Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
23rd Meeting: San Diego, USA, 19–26 February 2016

Document: JCTVC-W0106

Title:	**Evaluation of Backward-compatible HDR Transmission Pipelines**
Status:	Input Document to JCT-VC
Purpose:	Informative Document

Author(s) or Contact(s): Maryam Azimi[1], Ronan Boitard[1], Mahsa T. Pourazad[2,3], and Panos Nasiopoulos[1]

Tel: (604) 827-4878
Email: maryama@ece.ubc.ca
rboitard.w@gmail.com
mahsa.pourazad@gmail.com
panos@ece.ubc.ca

Source: [1]University of British Columbia (UBC), [2]TELUS Communications Inc., [3] The Institute for Computing, Information and Cognitive Systems (ICICS)

Abstract

In this work, we report the performance of two possible scenarios for distributing HDR content employing a single layer. One of them compresses HDR content (HDR10), with color conversion and tone mapping performed at the decoding stage to generate SDR, while the other scenario compresses tone-mapped SDR content with inverse tone-mapping and color conversion at the decoding side to generate HDR. We performed subjective tests to evaluate the viewing quality of the SDR videos on an SDR display. Note that previous work had revealed that the HDR10 approach had resulted in better HDR visual quality. The results showed that chroma subsampling is playing a very important role in determining the final visual quality of the SDR content. Several Tone Mapping Operators (TMOs) are tested. A non-invertible TMO, known to yield very high quality SDR, is shown to produce very good SDR results for the HDR10 pipeline.

1 Introduction

Several distribution schemes have been proposed to address both HDR and SDR [1]. These different distribution techniques can be classified as follows:

(a) Compression of HDR content (HDR10); color and tone mapping are performed at the decoding stage to generate SDR; additional metadata can be derived to improve the quality of the color and tone mapping,

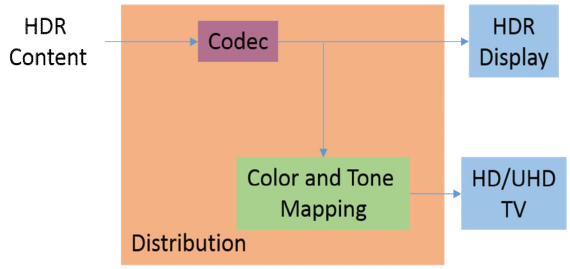
(b) Compression of SDR content; inverse color and inverse tone mapping are performed at the decoding stage to generate HDR; additional metadata can be derived to improve the quality of the inverse color and inverse tone mapping,

(c) Compression of HDR and SDR separately, denoted as simulcast (two separate single layers),

(d) Joint compression of HDR/SDR content, denoted as scalable.

Note that color mapping in our experiments only corresponds to converting BT.2020 content to BT.709 to address the display in SDR mode.

In [2], we compared the quality of HDR content in scheme (a) (HDR Single layer) with that of scheme (b) (SDR Single layer) using an HDR display (i.e., SIM2, 6,000 nits). In this informative document, we compare the SDR performance of schemes (a) and (b) using an SDR display (i.e., Samsung SUHDTV UN65JS9500). Schemes (a) and (b) are visualized in Figure 1-(a) and Figure 1-(b), respectively. The goal of this study is to report if a single layer backward-compatible pipeline can be employed with acceptable quality for both HDR and SDR content. In what follows we refer to scheme (a) as HDR10 and scheme (b) as the SDR scheme.



(a) (b)

Figure 1: (a) Distribution of HDR content and derivation of SDR content at the display stage, (b) distribution of SDR content and derivation of HDR content at the display stage.

2 Content

In this work we employed 4 HDR sequences included in the Call for Evidence (CfE) [3]: BalloonFestival, FireEater2, Market3, and Tibul2.

For the HDR10 scheme, the original HDR sequences of images (OpenEXR) have been perceptually encoded using the SMPTE ST 2084 [4] (PQ) with non-constant luminance and were quantized on 10 bits (using BT.1361, restricted range) after being chroma subsampled to 4:2:0 (see Figure 2) in order to generate HDR10 sequences. Note that in the Common Test Conditions (CTC) and CfE, the quantization is performed before the chroma downsampling.

