# Performance Evaluation of Reverse Tone Mapping Operators for Dynamic Range Expansion of SDR Video content

Hu Hao[1], Yang Zhang[1], Dimitris Agrafiotis[1], Matteo Naccari[2], Marta Mrak[2]

***Abstract*—When displaying Standard Dynamic Range (SDR) video on High Dynamic Range (HDR) displays a reverse tone mapping operation can be employed to expand the dynamic range of the SDR video to that offered by the display. This paper presents a subjective performance evaluation of existing reverse Tone Mapping Operators (rTMOs). The presented study evaluates the performance of rTMOs acting on both well exposed SDR video and video that exhibits exposure variation, as is often the case with user generated content (e.g. captured on mobile phones). The paper highlights flickering artefacts that arise in such cases and evaluates the effect on perceived quality of an adapted de-flickering method.**

## I. INTRODUCTION

High Dynamic Range (HDR) is one of the image features envisaged in the Ultra High Definition (UHD) format which will provide viewers with better pixels and enhanced quality of experience [1]. HDR video requires capture and display using HDR capable devices. Existing content (films, news, archive, home video) has been captured with Standard Dynamic Range (SDR) cameras. User generated content (e.g mobile phone captured video) is still being recorded primarily in SDR mode as HDR video would require multi-exposure techniques whereby the same picture is captured with a low and high exposure and then the two recordings are fused together. Displaying SDR on HDR displays can use reverse tone mapping in order to expand the dynamic range of the SDR sequence to the range offered by the display. With the increased availability of such displays this is a problem/opportunity that is of interest to both broadcasters and display manufacturers.

A number of reverse Tone Mapping Operators (rTMOs) have been recently proposed in the relevant literature including [2]–[13]. The aim of these rTMOs is to increase the brightness and contrast of conventional (SDR) content in order to display it on modern HDR displays. Several studies exist that evaluate the performance of existing rTMOs including [4], [5], [14]. The study conducted by Abebe et al. in [15], additionally explored the issue of colour reproduction when expanding the dynamic range of SDR images, but found that luminance reproduction is much more important. De Simone et al. [14] evaluated the fidelity of four rTMOs on SDR video sequences. The tested video sequences were well-exposed with no visible compression artefacts. It should be noted however, that typical user generated content may contain larger exposure variations which challenge the selected rTMO. Another issue to address when dealing with SDR to HDR video conversion is the fact that most of the rTMOs were originally designed for still images and their application on video is done on an independent frame by frame basis. This can lead to significant global changes of illumination between consecutive frames which in turn results in flicker.

In this context the novelty brought by this paper is twofold: on one hand it extends the study of [14] by including mobile phone captured content in subjective experiments. The subjective experiments are also extended by including the HDR reference material in order to assess the relative performance of existing rTMOs. On the other hand, the paper proposes a method to reduce flickering in the expanded HDR video which is based on the work of Guthier et al. [16] (originally developed for de-flickering tone-mapped HDR video). The presented studies rank the performance of existing methods and highlight the flickering effects that can arise when expanding the dynamic range of SDR video with exposure/brightness variations.

The remainder of the paper is structured as follows. Section II gives a brief review of related work. Section III presents the two subjective studies performed while Section IV describes the de-flickering method and its effect on the performance of the rTMOs. Section V concludes the paper.

## II. RELATED WORK

Reverse tone mapping operators can be broadly classified in two main categories: global rTMOs and local rTMOs. Global rTMOs expand the dynamic range of the input image/frame by applying the same expansion function to all pixels of the image. Akyuz et al. [4] evaluated different gamma valued curves for expanding the dynamic range of SDR images through two psychophysical experiments. The results of their study suggest that linear expansion is perceptually preferred. They also suggest that a simple dynamic range expansion function can yield good performance. In [5] Masia et al. found that simple linear expansion is only sufficient for well-exposed content. For over-exposed images with large areas of saturated pixels more sophisticated treatment may be required. They argue that an adaptive gamma curve should be applied to scale the dynamic range of an image through estimation of the image key. Recently, Masia et al. [6] proposed a multilinear model

