

Effect of Color Space on High Dynamic Range Video Compression Performance

Emin Zerman*, Vedad Hulusic*, Giuseppe Valenzise[‡], Rafał Mantiuk[†] and Frédéric Dufaux[‡]

*LTCI, Télécom ParisTech, Université Paris-Saclay, 46 rue Barrault, 75013 Paris, France

[†]The Computer Laboratory, University of Cambridge, Cambridge, UK

[‡]L2S, UMR 8506, CNRS - CentraleSupélec - Université Paris-Sud, 91192 Gif-sur-Yvette, France

Abstract—High dynamic range (HDR) technology allows for capturing and delivering a greater range of luminance levels compared to traditional video using standard dynamic range (SDR). At the same time, it has brought multiple challenges in content distribution, one of them being video compression. While there has been a significant amount of work conducted on this topic, there are some aspects that could still benefit this area. One such aspect is the choice of color space used for coding. In this paper, we evaluate through a subjective study how the performance of HDR video compression is affected by three color spaces: the commonly used Y’CbCr, and the recently introduced ITP (ICtP) and Ypu’v’. Five video sequences are compressed at four bit rates, selected in a preliminary study, and their quality is assessed using pairwise comparisons. The results of pairwise comparisons are further analyzed and scaled to obtain quality scores. We found no evidence of ITP improving compression performance over Y’CbCr. We also found that Ypu’v’ results in a moderately lower performance for some sequences.

I. INTRODUCTION

The human visual system (HVS) is capable of perceiving a much wider range of colors and luminous intensities present in our environment than the traditional standard dynamic range (SDR) imaging systems can capture and reproduce. High dynamic range (HDR) technology attempts to overcome these limitations by enabling capture, storage, transmission and display of such content, thus allowing more realistic and enhanced user experience [1], [2]. Nevertheless, this comes at the cost of large amount of data, that is difficult to store, transmit and reproduce. Therefore, efficient compression algorithms are being developed trying to balance the required bit rates and perceived quality.

HDR video compression algorithms can be generally classified into two categories as backward-compatible and non-backward-compatible [3]. The former [4], [5], [6] generally use a tone-mapping operator (TMO) in order to generate a base layer stream which can be viewed in SDR displays, and a residual stream which contains additional information for HDR video decoding. The latter [7], [8], [9] can encode videos with high bit-depth quantization and employs state-of-the-art video encoders [10], [11].

For SDR videos, it is common to transform the RGB signal to Y’CbCr color space prior to compression [12], as it is done in state-of-the-art video compression standards [10], [11]. Similarly, in the standardization efforts of MPEG [13], this color space transformation is utilized, while the Y channel is coded with Perceptual Quantization (PQ) [14], [15] instead of

the gamma correction function [16]. In addition to Y’CbCr color space transformation, Lu et al. [17] recently proposed the ITP (ICtP) color space transformation which shows better baseband properties than Y’CbCr for HDR/WCG compression with 10-bit quantization. LogLuv [7] color space transformation is another commonly used transformation among existing HDR video compression algorithms [8], [9]. This color space transformation has been slightly modified in this paper in order to use the same 10-bit encoding scheme and find the effects of it independently of the effects of bit depth. Therefore, we define Ypu’v’ which converts pixel values from RGB to Lu’v’, and encodes L channel with PQ EOTF [15] in order to get Yp, hence takes the name Ypu’v’. In this paper we investigate the effect of these three color spaces on HDR video compression.

To this end, we conduct a psychophysical study to compare video sequences coded at different bit rates with the three aforementioned color spaces. We employ a reduced-design pairwise comparison methodology to get as precise as possible results, comparing also stimuli across different bit rates with the goal of converting the obtained preferences to *quality scores*. The choice of compression levels (bit rates) in this case is crucial, and requires to carefully select test stimuli in such a way to avoid cases where viewers would unanimously prefer one stimulus over the other, or where they would not be able to observe any difference, between pairs of video sequences coded at two consecutive bit rates. Therefore, prior to the main study, we run a preliminary subjective test to select bit rates of stimuli for each content, for a given color space. Specifically, in the pre-study we present stimuli coded at different bit rates, with the goal to select compression levels spaced apart by a *just noticeable difference* (JND), i.e., such that 50% of participants could observe a quality difference between a pair of stimuli.