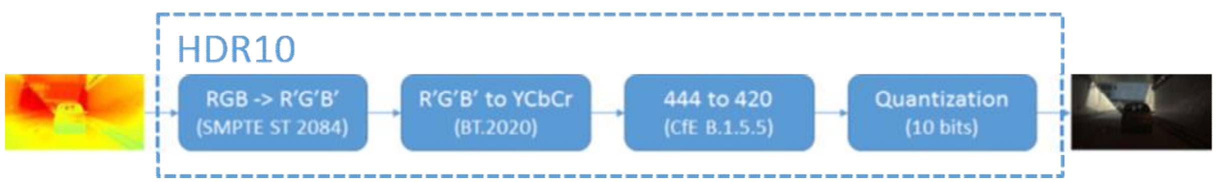


Figure 2: Preprocessing of original HDR source to obtain HDR10 content (scheme HDR10).

To obtain SDR versions of the test content for the SDR pipeline, we tone-mapped the HDR sequences using three different Tone Mapping Operators (TMOs): Mai et al. operator [5], Photographic Tone Reproduction (PTR) in its global version (Reinhard et al. [6]) and the Weighted-Least Square (WLS) (Farbman et al. [7]). We chose Mai and PTR, since they provide the best results in terms of MOS and coding bitrate, when viewed on SIM2 [2]. WLS is a local operator and by definition is not invertible (cannot reconstruct the HDR content in the SDR pipeline). However, we chose to add it to this test as it provides high visual quality SDR videos and can be a reference point for comparison. The snapshots of different TMO results for each sequence are provided in Annex A.

Figure 3: Preprocessing of original HDR content to obtain SDR sources (scheme SDR).

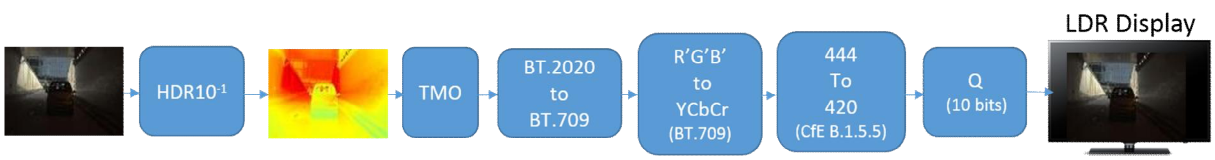
Once the SDR version of the content is generated, it is converted to Y'CbCr and after having been chroma downsampled to 4:2:0, it is quantized to 10 bits (see Figure 3).

3 Test Scenario

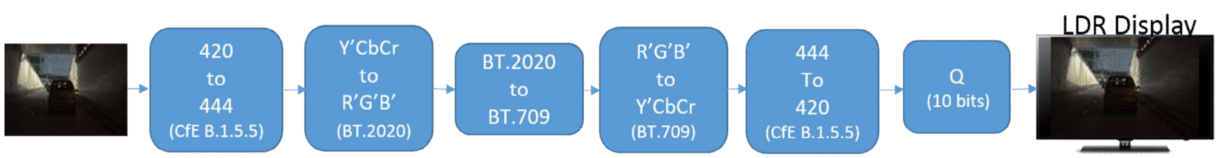
All subjective experiments reported in this document were performed on a 65" Samsung SUHDTV UN65JS9500 series 9 display, with 3840x2160 resolution, 1,000 nits peak luminance, and P3 color gamut.

We differentiate two different approaches to address this display:

1. **Y'CbCr 10-bit 4:2:0 in HEVC Main10 bit-streams**: content has been encoded using HEVC main 10 at QP0. In that case, the up-sampling from 4:2:0 to 4:4:4 is performed in the TV and is, thus, out of our control. Figure 4 (a) shows the workflow of the HDR10 pipeline while Figure 4 (b) shows the workflow of the SDR pipeline to address the display using this method.

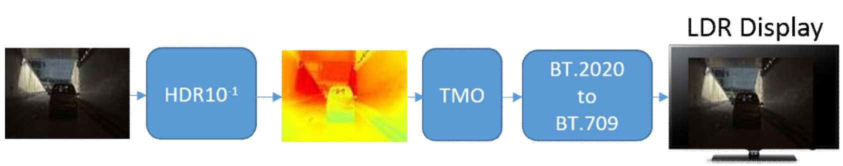


(a)

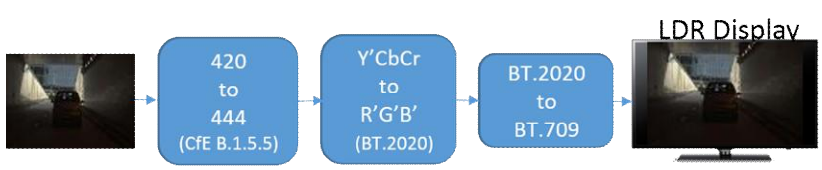


(b)