[1]Department of Electrical and Electronic Engineering, University of Bristol, Bristol, UK. E-mail: yang.zhang@bristol.ac.uk
[2]British Broadcasting Corporation Research and Development, London, UK.
978-1-5090-3649-3/17/$31.00 ©2017 IEEE

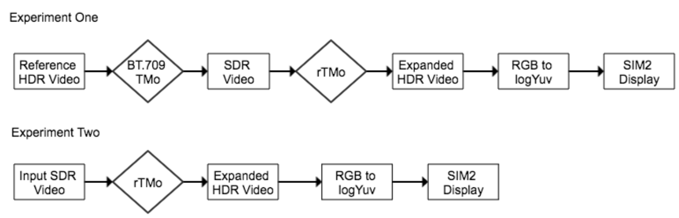



Fig. 1: Procedural diagram of our two subjective experiments.

for their gamma expansion function which outperforms their previous model for dynamic range expansion of underexposed images.

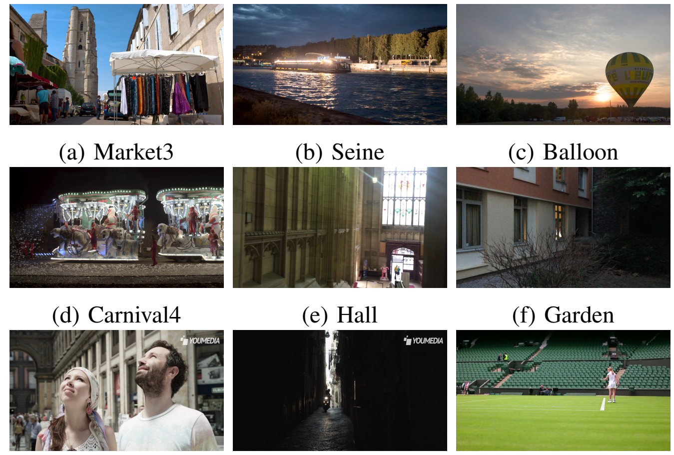
Local rTMOs compute an expansion map based on local/regional characteristics of the image/frame and use that to conditionally expand the dynamic range of pixels. Meylan et al. [7], [8] segment the input image using low pass filters and morphological operators before applying a piece-wise function to expand diffuse and specular regions differently. Their method has the advantage of being able to automatically determine the maximum diffuse threshold. Banterle et al. [2] employ the inverse of Reinhard's tone mapping operator [17] to expand the dynamic range of SDR images. They use a median cut algorithm to detect high luminance areas of the image and produce a so-called expand map by density estimation. This map is then used as a weighting factor for interpolating between the SDR image and an initial inverse tone-mapped image. Their framework is extended to support video content by employing temporal coherent density estimation [3]. Kovaleski and Oliveira [9] develop a bilateral grid to generate a Brightness Enhancement Function (BEF). In [10], they extend their work using a cross-bilateral step to compute smooth BEFs. Huo et al. [11] present a locally adaptive method based on a model of the HVS, whereby pixels in an SDR image are expanded using the inverse local retina response. Didyk et al. [12] design a system to manually classify the image into three regions - diffuse, reflection and light - with only the reflection and light regions being boosted. Wang et al. [13] proposed another solution to recover image details in over-exposed regions by extracting similar texture information from surrounding well-exposed areas. However, their approach as that of [12] is not automatic and requires manual input from users (e.g. segmentation).

## III. Performance evaluation of REVERSE TONE MAPPING OPERATORS

This section describes two psychophysics experiments conducted to assess the performance of the selected rTMOs and schematically represented in Figure 1.

### *A. Subjective experiment 1*

This experiment makes use of HDR reference video to evaluate the perceived quality of the expanded SDR content relative to that of the original HDR.