We analyze the results of the main study by scaling the preference probabilities for each pair of stimuli into global *just objectionable differences* (JOD) scores, as detailed in Section IV-D. One JOD difference between two stimuli corresponds to selecting one video as higher quality than the other in 75% of the trials. We employ the term JOD instead of JND in this case to stress that, in the main experiment, participants were asked to give a *quality* judgment (i.e., select the video which has better overall quality), rather than assess whether a difference between the stimuli exists (as in the pre-study). JOD can then be interpreted similarly to the DMOS concept, and enable to compare different methods using quality-rate curves. We complete the analysis by testing the statistical significance of JOD differences among different color spaces, and find that, overall, there is no substantial gain of ITP over Y’CbCr, while

Ypu'v' has slightly lower performance for some sequences.

II. RELATED WORK

The vast majority of existing video coding schemes entail converting RGB video into an opponent-channel (luminance-chrominance) representation, before carrying out compression, and transforming back to RGB after decoding. Since our chrominance contrast sensitivity exhibits low-pass characteristics, it is customary to subsample chrominance before compression, as a first form of removing redundancies. However, due to the enhanced perception of distortion in HDR content [18], chroma subsampling can produce annoying artifacts, which are mainly due to the correlation between luminance and chrominance channels.

The conventional constant-luminance Y'CbCr representation is commonly used in video compression; however, it has been found that it does not decorrelate optimally the luminance and chrominance information, especially in the HDR scenario. As a result, several alternative color spaces have been proposed in the past few years. The main purpose of these color space transformations is to reduce the correlation between luminance and chrominance. LogLuv [7] is another color space transform that is used as an alternative to Y'CbCr. In [8], Mantiuk et al. use a color space transform similar to LogLuv. First, RGB values are converted to XYZ, and then an 11-bit perceptually uniform luma channel L_p and 8-bit chroma channels u' and v' are found. In [9], Garbas and Thoma use $Lu'v'$ color space. They convert the real-world luminance to 12-bit luma L using a temporally coherent luminance mapping, where they use the weighted prediction feature of H.264/AVC, and 8-bit u' and v' chroma channels. In [19], Mahmalat et al. proposed a luminance independent chroma preprocessing method which uses $Yu'v'$ color space and a chroma transformation and validated their method by a subjective test. In addition to Y'CbCr and LogLuv based ones, Lu et al. [17] proposed another color space transformation called ITP (ICTcp). It features an opponent color model that mimics the human visual system. It is found to be more perceptually uniform than other color spaces.

A comparison between ITP and Y'CbCr color spaces for HDR video compression has been conducted in a recent work of Perrin et al. [20]. The two color spaces have been used in an HDR compression scheme where the RGB colors are converted to the mentioned color spaces. After the statistical tests, it was found that there was no preference for one color space over the other. Moreover, the videos with ITP color space were found to have lower perceived quality than Y'CbCr videos at high bit rates, and vice-versa at low bit rates. Our results confirm these findings, although we add an additional color space, and perform a more detailed statistical analysis of the data, including scaling preferences to quality scores and testing the statistical difference between methods.

III. PRE-STUDY: SELECTION OF TEST STIMULI

The pre-study was designed to find perceptually uniform distances between compressed HDR video sequences, rendered at different levels of compression. These distances are measured in just noticeable difference (JND) units. For each content, four JND steps, starting from the uncompressed sequence, were found. In the pre-study, only the sequences encoded using

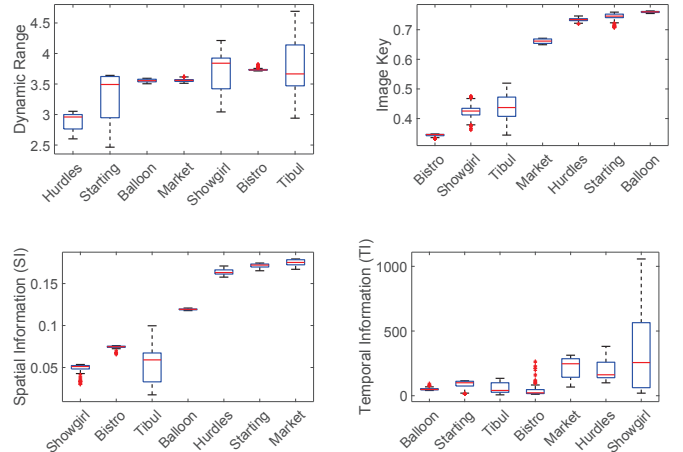


Fig. 1: Image statistics for selected scenes.