Figure 4. Workflow of (a) HDR10, and (b) SDR to address an SDR display (post processing) input: Y'CbCr 10-bit 4:2:0 in HEVC Main10 bit-streams



(a)



(b)

Figure 5. Workflow of (a) HDR10, and (b) SDR to address an SDR display (post processing) input: R'G'B'10-bits 4:4:4

2. **R'G'B'10-bits 4:4:4**: 10 bits content stored in a 16-bit Tiff file and played back using the scratch player. Figure 5 (a) shows the workflow of the HDR10 pipeline and figure 5 (b) shows the SDR pipeline to address the display using this method.

Note that in both approaches, every other frame of the sequence Market3 was dropped to avoid slow motion while playing at 25fps (original sequence at 50fps). In order to reproduce the same test conditions as in [2], we chose to address the display with a video signal of HD resolution (1920x1080). Thus, we cropped each sequence along the vertical axis (see Annex B for details).

The QPs used to encode the Market3, FireEater2, Tibul2 and BalloonFestival videos are reported in Annex B, Table I.

4 Impact of the HDR10 and SDR pipelines on SDR visual quality using chroma sampling for SDR withwith reference to the original quality reference

In our first experiment, we propose to assess the visual quality of decoded sequences compared to that of the original tone-mapped (SDR) video. One side of the display was the original (uncompressed) SDR video. The other side was the decoded test video. Both the original and the test videos were tone mapped with the same TMO.