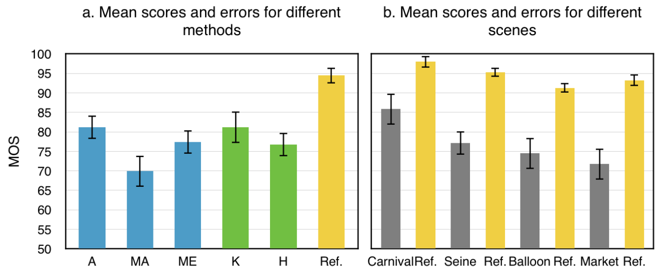


(a) Market3 (b) Seine (c) Balloon
(d) Carnival4 (e) Hall (f) Garden
(g) Naples (h) Vespa (i) Tennis

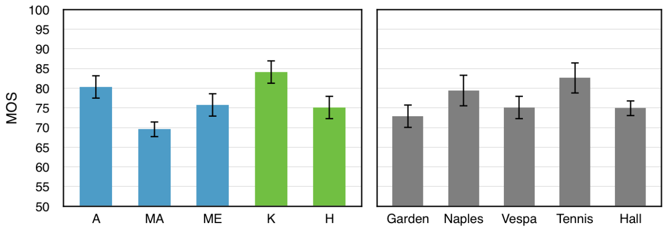
Fig. 2: Sample frames from the sequences used in our tests.

*1) Experimental setup:* Five rTMOs - three global and two local - were selected and evaluated in this subjective test as shown in the diagram of Figure 1. These were the following: the simple linear expansion of Akyuz et al. [4] (referred to as rTMO $A$); the multi-linear model based adaptive gamma expansion of Masia et al. [6] (rTMO $MA$); the brightness enhancement function of Kovaleski and Oliveira [10] (rTMO $K$); the retina response method of Huo et al. [11] (rTMO $H$); and the piecewise function expansion of Meylan et al. [8] (rTMO $ME$). The test material used comprises four sequences denoted as *Market3*, *Seine*, *Balloon* and *Carnival4*. The first three of these sequences, are sequences initially proposed for the MPEG ad hoc group (AHG) on support for HDR and wide colour gamut [18]. Sequence *Carnival4* is the sequence captured with a RED EPIC camera [19]. As can be seen in Figure 2, the test video sequences contain scenes ranging from bright to dark. The experiment was performed in a dark room, with all sequences displayed on a SIM2 HDR display (1920 $\times$ 1080 resolution and measured peak luminance of 2869 cd/m$^2$). The peak luminance value of the display is used as an input parameter to the rTMOs. The viewing distance for participants was set to 1.5 m (3 x screen height).

*2) Experimental procedure:* A total of 15 participants took part in this experiment. None of the participants had watched HDR content on an HDR display before. Five training presentations took place prior to the actual session in order to "stabilise" the viewers opinion as well as to familiarise them with the test procedure. Instructions about the assessment procedure and grading scale used were given at the beginning of the test. The SDR content at the input of each rTMO was produced by tone mapping HDR reference sequences using the gamma function defined in [20]. The choice of this particular tone mapping was motivated by our desire to emulate the tone mapping performed by SDR cameras/displays. The subjective study conducted followed the Double-Stimulus Continuous Quality-Scale (DSCQS) methodology, as recommended by ITU-R BT.500 [21]. Participants were shown only



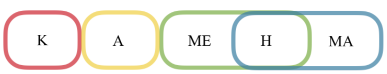

(a) First experiment




(b) Second experiment

Fig. 3: Mean opinion scores for the different operators and scenes. Error bars denote the 95% confidence interval.


(a) First experiment




(b) Second experiment

Fig. 4: Post-hoc analysis using Tukeys HSD. Any two operators within the same circle are statistically indistinguishable.

HDR stimuli, i.e. the HDR reference video and the expanded SDR versions, displayed on the SIM2 HDR display. Each participant watched a total of 5 video pairs twice (the HDR reference paired with each of the five expanded SDR versions) in randomised order. After each video pair presentation participants had to fill in a multiple choice questionnaire regarding their scores (e.g. *"why do you prefer video A or B?"* - *"video is brighter"*, *"obvious artefacts on video rated lower"*, etc. )

*3) Results and discussion:* The results of subjective Experiment 1 are shown in Figure 3a. Results are given in the form of Mean Opinion Scores (MOS) for all five tested rTMOs over the four test sequences. The differences in performance between the tested rTMOs were found to be significant by an ANOVA test ($p < 0.0001$). In terms of overall ranking, it can be observed that the reference (source) HDR outperforms all the rTMO-expanded SDR sequences (as was expected). The linear expansion method of Akyuz et al. [4] (rTMO $A$) has the highest overall performance across all four test sequences, with a mean score of 81.175. This supports previous claims by the authors of [4] that a simple dynamic range expansion operator can be perceptually preferred by viewers. We further performed post-hoc analysis using Tukeys HSD test that shows whether the performance of two methods with similar MOS is statistically indistinguishable or not. The results of this test are shown in Figure 4a at the 95 % significance level.