Y'CbCr color space were examined, and their corresponding bit rates were used as a reference for compression of the sequences in other two color spaces for the main study.

A. Design

The experiment was conducted in four sessions using two forced-choice evaluation, where the question was: “Can you observe any quality difference between the two displayed videos?”. In the study, the perceptual responses of participants were evaluated in a randomized design. The sequence and the compression rate were the independent variables. The dependent variable was the user preference.



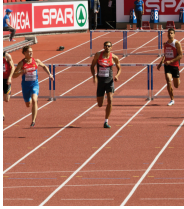



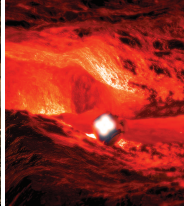
The dataset contained 7 video sequences, with a significant variance in image statistics, as described in section III-B. In each session, for each scene, five to seven sequences with different levels of compression were generated, so that $QP_i = QP_r + 2j$, where $j = \{0.5, 1, 2, 3, 4, 5, 7\}$. Each of these sequences, compressed using QP_i , were compared to the reference sequence with QP_r . In the first session, the uncompressed sequence was the reference and the lowest compression level was selected in the pilot study made with expert viewers. In subsequent sessions, the reference was the previously found sequence with one JND from its reference.

In each trial, two videos of the same content but different compression level were displayed in a side-by-side fashion. Upon the video presentation, the voting sign was displayed allowing the participants to make their choice. They were asked if they can perceive any difference in quality with respect to compression artifact, previously demonstrated during the training session. The voting time was not restricted. Next set of stimuli was presented one second after the user voted.

B. Material

7 HDR video sequences, as shown in Table I, were used in the study. The sequences were selected based on the image statistics and the pilot study, so that the dynamic range (DR), image key (IK), spatial (SI) and temporal (TI) perceptual information measures and image content vary and are evenly distributed across the data set, see Figure 1.

TABLE I: Starting frames, corresponding frame rates and horizontal crop windows (in pixels) of the test sequences used in the pre-study. The Balloon, Market and Tibul sequences were proposed in MPEG by Technicolor and CableLabs [21]; the Bistro and Showgirl sequences are from the Stuttgart HDR Video Database [22]; and Hurdle and Starting sequences are from EBU Zurich Athletics 2014 (<https://tech.ebu.ch/testsequences/zurich>). Showgirl and Tibul scenes were not used in the main study.

Sequence	Balloon	Bistro	Hurdles	Starting	Market	Showgirl	Tibul
							
fps	24	30	50	50	50	30	30
H-crop	921-1872	855-1806	1-952	541-1492	471-1422	429-1380	481-1432

Test sequences were generated using the following chain of operations: First, the RGB HDR frames were encoded using PQ EOTF and then transformed to $Y'CbCr$ color space. After 4:2:0 chroma subsampling, $Y'CbCr$ frames were encoded using HEVC Main-10 profile with HM 16.5 [11]. The encoded bit streams were then decoded and both the color transformation and EOTF encoding have been inverted. The resulting frames were stored in an AVI file as raw video frames. After JND is found for that level, the set of videos for the next level with different QPs were generated, as described in section III-A.

The experiments were conducted in a dark, quiet room, with the luminance of the screen when turned off at 0.03 cd/m^2 . The stimuli were presented on a calibrated HDR SIM2 HDR47ES4MB 47" display with 1920×1080 pixel resolution, peak brightness of 4000 cd/m^2 , used in its native HDR mode. The distance from the screen was fixed to three heights of the display, with the observers' eyes positioned zero degrees horizontally and vertically from the center of the display [23].