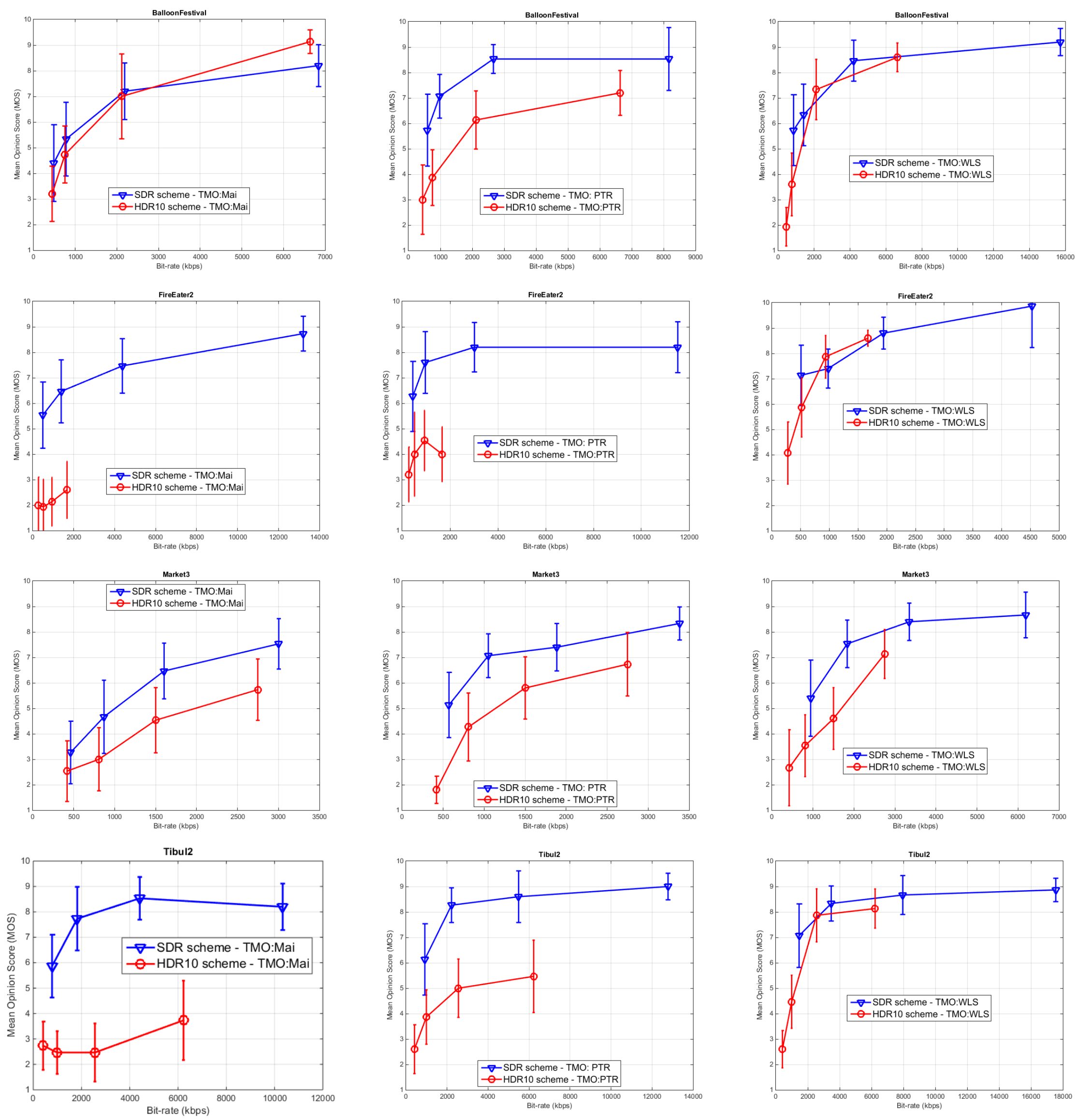


Figure 6. MOS-rate comparison of videos with HDR10 and SDR pipelines for BalloonFestival, FireEater2, Market3, and Tibul2 sequences, input: Y’CbCr 10-bit 4:2:0 in HEVC Main10 bit-streams.

The subjects were asked to evaluate the distortion of the decoded content by comparing it to that of the original. A scale from 1 to 10 was used to assess the distortion with score 1 referring to very annoying and 10 representing imperceptible distortion. A training session was organized before the actual experiment to describe compression artifacts. The subjects rated 96 test streams (4 sequences × 4 QPs × 3 TMOs × 2 schemes = 96), which were randomly ordered so that two same sequences would not be shown following each other.

The Mean Opinion Score (MOS) for each test video was calculated by averaging the scores over all the subjects with 95% confidence interval. Figure 6 shows the results in MOS values versus the required bitrate for encoding the videos using the HDR10 and SDR pipelines. As it can be seen, SDR yields mainly higher subjective quality compared to HDR10 scheme.

The SDR videos resulted from HDR10 pipeline scheme, has been through two steps of chroma downsampling in two different domains (one before the codec on HDR10 data and one after it on tone mapped data to address the display), while in the SDR pipeline both chroma subsampling steps were performed in the tone mapped domain. Note that for the SDR pipeline, upsampling and downsampling was performed in order to convert BT.2020 to BT.709. In most cases, these two additional steps would not need to occur that is to say if the gamut conversion is performed before the compression. However, these two steps are mandatory for the HDR10 pipeline, since tone mapping is applied on RGB 4:4:4 values.

The two subsampling steps in two different domains (HDR10 and SDR) may cause chroma artifacts and in turn noticeable visual artifact on the viewed SDR video. This may be the reason why HDR10 pipeline is reportedly performing lower than the SDR one. In order to check this assumption and evaluate the visual effect of applying two subsampling steps in two different domain, we prepared the HDR10 content in R'G'B' 4:4:4 format (see Figure 5), bypassing the internal (display) chroma upsampling. Figure 8 shows an example of the chroma artifacts due to the additional subsampling process.

We conclude that by eliminating the second chroma subsampling process in the HDR10 pipeline, the visual quality of the decoded SDR videos increases. Although at this point in time this may not be possible as many displays are addressed using Y'CbCr 4:2:0, we believe that designing a better subsampling filter may drastically reduce these artifacts.

5 Impact of the HDR10 and SDR pipelines on the SDR visual quality without reference to the original quality

Our first experiment hinted that the SDR pipeline might be more efficient than the HDR10 one. However, in that experiment we were evaluating the impact of video coding on SDR video. However, subjects reported that in many cases it was complicated to evaluate the fidelity since the viewed quality of some of the tone mapping operators was below acceptable quality to begin with (please refer to Annex 1 for snapshots of tone mapped videos).

That is why we conducted a second experiment where the subjects were asked to rate the quality of the test videos without any reference to the original video. We addressed the display using the R'G'B' 4:4:4 format (see Figure 5), thus avoiding the second chroma subsampling step. Similar to the previous test (Section 4), 96 test videos were presented to the subjects along with the original video. Again, a discrete rating scale from 1 (worst quality) to 10 (best quality) was given to the subjects. The Scratch player [9] was used to view the contents in 10 bits.