The gamma expansion rTMO of [6] (rTMO $MA$) is designed to cope with over-exposed content and as a result it seems to perform poorly with under-exposed sequences like *Seine*. The performance of all five rTMOs was found to be scene dependent. Higher average scores were observed for darker content - *Carnival4* and *Seine*. This observation is different from the findings of [15], but their study, unlike ours, involved images and not video. In our study viewers observed flickering artefacts in the brighter scenes which explains the lower average scores. The presence of flickering was partly due to tone mapping the reference HDR down to SDR, but mainly because of the expansion of the dynamic range of the SDR video which exacerbated and /or introduced flickering artefacts. Brighter scenes tend to have a greater span in dynamic range, with the maximum, minimum and mean frame luminance values varying a lot more from frame to frame compared to darker scenes. We discuss flickering artefacts further in Section IV.

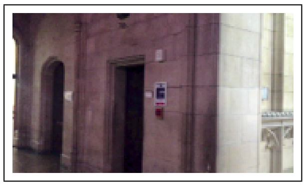
### B. *Subjective experiment 2*

In the second experiment we have no reference HDR video as the test sequences used were SDR clips. The aim of this experiment is to study the relative performance of the five tested rTMOs.

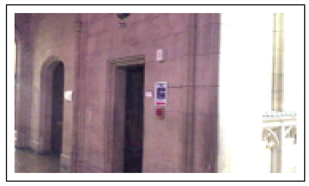
*1) Experimental Design:* Experiment 2 follows the Stimulus-comparison (SC) methodology described in [21]. General experimental arrangements, such as viewing conditions and assessment procedure were the same as in Experiment 1. The same subjects took part in our second experiment. Five SDR test sequences were used. Two of these sequences (*Hall* and *Garden*) were captured by the authors with a mobile phone. The remaining three (*Naples*, *Vespa* and *Tennis*) are the same HD sequences used in the study of [14]. Participants watched 20 video pairs for each test sequence ($n \times (n-1)$, where $n$ is the number of tested rTMOs). A break was offered to all participants every ten minutes.

*2) Results and Discussion:* Figure 3b shows the results of Experiment 2 in the form of mean opinion scores for all test sequences and rTMOs. It can be observed that the methods proposed by Kovaleski and Oliveira et al. [10] - rTMO $K$ - and Akyuz et al. [4] - rTMO $A$ - have the highest MOS (scores of 84.14 and 80.27 respectively). The analysis of variance (ANOVA) that was performed confirmed that a significant difference exists between the performance of the different rTMOs ($p < 0.0001$). Post-hoc analysis (Figure 4b) showed that the performance advantage of rTMO $K$ was statistically distinguishable. The simple linear expansion, rTMO $A$, offers very good performance.

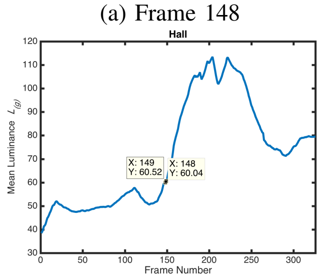
As expected, video sequences shot using professional cameras (*Naples*, *Vespa* and *Tennis*) have higher scores than those captured by mobile devices (*Garden* and *Hall*). Viewers



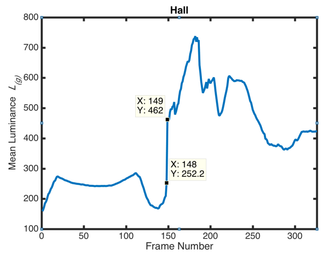
(a) Frame 148



(b) Frame 149




(c) Frame Luminance trace of original SDR




(d) Frame Luminance trace of Expanded SDR

Fig. 5: Frames 148 and 149 from test video, *Hall* expanded by rTMO *K* (top). Mean frame luminance trace of the original SDR version of *Hall* (bottom left) and the expanded SDR (bottom right).