C. Participants and Procedure

33 people (M=20 and F=13) with an average age of 33.6, volunteered for the experiment. In each of the four sessions there were 13, 17, 11 and 12 participants respectively, among whom most took part in two nonconsecutive sessions. All of them reported normal or corrected-to-normal visual acuity.

Prior to the experiment, the participants were debriefed about the purpose and the experimental procedure. This was followed by a short training session with 8 sample trials. At this time, two sequences that were not used in the study, rendered at several levels of compression, were utilized and the nature of the artifacts was explained. Following the training, the experiment commenced and no further interaction between the participant and the experimenter occurred until debriefing once all trials were conducted.

D. Results

During the analysis, 37 out of 273 comparisons (per participant and per scene) were removed due to inconsistencies in the responses. Mainly, these were the cases where all the test sequences were found either the same of different as the target stimulus, or when the participant showed inconsistent behavior, e.g., observing a quality difference for a small bit rate

difference but not for larger ones. The JND levels were then obtained by using a logistic fitting and finding thresholds at which 50% of the participants could see the difference between pairs of stimuli.

IV. MAIN-STUDY: COLOR SPACE EFFECT ON COMPRESSION

In the main study the compression performance when coding HDR video sequences using different color spaces was investigated. The bit rates were selected based on the results of the pre-study.

A. Design

For the main experimental task we choose paired comparisons, which provide higher sensitivity, and easier experimental task than direct rating. However, this method may require comparing an excessive number of pairs when a large number of conditions is involved [24], such as in our case. At the same time, comparing stimuli with significantly high perceptual difference leads to obvious and unneeded results. Therefore, an incomplete design, in which only the relevant pairs are compared, is employed.

In the main study HDR video sequences were compared across bit rates for the same color space, and across color spaces using the same bit rate. For the former, only the sequences compressed at the neighboring bit rates were compared, e.g. BR1 vs BR2, or BR3 vs BR4. The uncompressed sequence was compared only with the three videos compressed at the highest bit rates. In each trial the participants had to select the sequence with higher quality, i.e. with lower magnitude and amount of perceivable artifacts.

B. Materials

Due to the high inconsistencies of the pre-study results for Tibul and Showgirl scenes, these were discarded in the main study. The test sequence generation is done similar to the description made in section III-B. RGB videos are either transformed to $Y'CbCr$ after PQ EOTF encoding, to ITP, or to $Ypu'v'$. After 4:2:0 chroma subsampling, converted frames have been encoded using HEVC Main-10 profile with HM 16.5 [11]. The encoded bit streams are then decoded and color transformation and EOTF encoding have been inverted. The