It is observed from the reported results in Figure 7 that regardless of the pipeline used to transmit the videos, the Mai and PTR TMOs achieved mostly MOS values below 6, even at high bitrates. We could conclude that for acceptable visual quality these two TMOs are not recommended for supporting SDR backward compatibility. However, the WLS TMO on average achieves an acceptable subjective quality regardless of the bitrate.

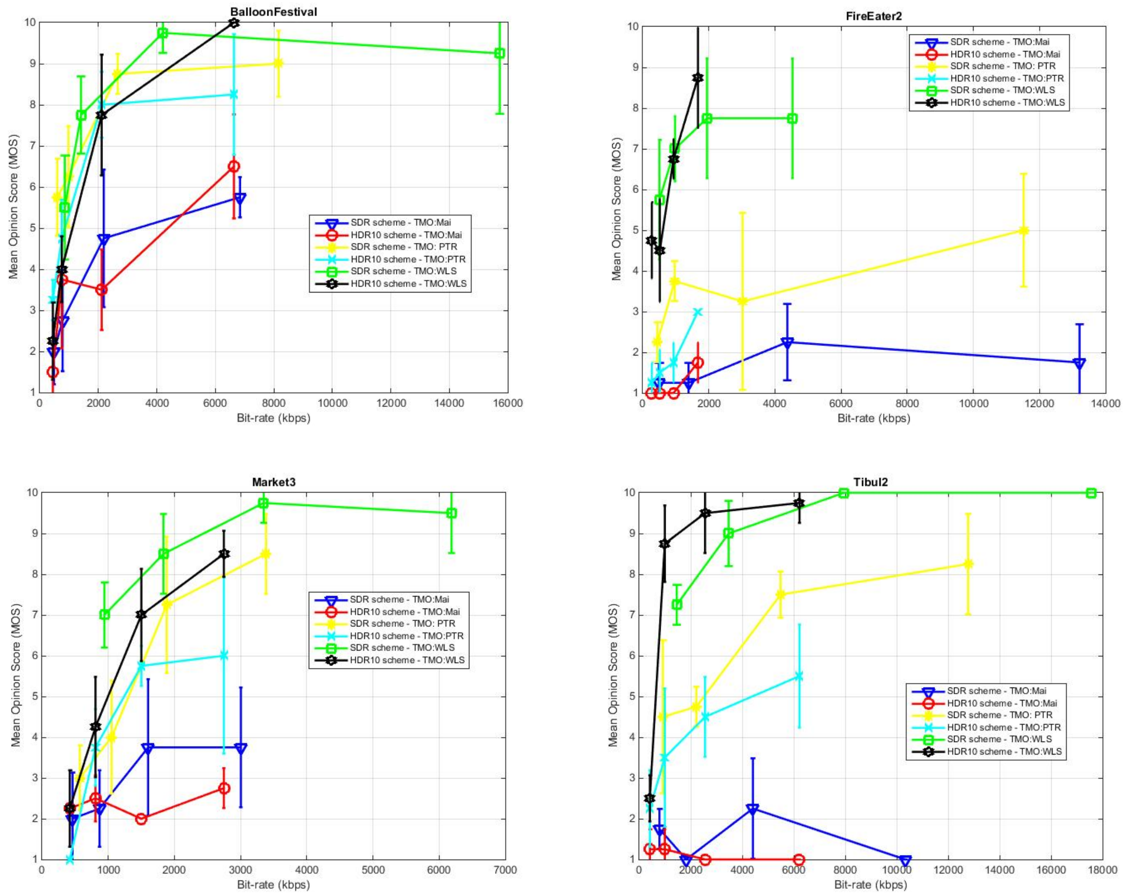
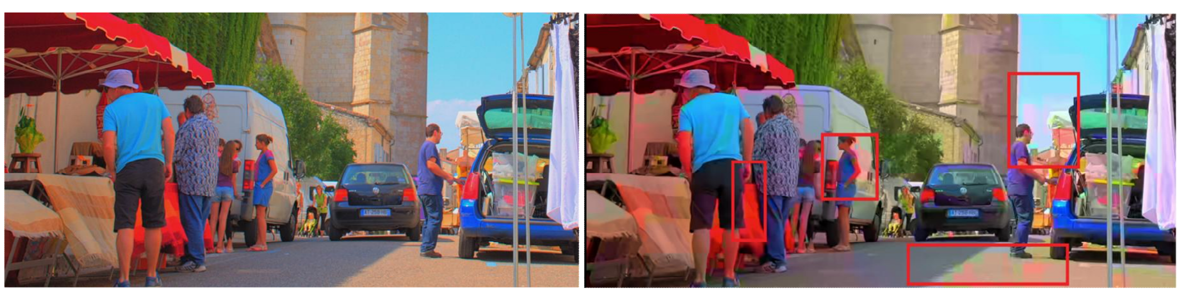


