AN ADAPTIVE PERCEPTUAL QUANTIZATION METHOD FOR HDR VIDEO CODING

Y. Liu, N. Sidaty, W. Hamidouche, O. Déforges

G. Valenzise[1] and E. Zerman[2]

IETR Lab, CNRS 6164
INSA de Rennes, France

[1]L2S UMR 8506 CNRS, CentraleSupelec, Université Paris-Sud
[2]LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France

ABSTRACT

This paper presents a new adaptive perceptual quantization method for the High Dynamic Range (HDR) content. This method considers the luminance distribution of the HDR image as well as the Minimum Detectable Contrast (MDC) thresholds to preserve the contrast information during quantization. Base on this method, we develop a mapping function for HDR video compression and apply it to a HEVC Main 10 Profile-based video coding chain. Our experiments show that the proposed mapping function can efficiently improve the quality of the reconstructed HDR video in both objective and subjective assessments.

Index Terms— High Dynamic Range (HDR), minimum detectable contrast (MDC), HEVC, HDR video coding

1. INTRODUCTION

Thanks to the development of display technology, the recent commercial televisions are able to display High Dynamic Range (HDR) content [1]. However, HDR imaging often adopts a high bit-depth or floating point representation [2], which requires a large storage capacity and is not compatible with conventional image/video coding systems that process integer digits. Therefore, converting the original HDR format to the integer one is desirable for the application of present coding technologies to the processing of the HDR content.

To this end, Perceptual Transfer Functions (PTFs) were applied to quantize the physical luminance values to integers. Mantiuk et al. proposed a PTF based on the threshold versus intensity (t.v.i) models [3]. It requires a 10 to 11 bits precision to quantize the luminance from $10^{-4}cd/m^2$ to $10^9 cd/m^2$. Later, the peaks of the Contrast Sensitivity Function (CSF) at different luminance adaption levels were used in place of t.v.i [4]. Miller et al. also designed a PTF named Perceptual Quantizer (PQ) [5], which was derived from Barten's CSF [6] and is different from [4]. PQ was reported to use integers with 11-bit depth to quantize the luminance range from $10^{-3}cd/m^2$ to $10^4 cd/m^2$, while 10-bit integers were deemed to be sufficient to represent natural images without visible loss [5]. However, Boitard et al. evaluated several transfer functions and observed that the minimum bit-depth of luminance encoding by PQ is more than 10-bit to avoid perceptual loss [7]. Towards an HDR video transmission solution, ISO/IEC JTC 1 launched a Call for Evidence (CfE) and recommended an HDR coding chain based on HEVC [8] to process 10-bit integers quantized by PQ [9, 10]. The 10-bit depth is a practical choice for both production and distribution of HDR content, due to its compatibility with the existing production and

display equipments [11]. However, the 10-bit PQ encoding is not adequate to completely preserve the perceptual information in a wide luminance range [5, 7]. Besides, current HDR images and videos seldom feature content up to $10^4 cd/m^2$, thus many codewords are actually not used in practice. Therefore, it is necessary to adjust the PQ encoding according to the actual HDR content in order to improve the quality of the quantized HDR content, as well as the efficiency of the HDR video coding. Yu et al. proposed an adaptive PQ for HEVC Main 10 Profile-based video coding and reported a bitrate gain compared to PQ solution [12].

In this paper, we propose a new adaptive perceptual quantization method for HDR content that enables to better employ the available 10 bits used in the standard HDR coding chain. This method is designed to preserve better the contrast information of original HDR content. The so-obtained integer HDR image presents visible enhancement in terms of contrast. The application of the proposed method in the HEVC Main 10 Profile-based coding chain contributes to obvious quality improvements of the decoded HDR videos measured by both objective and subjective evaluations.

The rest of the paper is organized as follows. In Section 2, we present an overview of the HDR video coding chain and the standard PQ PTF as background. The proposed method is described in detail in Section 3. Section 4 presents and analyzes the experimental results. Finally, Section 5 concludes this paper.

2. SYSTEM DESCRIPTION

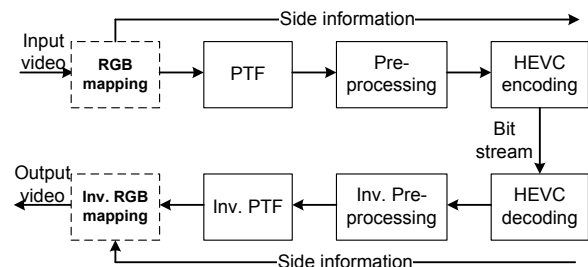### 2.1. Overview of the proposed HDR video coding scheme

Fig. 1. Diagram of the proposed HDR video coding scheme.