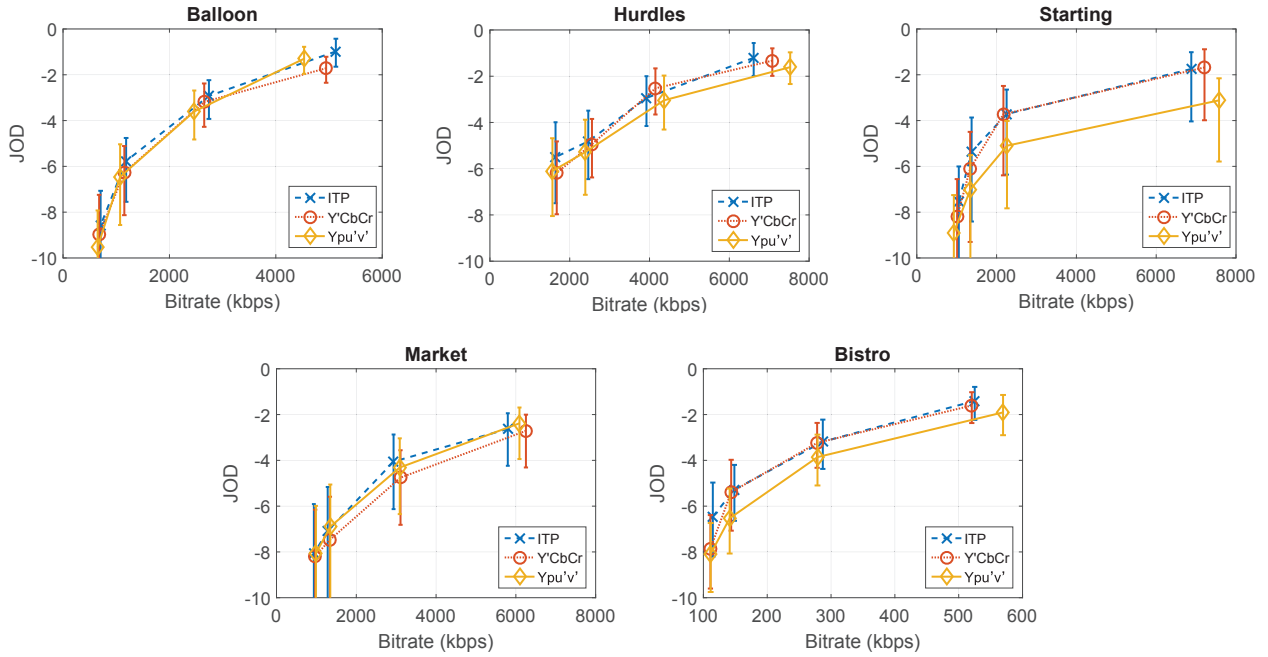


Fig. 2: Image scores obtained by scaling preferences to relative quality distances (in JOD units) for the three tested color spaces.

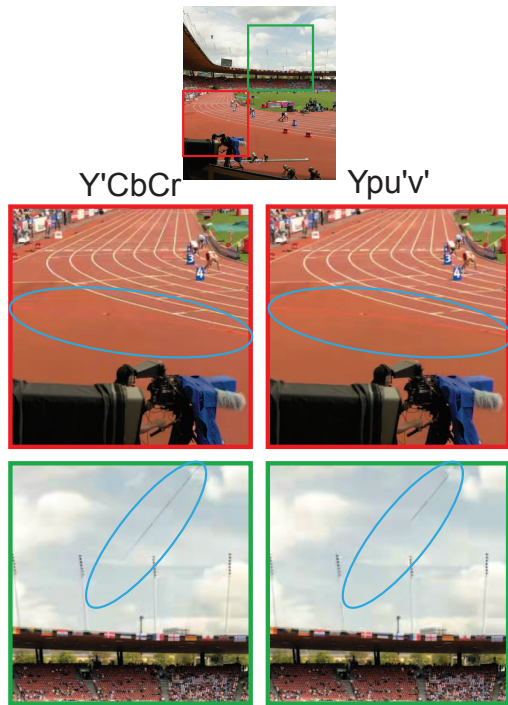


Fig. 3: An example of the difference in compression performance between the $Y'CbCr$ and $Ypu'v'$ color spaces, compressed at the corresponding bit rates (JOD2 level). The compression artifacts are more obvious in the bottom-left (red) corner of the scene when using $Y'CbCr$, while the details in the top (green) part of the scene where more preserved when using this color space.

resulting frames have been stored in an AVI file as raw video frames. As described in previous section, JND levels are found using only $Y'CbCr$ color space transformation. The QP values for ITP and $Ypu'v'$ are found by finding similar bit rates to selected $Y'CbCr$ videos corresponding to different JND levels.

C. Participants and Procedure

18 people (M=14 and F=4) with an average age of 29.44, volunteered for the experiment in the main study. All of them reported normal or corrected-to-normal visual acuity. This time, the participants were asked to select the sequence with the higher quality and thus less compression artifacts, or otherwise make the best guess. 14 participants took part in two sessions, composed of the same pairs but displayed in different order, i.e. A vs B, and B vs A. This way, the total number of user responds per pair was 32. The other aspects of the procedure and the framework were the same as in the pre-study.

