. The paper concluded that participants gave more importance to contrast and naturalness over details and colour saturation in bright environments [8].

Tone mapping operators considered

As this study focuses on HDR video tone-mapping, only time-dependent TMOs were considered. The TMOs were selected based on the previous work by Melo et. al. [8]. From this study 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). The choice was made since the two TMOs can give insights on reflections 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 and as such Pat seems to produce more natural and contrasted images while Man produces more detailed and saturated images.

The display adaptive TMO is, as the name suggests, a TMO that is able to adapt according to the display features. The 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. After a proper calibration of the TMO, it 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 time-dependent visual adaptation TMO 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 adaption model based on Hunt's static model of colour vision and uses the retina response signals for rods for

calculating the luminance information and the response vector for colour information. In addition temporal coherency is added. The method is not computationally complex and thus is suitable to use in real-time applications.

## 1 Experimental setup

This section describes the experiments undertaken. In the experiments 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).

### 1.1 Method

An experimental framework that makes use of randomisation of the videos and TMO combinations was used in order to minimise 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. Fig. 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 3 independent variables: the set of TMOs used, the reflections on the display, and the scene groups. The TMOs and the reflections on the display were in-between participant independent variables; the scene groups was a within-participant variable.

The overall results were analysed using a 2 (scene groups) × 3 (reflections on display) × 4 (TMOs) mixed factorial ANOVA. The main effects calculated across all videos were of the group (each viewed 3 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 analysed using a 2 (scene groups) × 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 a combination of both. The reflections on the screen variable consisted of 3 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, dividing the video vertically in half and applying Man to one side and Pat to the other (referred from now on as ManPat), and by 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 in dividing participants in two groups 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 meters since the suggested viewing distance for high-definition displays is 3 times the height of the display (which was approximately 60 cm) according to the International Telecommunication Union Recommendation BT.500-13 [19]. A further concern was the luminance adaption 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.

## 1.2 Materials

The generic experimental setup for all the experimental scenarios is presented in Fig. 3. The "join" between the 2 TMOs was attenuated by a fade between the TMOs so there was no visible division. The photographic softbox was always placed so it was not between the participant and the reference and on 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, labelled "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 $cd/m^2$ 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.

Specific experimental software was devised for ranking the TMOs (Figure 4). For each video, the participants were able to see all of the combinations of TMOs applied simultaneously on the iPad and on the HDR display. 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 4 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 has finished.

The optimal TMO settings were determined by a group of three experts (who each had at least 2 years' experience in HDR imaging). The global version of Pat was used since the local version the time-dependent effects. The configurable settings for this TMO are the adaptation levels for cones and the adaptation levels for cones that were calculated for each case using the average luminance. For Man, when under reflections-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 $E_{amb} = 1450$.

### 1.3 Participants

A total of 90 participants, 51 men and 39 women aged between 19 and 28 years, were randomly assigned between the different experiment scenarios and between the six scenes (half of them evaluated a set of three videos and the other half the set containing the remaining three videos). 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.

### 1.4 Procedure

The participant stood at approximately 1.8 meters 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. Next, the evaluation software was shown to the participants so they could familiarise themselves with it. The average time for each experiment was around 10 to 15 minutes.

## 2 Results

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

As mentioned on 1.1, overall results were analysed using a using a 2 (scene groups) × 3(reflections on display) × 4(TMOs) mixed factorial ANOVA. The main effects calculated across all videos were of the group (each viewed 3 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 analysed using a 2(scene groups) × 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 amongst participants where W=1 signifies perfect agreement amongst participants and $W = 0$ complete disagreement. Table 3 summarizes the obtained results. We highlight the fact that despite the fact that Kendall's Coefficient of Concordance does not range high in values (it is between 0.14 and 0.24), it is significant for $p < 0.05$.

## 2.1 Overall results

For the different scenarios that were evaluated as well for the TMOs, the statistically significant difference is given by their main effect that was $F(6, 252) = 2.13(p < 0.05)$ and $F(2.93, 245.952) = 20.98(p > 0.05)$ respectively. Regarding the group main effect, the value reported was $F(3, 252) = 6.59(p > 0.05)$. 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 evaluated. 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 closer to 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.

## 2.2 Scenario 1 - reflections across 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 $F(2.64, 74.97) = 5.31$ with 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 $F(3, 84) = 2.55(p < 0.05)$. The Kendall's Coefficient of Concordance was significant at $W = 0.14$.

## 2.3 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 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.

The results reported for TMO main effect were $F(2.82, 81.23) = 2.55$ (Greenhouse-Geisser correction as sphericity was violated). The reported group main effect was not significant $F(3, 84) = 3.58(p < 0.05)$. The computed Kendall's Coefficient of Concordance was $W = 0.24(p < 0.05)$.

## 2.4 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 on 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). As before, there was no significant difference between the groups' rankings as the main effect was $F(3, 84) = 1.57(p < 0.05)$. The Kendal's Coefficient of Concordance on this scenario was $W = 0.19(p < 0.05)$.

## 3 Discussion

The main goal of the experiment was to study the impact of reflections on mobile devices displays and to provide some insights 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 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, what gives more strength to results. A first conclusion that can be extracted from 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.

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 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 SDD. Future work will consider the effect of reflection on mobile devices with different screen sizes.

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 that condition. This may indicate that when the display is fully exposed to reflections a TMO's perceptual accuracy can be compromised and the visualisation 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 which were $W = 0.24$ for this work against $W = 0.89$ 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.