Figure 7. MOS-rate comparison of videos with scheme 1 and 2 for BalloonFestival, FireEater2, Market3, and Tibul2 sequences, input: R’G’B’10-bits 4:4:4



(a)

(b)

Figure 8. (a) Market3 with Mai TMO with additional subsampling process Y’CbCr 10-bit 4:2:0 in HEVC Main10 bit-streams versus the original on the left (b) Market3 with WLS TMO with additional subsampling process Y’CbCr 10-bit 4:2:0 in HEVC Main10 bit-streams versus the original on the left

6 Conclusion and Recommendations

In this work, we compared two distribution pipelines for delivering HDR content to SDR displays. In our first experiment, results showed that the distortion resulting from the HDR transmission scheme was higher than that of the SDR scheme. We showed that the main cause for this outcome is the application of two chroma subsampling steps in the two different domains (HDR10 and SDR tone-mapped). By removing one of the subsampling processes in the HDR10 transmission scheme, chroma artifacts were greatly reduced. These experiments indicate that the current chroma subsampling process is not adequate and needs to be improved.

Another test indicated that TMOs that were performing well reproducing HDR did not produce pleasantly viewable SDR, despite the fact that coding introduced negligible distortion. This test also indicated that using the HDR transmission pipeline allows the flexibility to implement non-invertible TMOs at the decoder side, which can greatly improve the visual quality of the resulting SDR.

Combining the findings of this contribution with those from [2] we can conclude that in the case of transmitting a backward-compatible single layer signal, the HDR10 pipeline is preferable over that of SDR. The reasons behind this conclusion are the following:

- Transmitting HDR10 reportedly introduces **less** distortion compared to SDR10 when addressing an HDR display,
- Transmitting HDR10 reportedly introduce **more** distortions compared to SDR10 when addressing an SDR display for a given TMO, however the performance of the TMOs is important as using a non-invertable TMO results in a high quality content.
- Transmitting HDR10 allows flexibility on the choice of the TMO performed at the decoder stage, thus non-invertible TMOs that can achieve higher quality SDR results can be used. In this case the HDR10 pipeline offers better HDR quality than the SDR pipeline (previous contribution) and comparable SDR quality to the SDR pipeline.

Finally, results reconfirm the importance of an efficient subsampling process.

7 References

[1] R. Boitard, M. T. Pouzarad, and P. Nasiopoulos, J. Slevinsky, "Demystifying High-Dynamic-Range Technology: A new evolution in digital media," in Consumer Electronics Magazine, IEEE , vol.4, no.4, pp.72-86, Oct. 2015.

[2] M. Azimi, R. Boitard, M. T. Pouzarad, and P. Nasiopoulos, " Evaluation of PQ versus tone-mapping for single layer HDR video compression," in ISO/IEC JTC1/SC29/WG11 MPEG2015/ M37318, 2015.

[3] A. Luthra, E. Francois, and W. Husak, “Call for Evidence (CfE) for HDR and WCG Video Coding,” in ISO/IEC JTC1/SC29/WG11 MPEG2015/N15083, 2015.

[4] Society of Motion Picture & Television Engineers, "High Dynamic Range Electro- Optical Transfer Function of Mastering Reference Displays Table," in SMPTE ST 2084, 2014.

[5] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward and W. Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression," Image Processing, IEEE Transactions on, vol. 20, pp. 1558-1571, 2011.

[6] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, vol. 21, pp. 267-276, July 2002.

[7] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.