reported that several artefacts including white noise, coding noise and flickering were present in the expanded version of the two mobile phone sequences *Garden* and *Hall*. Out of these two mobile phone video sequences, the darker one (*Garden*) received lower MOS compared to *Hall*. This is in line with the findings of [14], where it was suggested that rTMOs amplify noise present in the video more in the case of dark SDR input, making any existing artefacts even more visible when displayed on an HDR monitor.

The two mobile phone sequences additionally suffered from flickering which contributed to the lower scores. Flicker in these two sequences is mainly due to the autoexposure function of the mobile phone camera and although not obvious on the mobile screen, reverse tone mappers enhance it and make it visible on the HDR display. As stated in the Introduction, this flickering appears mainly because of the rTMOs used which were mainly designed for images. Figure 5 illustrates the problem with sequence *Hall*. The mean luminance of the expanded SDR video increases significantly from frame 148 (mean luminance of 255) to frame 149 (mean luminance of 462). The quick luminance variation happens as the camera pans horizontally from a dark scene to a bright scene (a large mean luminance drop takes place from frame 186 to 187 as the camera pans back to the starting dark scene). The brightness change can be easily seen on the displayed frames.

## IV. FLICKER REDUCTION

The two subjective studies presented in the previous section have highlighted the need for the reverse tone mapping operation to consider de-flickering, especially when the original SDR video exhibits changes in illumination. Work done in [16] suggested a method for detecting and removing flicker from tone mapped HDR video sequences (HDR material converted to SDR). The same method is followed here for removing flicker from the expanded SDR video sequences. The method is briefly described in the following.

The de-flickering method makes uses of the mean luminance of a frame. We use the geometric mean luminance as this reduces outliers [22]. Geometric mean luminance is given by the following equation:

$$L_{(g)} = \exp(\frac{1}{N}\sum_{i,j=1}^{N} log(L_{i,j}+\varepsilon)), \tag{1}$$

where $N$ denotes the number of pixels in an image, $L(i,j)$ represents the luminance value of pixel with coordinate $(i,j)$, and $\varepsilon$ is a small positive value to avoid singularity.

Equation (2) describes the function used to detect and remove flickering. The function follows Stevens's power law [23] which describes the relationship between background stimuli and minimum noticeable difference:

$$\psi(I_t) = k \cdot (I_{(t-1)})^{\alpha}, \tag{2}$$

where $I_{(t-1)}$ represents the geometric mean luminance of the previous frame, $\psi(I_t)$ is the just noticeable difference threshold (brightness changes that can be perceived by the HVS). The value of the exponent is defined as 0.33 for brightness stimuli. The proportionality constant $k$ has to be determined. Following subjective tests that we performed with a few expert viewers we set $k = 1.21$. The value of $\psi(I_t)$ acts as flicker detector. If the difference in mean luminance between the current and previous frame exceeds $\psi(I_t)$, then flickering is likely to be observed in the current frame. In this case, the mean luminance of the current frame will be adjusted to the nearest tolerable border, for example modify to $I_{(t-1)} \pm \psi(I_t)$. The ratio between modied luminance and real luminance should be obtained, and then apply the ratio to all the pixels in the frame that suffers flicker. Larger variation could happen, as shown in Figure 5, the variation of the frame luminance exceed the value far more than $I_t$. Flicker is observed in both Exp.1 and Exp.2 and we proved it is the most obtrusive artefact for expanded HDR videos, and we noticed the frame luminance upraise or drop at the moment when flicker happens.