D. Results

The pairwise comparison results were scaled using publicly available software¹. The software uses a Bayesian method, which employs a maximum-likelihood-estimator to maximize the probability that the collected data explains JOD-scaled quality scores under the Thurstone Case V assumptions. The optimization procedure finds a quality value for each pair of stimuli that maximizes the likelihood, which is modeled by the binomial distribution. Unlike standard scaling procedures, the Bayesian approach can robustly scale pairs of conditions for which there is unanimous agreement. Such pairs are common

¹*pwcmp* toolbox for scaling pairwise comparison data <https://github.com/mantiuk/pwcmp>

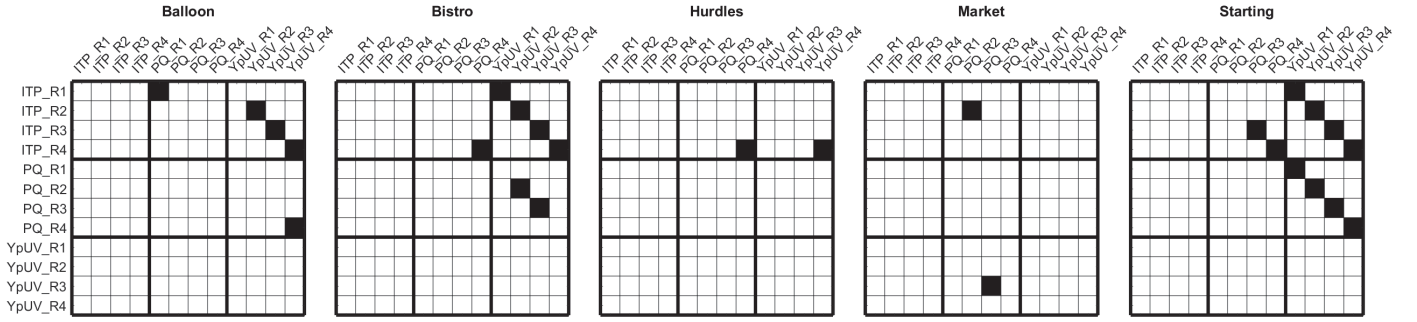


Fig. 4: Difference between test conditions after significance test on JODs. Only the conditions at the same bit rate are reported. Black entries at position (i, j) indicate that stimulus i has been found to be significantly better than stimulus j , at 95% confidence. Similar results are obtained by performing a pairwise binomial test on raw (unscaled) data (not reported due to space limitation).