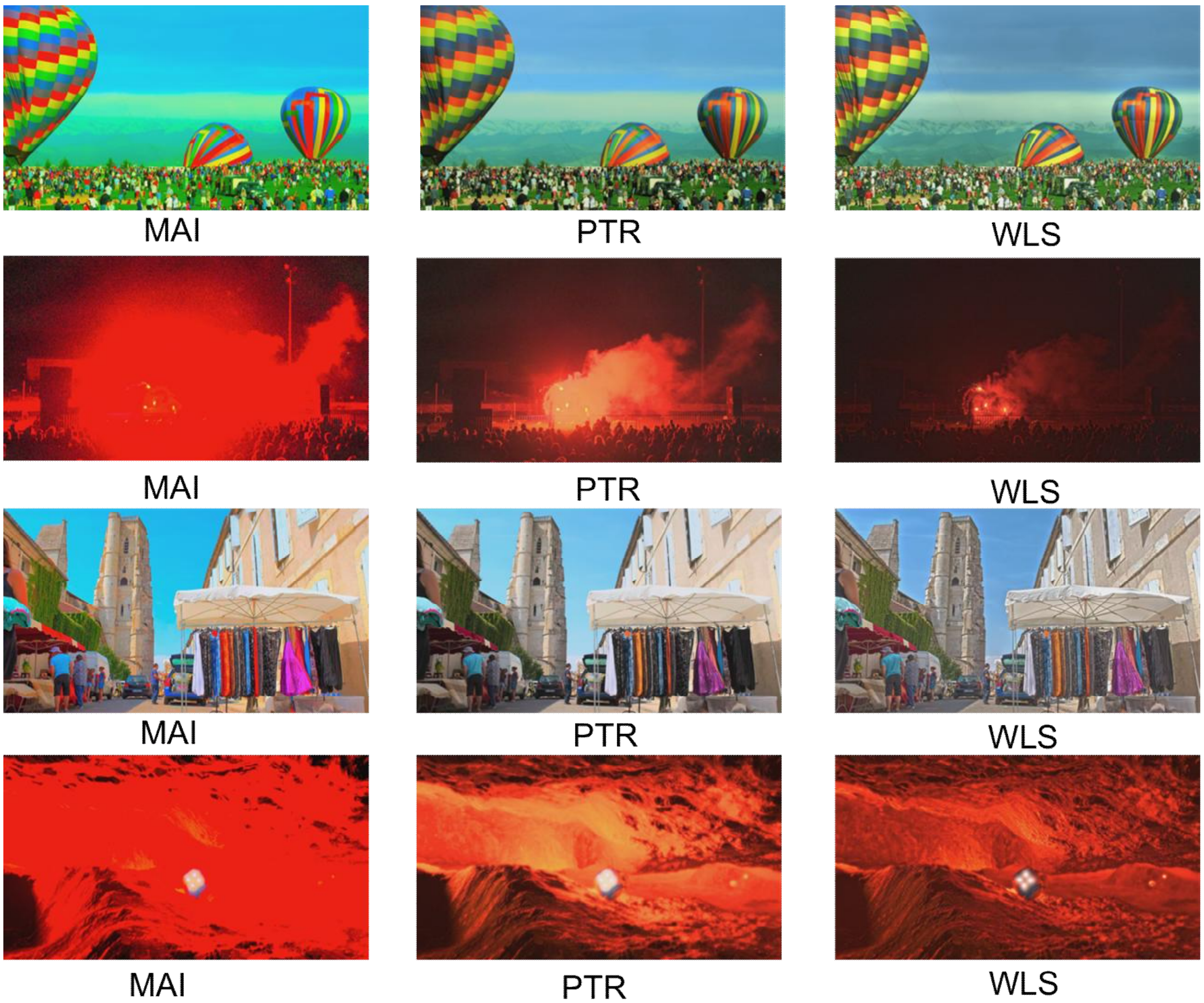
[8] [13] ITU, “Methodology for the subjective assessment of the quality of television pictures BT Series Broadcasting service,” in Recommendation ITU-R BT.500-13, vol. 13, 2012.

[9] cratch Player: http://www.assimilateinc.com/.

8 Patent rights declaration(s).

University of British Columbia (UBC) & TELUS Communications Inc. may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).

9 Annex A: TMOs



10 Annex B: Cropping

Table I: Cropped area per sequences.

(a) (b) (c) (d)

Figure 9. Cropped versions of Market3 (a), FireEater2 (b), BalloonFestival (c) and Tibul2 (d) for subjective test.

Sequence name	**Cropped area**	**QPs**
FireEater2	550 - 1497	21, 25, 29, 33
Tibul2	1 - 948	19, 24, 29, 33
Market3	800 - 1747	29, 33, 37, 41
BalloonFestival	100 - 1047	18, 26, 34, 38