To test the effect that the above de-flickering approach has on the perceived quality of expanded SDR video we performed a subjective experiment using sequences *Market*, *Hall*, *Garden*, and *Balloon* (same participants and setup as in Experiment 2). These sequences were flagged as exhibiting flickering by the participants of our previous experiments. We tested the performance of the two rTMOs that performed best in our previous experiments (rTMOs $K$ and $MA$) with and without de-flickering post-processing. The mean opinion scores given in Figure 6 show that de-flickering improves the performance of both rTMOs (the differences between the four processing methods were found to be statistically significant). Post-hoc analysis (Figure 7) shows that this difference is statistically distinguishable. Mean luminance traces for the four tested

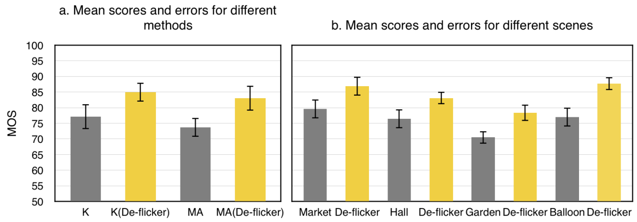



Fig. 6: Mean opinion scores for expanded SDR with and without de-flickering. Left: average for all sequences. Right: individual sequences


Fig. 7: Post-hoc analysis of de-flickering results

sequences (*Market*, *Hall*, *Balloon* and *Garden*) before and after de-flickering are shown in Figure 8.

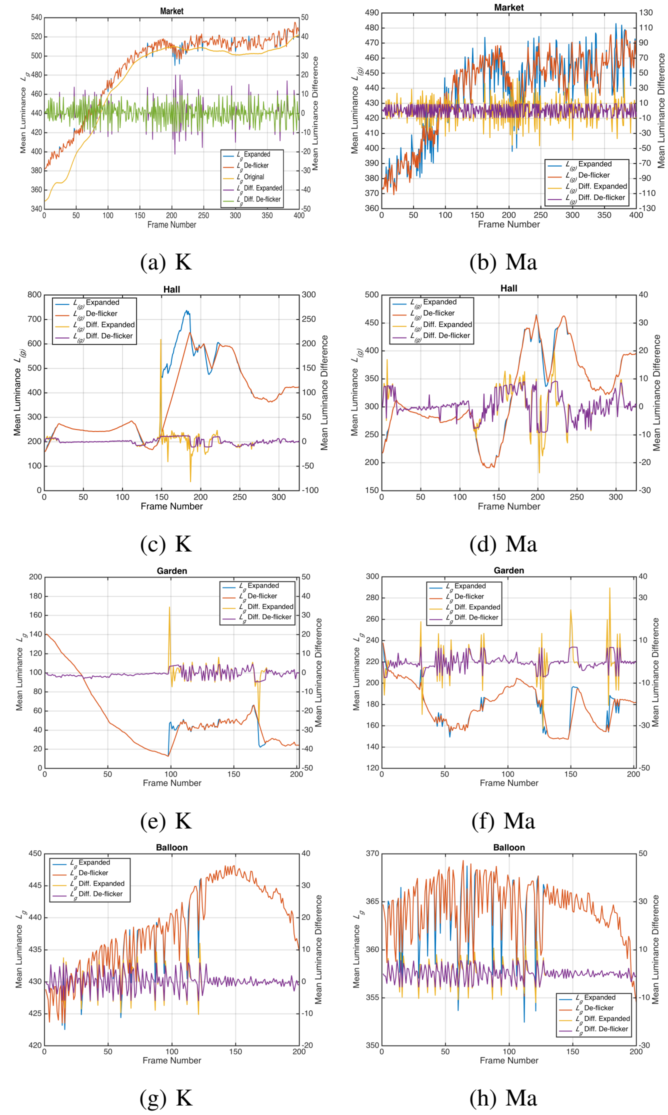



Fig. 8: Mean luminance frame trace for *Market*, *Hall*, *Garden* and *Balloon* with and without de-flickering

## V. Conclusion

This paper presented two subjective studies to evaluate the performance of state-of-the-art rTMOS, used for SDR to HDR conversion. Our tests extended previous works done in the area by including some user generated content which exhibit larger exposure variation. Our subjective tests evaluated five representative rTMOs and have found the rTMO of [10] to offer the best performance. The simple linear rTMO of [4] was found to also offer very good performance. Moreover, our studies have highlighted flickering artefacts that can arise when rTMOs are applied to SDR content of varying exposure. These can have a significant effect on the perceived quality of expanded SDR video. To address this issue, we adapted the method of [16] (originally developed for de-flickering tone-mapped HDR video) for de-flickering expanded SDR video. Subjective tests that we performed showed that this process brings significant perceived quality benefits.