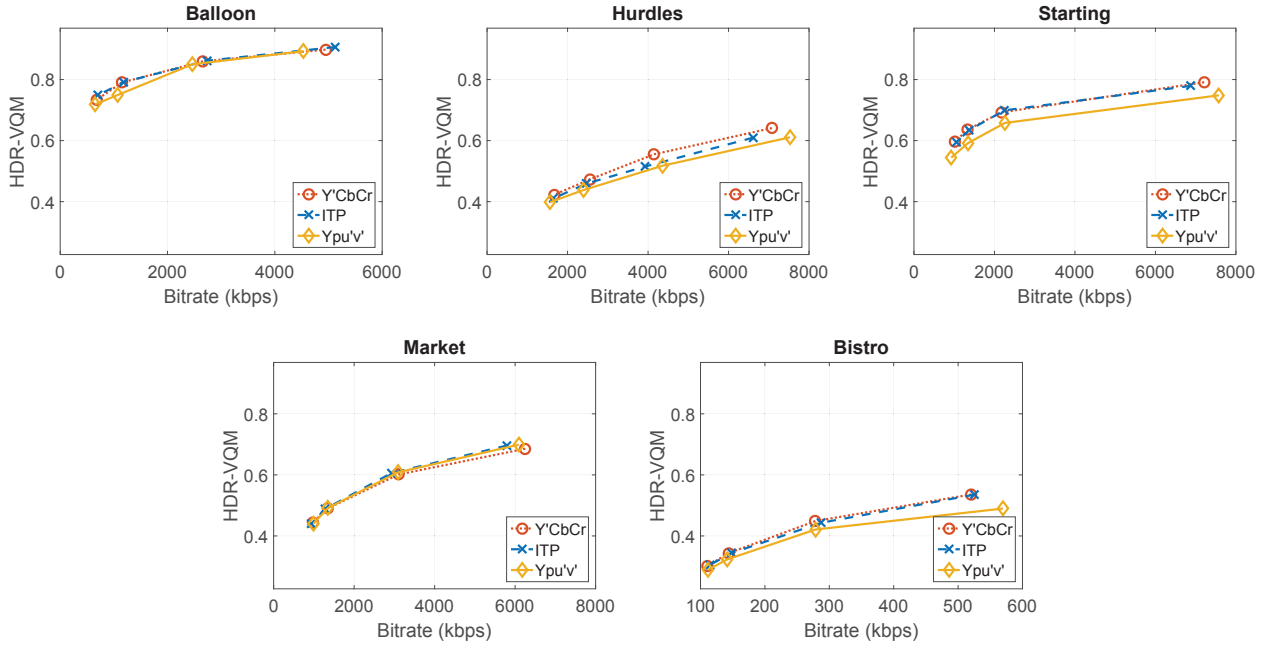


Fig. 5: The results obtained by comparing the all scenes for the three color spaces using the HDR-VQM metric. All scores are normalized, where 1 means perfect quality and lower scores represent a decrease in subjective quality.

when a large number of conditions are compared. It can also scale the result of an incomplete and unbalanced pairwise design, when not all the pairs are compared and some pairs are compared more often than the others. As the pairwise comparisons can provide only relative information about the quality, the JOD values are also relative. To maintain consistency across the video sequences, we always fix the starting point of the JOD scale at 0 for different distortions and thus the quality degradation results in negative JOD values. The confidence intervals were found using bootstrapping. In addition to subjective results, the video quality was predicted using an objective quality metric for HDR video — HDR-VQM [25].

1) *Scaling the pairwise data:* Looking at the scaled data (Figure 2), overall, there is no significant difference between the video compression performance using tested color spaces.

However, there are a few cases where a preference of using one color space over the other is evident, e.g. *Starting* scene at higher bit rates. In this sequence, there were two predominant regions of interest, where the compression artifacts appeared to be the most obvious, see Figure 3. At lower bit rates (JOD1 and JOD2), user ratings were highly dependent on which of these two regions they were focusing while making the comparison. Due to the conflicting appearance of the artifacts, large confidence intervals can be seen on the plots. At higher bit rates (JOD3 and JOD4), the details in the bottom (red) part were corrupted in all the methods, resulting in more uniform responses and smaller confidence intervals. Testing for the evidence of a significant difference between the methods, several cases were found as significantly better than their counterpart, see Figure 4. This test was in accordance with the results from Figure 2, showing Ypu'v' color space mainly

has the worst effect on compression performance, while ITP is not significantly better than Y’CbCr except for a few cases.

2) *Comparison using HDR-VQM*: Comparing the same stimuli using the HDR-VQM objective metric, almost identical results are obtained, see Figure 5. In most of the cases, compression with Ypu’v’ color space results with the lowest quality, except for the Market scene where there is almost no difference in scores. Similar situation is with the Balloon scene, where only at low bit rates, a minor difference between the three methods is found. Notice that we selected the HDR-VQM metric for objective evaluation since this is the only HDR full-reference metric specific to *video*. While it is a color-blind metric, color difference metrics such as ΔE_{2000} have been found to predict poorly visual quality. Nevertheless, we observe that the prediction accuracy of HDR-VQM is not sufficient to predict precisely the quality scores found in the subjective experiment, e.g. *Hurdles* and *Bistro* have significantly lower predicted quality than the actual one.

V. DISCUSSION

The results of our test on the effect of color space on compression performance for HDR video reveal that the influence of color space on coding performance is, in general, little. With the exception of a few specific, content-dependent cases, we did not find evidence of the ITP color space being significantly better than Y’CbCr. Instead, we observed that Ypu’v’ has in general lower performance for coding, although the differences in quality are generally small.

Even in those cases where a difference can be observed, we found that this is strongly content dependent, and is highly influenced by the visual attention patterns of each observer. This produces larger confidence intervals in the estimated quality scores, indicating that the problem of assessing visual quality for small differences in the magnitude of the distortion across stimuli (such as those produced by changing the color space) can be strongly subject-dependent and requires both a careful choice of test material and appropriate analysis tools. We did so in this paper by selecting test stimuli through a preliminary subjective study, aimed at well conditioning the scaling procedure carried out after the main study to find JOD quality scores.

Our results also confirm that HDR video quality metrics can somehow predict the general trend and ranking between stimuli, but are not sufficiently precise to discriminate very tiny perceptual differences and predict absolute quality levels. This motivates further studies in that direction, e.g., to predict not MOS values but JODs.