The diagram in Fig.1 presents our HDR video coding scheme. The input video refers to HDR videos in RGB float values. The RGB value is firstly mapped into a luminance range. This mapping is based on the proposed adaptive quantization introduced in the next section. A series of side information is also generated in this block. The PQ PTF is applied to the mapped R, G and B channels to get R'G'B' as relative luminance. In the Pre-processing block,

ICIP 2017

the R'G'B' values are converted to YUV color space. The result is quantized to 10-bit integers and downsampled to YUV 420 format. The HEVC encoder encodes the 10-bit YUV video into bit stream. The bit stream feeds the inverse chain to reconstruct the output video with the help of the side information.
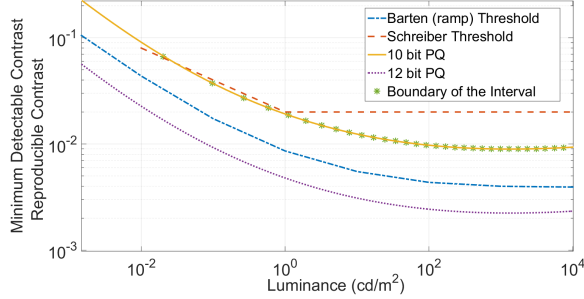
2.2. Perceptual Quantizer

Fig. 2. Reproducible contrast curves of 10-bit and 12-bit PQ.

PQ gives a relationship between the codeword used for integer representation of the signal and the output display luminance [5]:

$$Y_i = f(V_i) = L \left(\frac{max\left[V_i^{\frac{1}{m}} - c_1, 0\right]}{c_2 - c_3 V_i^{\frac{1}{m}}} \right)^{\frac{1}{n}}, \qquad (1)$$

where m, n, c_1, c_2 and c_3 are given coefficients, Y_i is the displayed luminance, L the peak luminance, V_i the relative luminance indicated by the integer codeword i. Using k-bit precision,

$$V_i = i/(2^k - 1), \;\; 0 \leq i \leq 2^k - 1, i \in \mathbb{Z}. \qquad (2)$$

The contrast m_c between two consecutive relative luminance levels is defined as follows [5]

$$m_{c_i} = (Y_{i+1} - Y_i)/(Y_{i+1} + Y_i). \qquad (3)$$

Fig.2 presents the reproducible contrast curves of PQ with 10-bit and 12-bit precisions, as well as the Minimum Detectable Contrast (MDC) thresholds [5]. The quantization artifact is visible when the reproducible contrast is larger than MDC [5]. Compared to the Schreiber MDC threshold [13], the Barten MDC threshold provides a more precise and strict bound [14]. The PQ curves have a similar trend as the Barten threshold, while the 12-bit one is below it. The 10-bit PQ has higher levels of quantization and leads to perceptual contrast loss [7]. In this case, we propose an adaptive HDR quantization method according to the luminance and the MDC thresholds.

3. PROPOSED ADAPTIVE PERCEPTUAL QUANTIZATION

In order to efficiently employ the integer codeword i, we propose a codeword allocation scheme based on the luminance distribution of pixels of HDR images and on the Barten' MDC threshold. The proposed allocation contributes to the contrast information without introducing further distortion compared to PQ. The allocation results are finally used to conduct the quantization of the HDR image.

An initial codeword allocation is firstly calculated in function of the luminance distribution of the pixel. That is, the number of

Fig. 3. Image 'FireEater'. The HDR image in floating point cannot be displayed directly in the conventional monitor, this image is one example of 'FireEater' with a specific exposure.

codewords used to quantize a certain luminance interval is fixed proportionally to the number of pixels having luminance in that range. Specifically, the total luminance range, $0 \sim 10^4 \; cd/m^2$, is segmented into 32 intervals in PQ domain for perceptual uniformity [7]. Each interval has $2^k/32$ relative luminance levels by the k-bit PQ quantization. For example, using 10-bit precision, each interval has $2^{10}/32 = 32$ levels and the jth interval refers to the display luminance range $[Y_{32j}, Y_{32j+31}]$, where $0 \leq j \leq 31$, $j \in \mathbb{Z}$. The boundaries of the intervals are indicated in Fig.2 by star symbols. The luminance distribution is represented by the number of pixels in each interval. Let N denote the number of pixels, Y_l the luminance of the pixel l. The proportion p_j in the jth interval is defined as

$$p_j = \frac{\sum_{l=1}^{N} g_j(Y_l)}{N},$$
$$where \; g_j(Y_l) = \begin{cases} 1, & Y_{32j} \leq Y_l < Y_{32j+31} \\ 0, & otherwise \end{cases} \qquad (4)$$