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 rather than a plain usage of the second-ranked TMO. As in Scenario 2, this needs to be confirmed with eye tracking.

## 4  Conclusions

In this paper we set out to undertake a evaluation with a novel approach in order to study the impact of reflections on screen by understanding how different HDR video TMOs perform under reflective conditions. On top of that, this study intends to clarify if there is a need to develop a new or hybrid TMO that can deal with reflections better.

In the overall, 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. 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, ie 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 colour 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 that 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 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.

**Author details**
[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. [3]WMG, University of Warwick, Gibbet Hill Road, CV4 7AL Coventry, UK. [4]goHDR Ltd., Gibbet Hill Road, CV4 7AL Coventry, UK.

**References**
 1. Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., Vorozcovs, A.: High dynamic range display systems. ACM Trans. Graph. **23**(3), 760–768 (2004). doi:10.1145/1015706.1015797
 2. Union, I.T.: The World in 2013: ICT Facts and Figures. http://www.itu.int/ITU-D/ict/facts/material/ICTFactsFigures2011.pdf. Accessed: 21/11/2013 (2011)
 3. CISCO: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017. http://www.cisco.com/c/en/us/solutions/collateral/serv ice-provider/visual-networking-index-vni/white_paper_c11-520862.html. Accessed: 21/11/2013 (2013)
 4. Ooyala: Global video index q3 2014. Report, International Telecommunication Union, Silicon Valley, USA (2014)
 5. Ross, J., Simpson, R., Simpson, B.: Media richness, interactivity and retargeting to mobile devices: a survey. International Journal of Arts and Technology **4**(4), 442–459 (2011). doi:10.1504/IJART.2011.043443
 6. Pattanaik, S.N., Tumblin, J., Yee, H., Greenberg, D.P.: Time-dependent visual adaptation for fast realistic image display. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '00, pp. 47–54. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (2000). doi:10.1145/344779.344810
 7. Mantiuk, R., Daly, S., Kerofsky, L.: Display adaptive tone mapping. ACM Trans. Graph. **27**(3), 68–16810 (2008). doi:10.1145/1360612.1360667
 8. Melo, M., Bessa, M., Debattista, K., Chalmers, A.: Evaluation of Tone Mapping Operators for HDR video under different ambient luminance levels. To be published (2015)
 9. Banterle, F., Artusi, A., Debattista, K., Chalmers, A.: 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. Mantiuk, R., Myszkowski, K., Seidel, H.-P.: Visible difference predicator for high dynamic range images. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pp. 2763–2769 (2004)
11. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: 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. Drago, F., Martens, W., Myszkowski, K., Seidel, H.-P.: Perceptual evaluation of tone mapping operators with regard to similarity and preference. Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany (August 2002)
13. Ledda, P., Chalmers, A., Troscianko, T., Seetzen, H.: Evaluation of tone mapping operators using a high dynamic range display. In: ACM SIGGRAPH 2005 Papers. SIGGRAPH '05, pp. 640–648. ACM, New York, NY, USA (2005). doi:10.1145/1186822.1073242

14. Čadík, M., Wimmer, M., Neumann, L., Artusi, A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. Computers & Graphics **32**(3), 330–349 (2008)
15. Eilertsen, G., Unger, J., Wanat, R., Mantiuk, R.: Survey and evaluation of tone mapping operators for hdr video. In: ACM SIGGRAPH 2013 Talks (2013)
16. Urbano, C., Magalhães, L., Moura, J., Bessa, M., Marcos, A., Chalmers, A.: Tone mapping operators on small screen devices: An evaluation study. Computer Graphics Forum **29**(8), 2469–2478 (2010). doi:10.1111/j.1467-8659.2010.01758.x
17. Akyüz, A.O., Eksert, M.L., Aydin, M.S.: An evaluation of image reproduction algorithms for high contrast scenes on large and small screen display devices. Computers & Graphics **37**(7), 885–895 (2013). doi:10.1016/j.cag.2013.07.004
18. Melo, M., Bessa, M., Debattista, K., Chalmers, A.: Evaluation of hdr video tone mapping for mobile devices. Signal Processing: Image Communication **29**(2), 247–256 (2014). doi:10.1016/j.image.2013.09.010. Special Issue on Advances in High Dynamic Range Video Research
19. Union, I.T.: Methodology for the subjective assessment of the quality of television pictures. Recommendation bt.500-13, International Telecommunication Union, Geneva, Switzerland (2012)

figures/screens

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

figures/division

**Figure 2 Mobile device with half of the screen under reflection.**

figures/scenarios

**Figure 3 Experimental setup scheme for the 3 scenarios.** a) refers to a scenario where the same TMO is applied to the whole image and b) where 2 TMOs are applied to the same image.

figures/software

**Figure 4 Software used on the experiments.**

**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 | > 4 000 $cd/m^2$ | 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 (seconds) | Avg. DR (Log units) | Capture |
|---|---|---|---|
| CGRoom | 7 | 16.6 | CG |
| Jaguar | 13 | 13.4 | Canon 1D Mark II |
| Kalabsha | 11 | 18.5 | CG |
| Morgan Lovers | 15 | 19.5 | Spheron HDRv |
| Explosion | 8 | 12 | Canon 5D |
| Medical | 4 | 15.2 | Spheron HDRv |

**Table 3** Overall results obtained for each scenario. Coloured 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 | $1^{st}$ | $2^{nd}$ | $3^{rd}$ | $4^{th}$ |
|---|---|---|---|---|---|
| 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 halt of the screen | 0.19 | Pat | PatMan | ManPat | Man |