REFERENCES

- [1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced high dynamic range imaging: theory and practice*. CRC Press, 2011.
- [2] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video. From Acquisition, to Display and Applications*. Academic Press, 2016.
- [3] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers, “Objective and subjective evaluation of high dynamic range video compression,” *Signal Processing: Image Communication*, vol. 47, pp. 426–437, 2016.
- [4] G. Ward and M. Simmons, “Subband encoding of high dynamic range imagery,” in *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*. ACM, 2004, pp. 83–90.
- [5] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, “Backward compatible high dynamic range MPEG video compression,” in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 713–723.
- [6] C. Lee and C.-S. Kim, “Rate-distortion optimized compression of high dynamic range videos,” in *Signal Processing Conference, 2008 16th European*. IEEE, 2008, pp. 1–5.
- [7] G. W. Larson, “Logluv encoding for full-gamut, high-dynamic range images,” *Journal of Graphics Tools*, vol. 3, no. 1, pp. 15–31, 1998.
- [8] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, “Perception-motivated high dynamic range video encoding,” in *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 733–741.
- [9] J.-U. Garbas and H. Thoma, “Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC,” in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 829–832.
- [10] A. M. Tourapis, A. Leontaris, K. Suhring, and G. Sullivan, “H.264/14496-10 AVC reference software manual,” Doc. JVT-AE010, Tech. Rep., 2009.
- [11] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [12] ITU, “ITU-R BT.709-6 Parameter values for the HDTV standards for production and international programme exchange,” International Telecommunication Union, Tech. Rep., 2015.
- [13] A. Luthra, E. Francois, and H. W., “Call for evidence (CfE) for HDR and WCG video coding,” in *ISO/IEC JTC1/SC29/WG11 MPEG2014/N15083*, Feb. 2015.
- [14] S. Miller, M. Nezamabadi, and S. Daly, “Perceptual signal coding for more efficient usage of bit codes,” *SMPTE Motion Imaging Journal*, vol. 122, no. 4, pp. 52–59, 2013.
- [15] SMPTE, “High dynamic range electro-optical transfer function of mastering reference displays,” SMPTE ST 2084, 2014.
- [16] ITU, “ITU-R BT.1886 Reference electro-optical transfer function for flat panel displays used in HDTV studio production,” International Telecommunication Union, Tech. Rep., 2011.
- [17] T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, J. Pytlarz, R. Atkins, J. Fröhlich, and G. Su, “ITP colour space and its compression performance for high dynamic range and wide colour gamut video distribution,” *ZTE Communications*, Feb. 2016.
- [18] T. O. Aydın, R. Mantiuk, and H. Seidel, “Extending quality metrics to full dynamic range images,” in *Proc. of SPIE Electronic Imaging: Human Vision and Electronic Imaging XIII*, San Jose, USA, January 2008, pp. 6806–10.
- [19] S. Mahmalat, N. Stefanoski, D. Luginbühl, T. O. Aydın, and A. Smolic, “Luminance independent chromaticity preprocessing for HDR video coding,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1389–1393.
- [20] A.-F. N. M. Perrin, M. Rerabek, W. Husak, and T. Ebrahimi, “Evaluation of ICtCp color space and an Adaptive Reshaper for HDR and WCG,” *IEEE CONSUMER ELECTRONICS MAGAZINE*, 2016.
- [21] D. Touze and E. Francois, “Description of new version of HDR class A and A’ sequences,” *ISO/IEC JTC1/SC29/WG11 MPEG2014 M*, vol. 35477, 2015.
- [22] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays,” 2014. [Online]. Available: <http://spiedigitallibrary.org>
- [23] ITU, “ITU-R BT.710-4 Subjective assessment methods for image quality in high-definition television,” International Telecommunication Union, Tech. Rep., 1998.
- [24] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, “Comparison of four subjective methods for image quality assessment,” in *Computer Graphics Forum*, vol. 31, no. 8. Wiley Online Library, 2012, pp. 2478–2491.
- [25] M. Narwaria, M. P. Da Silva, and P. Le Callet, “HDR-VQM: An objective quality measure for high dynamic range video,” *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.