## References


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

[2] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse Tone Mapping," in *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, New York, NY, USA, 2006, GRAPHITE '06, pp. 349–356, ACM.

[3] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Expanding Low Dynamic Range Videos for High Dynamic Range Applications," in *Proceedings of the 24th Spring Conference on Computer Graphics*, New York, NY, USA, 2010, SCCG '08, pp. 33–41, ACM.

[4] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do HDR Displays Support LDR Content?: A Psychophysical Evaluation," in *ACM SIGGRAPH 2007 Papers*, New York, NY, USA, 2007, SIGGRAPH '07, ACM.

[5] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of Reverse Tone Mapping Through Varying Exposure Conditions," in *ACM SIGGRAPH Asia 2009 Papers*, New York, NY, USA, 2009, SIGGRAPH Asia '09, pp. 160:1–160:8, ACM.

[6] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools and Applications*, pp. 1–18, Nov. 2015.

[7] L. Meylan, S. Daly, and S. Süsstrunk, "The Reproduction of Specular Highlights on High Dynamic Range Displays," *Color and Imaging Conference*, vol. 2006, no. 1, pp. 333–338, Jan. 2006.

[8] L. Meylan, S. Daly, and S. Süsstrunk, "Tone mapping for high dynamic range displays," 2007, vol. 6492, pp. 649210–649210–12.

[9] R. P. Kovaleski and M. M. Oliveira, "High-quality brightness enhancement functions for real-time reverse tone mapping," *The Visual Computer*, vol. 25, no. 5-7, pp. 539–547, Mar. 2009.

[10] R. P. Kovaleski and M. M. Oliveira, "High-Quality Reverse Tone Mapping for a Wide Range of Exposures," in *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug. 2014, pp. 49–56.

[11] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *The Visual Computer*, vol. 30, no. 5, pp. 507–517, Sept. 2013.

[12] P. Didyk, R. Mantiuk, M. Hein, and H.P. Seidel, "Enhancement of Bright Video Features for HDR Displays," *Computer Graphics Forum*, vol. 27, no. 4, pp. 1265–1274, June 2008.

[13] L. Wang, L. Wei, K. Zhou, B. Guo, and H. Shum, "High Dynamic Range

Image Hallucination," in *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, Aire-la-Ville, Switzerland, Switzerland, 2007, EGSR'07, pp. 321–326, Eurographics Association.

[14] F. De Simone, G. Valenzise, P. Lauga, F. Dufaux, and F. Banterle, "Dynamic range expansion of video sequences: A subjective quality assessment study," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2014, pp. 1063–1067.

[15] M. A. Abebe, T. Pouli, and J. Kervec, "Evaluating the Color Fidelity of ITMOs and HDR Color Appearance Models," *ACM Trans. Appl. Percept.*, vol. 12, no. 4, pp. 14:1–14:16, Sept. 2015.

[16] B. Guthier, S. Kopf, M. Eble, and W. Effelsberg, "Flicker reduction in tone mapped high dynamic range video," 2011, vol. 7866, pp. 78660C–78660C–15.

[17] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic Tone Reproduction for Digital Images," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 2002, SIGGRAPH '02, pp. 267–276, ACM.

[18] S. Lasserre, F. LeLéannec, and E. Francois, "Description of HDR sequences proposed by Technicolor," *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228], IEEE, San Jose, USA*, 2013.

[19] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High Dynamic Range Video Compression Exploiting Luminance Masking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 950–964, May 2016.

[20] ITU-R, "Parameter values for the HDTV standards for production and international programme exchange," June 2015.

[21] ITU-R, "Methodogy for the subjective assessment of the quality of television pictures," Jan. 2012.

[22] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, CRC Press, Feb. 2011, Google-Books-ID: LL5orppYlJsC.

[23] J. C. Stevens and S. S. Stevens, "Brightness Function: Effects of Adaptation*," *Journal of the Optical Society of America*, vol. 53, no. 3, pp. 375, Mar. 1963.