By definition of p_j, the number of relative luminance levels in the jth interval is reallocated and calculated by

$$N_j' = \lfloor 2^k \cdot p_j + 0.5 \rfloor. \qquad (5)$$

N_j' is proportional to p_j and is treated as the initial allocation result. A large N_j' indicates a small gap between two relative luminance levels in the jth interval. Since the reproducible contrast is related to the step of the displayed luminance between consecutive levels [5, 14], a high p_j results in a low reproducible contrast value and vice versa. However, considering p_j alone brings visible artifacts in contrast.
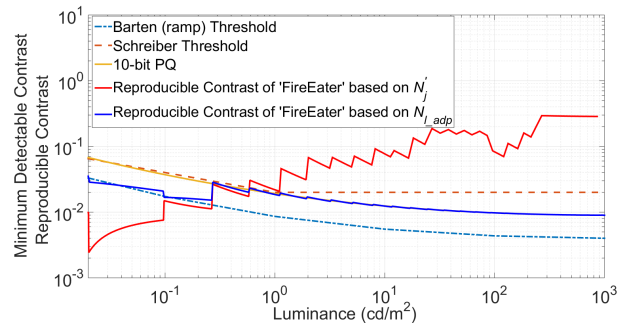

Fig. 4. Reproducible contrast of 'FireEater' with 10-bit precision.

Fig.3 shows one example of the HDR image 'FireEater' while Fig.4 presents the reproducible contrast of 'FireEater' based on N_j' only. Compared to 10-bit PQ, the reproducible contrast by the use of N_j' decreases in the low luminance region while rises and exceeds

the MDC thresholds in high luminance. Although a greater precision in contrast is achieved in shadow, the reproducible contrast is below the Barten MDC and does not contribute to a better viewing experience [5], indicating that this approach alone is not effective in optimizing the use of codewords. In order to use the relative luminance levels more efficiently, the MDC is taken into account.

The MDC of Barten threshold is calculated by the Barten CSF at a specific luminance [5, 6]. In the jth interval $[Y_{32j}, Y_{32j+31}]$, we use equation (3) to compute the higher luminance level $Y_{i+1,j}$ from the lower one $Y_{i,j}$, until $Y_{i+1,j}$ approaching Y_{32j+31}:

$$Y_{i+1,j} = Y_{i,j} \cdot \frac{1 + MDC_{i,j}}{1 - MDC_{i,j}}, \; Y_{1,j} = Y_{32j}, i \geq 0, i \in \mathbb{Z} \quad (6)$$

where $MDC_{i,j}$ is the MDC of $Y_{i,j}$. Let N_{j_MDC} denote the number of levels determined by Barten CSF in the jth interval.

$$N_{j\_MDC} = i, \; when \; Y_{i,j} \leq Y_{32j+31} < Y_{i+1,j} \; . \quad (7)$$

$Y_{i,j}$ is no more than Y_{32j+31}. Then, the adaptive number of the relative luminance levels in jth interval is determined by

$$N_{j\_adp} = \begin{cases} 2^k/32, & N'_j \leq 2^k/32 \\ N'_j, & 2^k/32 < N'_j \leq N_{j\_MDC} \\ N_{j\_MDC}, & N'_j > N_{j\_MDC} \end{cases} \quad (8)$$

The luminance in the interval $[Y_{32j}, Y_{32j+31}]$ can then be uniformly quantized into N_{j_adp} relative luminance levels. Fig.4 also presents the reproducible contrast derived from N_{j_adp} for 'FireEater'. The contrast information is reserved more than 10-bit PQ in the low luminance region, while close to PQ in high luminance.

For video coding, we applied N_{j_adp} in the RGB mapping function in Fig.1. The total luminance range is still separated into 32 intervals, while the minimum Y_{j_adp} and maximum Y_{j+1_adp} luminance in jth interval is determined by

$$Y_{j\_adp} = f\left(\frac{\sum_{r=1}^{j-1} N_{r\_adp}}{2^k - 1}\right), 0 \leq j \leq 31, \; j \in \mathbb{Z} \quad (9)$$

$f(\cdot)$ refers to Eq.(1). The mapped RGB value is calculated by

$$val' = (Y_{j+1\_adp} - Y_{j\_adp})\frac{val - Y_{32j}}{Y_{32j+31} - Y_{32j}} + Y_{j\_adp} \quad (10)$$

where val is the original RGB value. N_{j_adp} of each interval is sent as the side information to reconstruct the HDR video at user side. The bitrate cost of the side information is $3.88 \sim 7.75$ kbps, which is relative to the frame rate in our implementation. Fig.5 presents the quantized YUV 'FireEater' images with 10-bit precision. The image quantized with mapping shows more details in shadow than the one without it.

4. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed adaptive quantization method for HDR video coding, we implemented the mapping function in the HDR video coding chain [9]. The HEVC is performed in the HM 16.2 model [15] with the main 10-profile random access configuration [9]. Six HDR videos (size:1920×1080), 'Balloon', 'Market3', 'Hurdles', 'Starting', 'FireEater2' and 'Tibul2' are tested. Each video is processed by the coding chain in Fig.1. The quality of the reconstructed HDR video is evaluated through both objective metrics and subjective experiments. In the objective test,
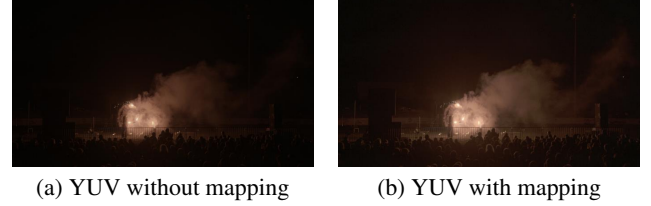

(a) YUV without mapping   (b) YUV with mapping

Fig. 5. Quantized HDR images with 10-bit precision.

the reconstructed video is measured by HDR VQM [16], which has been demonstrated to be reliable to assess the perceived HDR content quality [17], as well as tPSNR and PSNR_DE [9], which are used in [12] for comparison. In the subjective test, the reconstructed videos from the coding chains with conventional PQ and with the proposed adaptive mapping function are displayed in the HDR display for paired comparison. The bitrates of encoding test videos are in Table 1. The cost of the side information has been considered in bitrates of the mapping solution.

4.1. Objective evaluation results

The comparison on the HDR VQM score is shown in Fig.6 where a low score indicates a quality close to the original video. The results of the mapping function are better than those without mapping for 'Balloon', 'Hurdles', 'Starting', 'FireEater2' and 'Tibul2', while close for 'Market3'. The advantage of the mapping function is greater at high bitrate than at low bitrate evaluated by HDR VQM.

Yu et al. proposed an adaptive PQ for HDR video coding and tested it with four videos to show BD-rate gains to PQ in tPSNR and PSNR_DE. We also calculated BD-rate gains of the mapping function in Table 2. A negative value indicates a BD-rate gain measured by a quality metric. The proposed mapping function, which modifies quantized levels to each luminance interval, has better average results than [12], which adapts the quantization based on the maximum and minimum luminance of the HDR content. Particularly for 'FireEater', which has many pixels with low luminance, the proposed mapping function preserves better details in dark content and contributes much BD-rate gain.

4.2. Subjective evaluation results

The subjective test has been conducted in a dark, quiet room. The videos were presented on a calibrated HDR SIM2 display with peak brightness of $4000 \; cd/m^2$ [18]. The distance from the screen is fixed to three heights of the display, with observers' eyes positioned zero degrees horizontally and vertically from the center of the display.

Four videos of 8 sec duration each: 'Balloon', 'Market3', 'Hurdles' and 'Starting', all encoded at four bitrates shown in Table 1, were evaluated. 'FireEater' and 'Tibul2' were used for training observers to vote. In one specific Rate, the reconstructed HDR videos processed by PQ only and by the proposed mapping function were displayed side by side twice, having each video on both sides (A vs B, and B vs A). In each trial the observer was asked to select the one with higher quality. If the observer made two opposite choices, we treat the two videos as a tie judged by this observer. Otherwise, the chosen video is considered to have a higher perceivable quality.

16 observers, aging from 23 to 39, took part in the test. Their voting results are converted to subjective quality scores by the use of Paired Evaluation via Analysis of Reliability (PEAR) method [19].

Video	Rate1 (kbps)			Rate2 (kbps)			Rate3 (kbps)			Rate4 (kbps)		
	PQ	Proposed	ΔR %	PQ	Proposed	ΔR %	PQ	Proposed	ΔR %	PQ	Proposed	ΔR %
Balloon	6639.6	6623.7	+0.239	3763.4	3734.3	+0.773	2154.5	2144.1	+0.483	1275.4	1267.0	+0.659
Market3	7914.5	7893.9	+0.260	4219.6	4209.9	+0.230	2313.0	2301.4	+0.514	1248.1	1241.9	+0.497
Starting	3213.2	3195.8	+0.542	1730.0	1725.0	+0.289	1028.5	1020.1	+0.817	618.5	607.0	+1.859
Hurdles	6311.8	6271.2	+0.643	3312.5	3283.5	+0.875	1818.9	1773.0	+2.524	1052.2	1048.6	+0.342
FireEater2	1912.3	1904.8	+0.392	1257.2	1256.0	+0.095	808.6	801.5	+0.878	519.8	517.0	+0.539
Tibul2	6090.6	6082.0	+0.141	2499.8	2480.6	+0.768	970.1	969.0	+0.113	402.4	401.7	+0.174

Table 1. Bitrates of test videos used in the evaluation. $\Delta R = (Rate(PQ) - Rate(Proposed))/Rate(PQ)$.
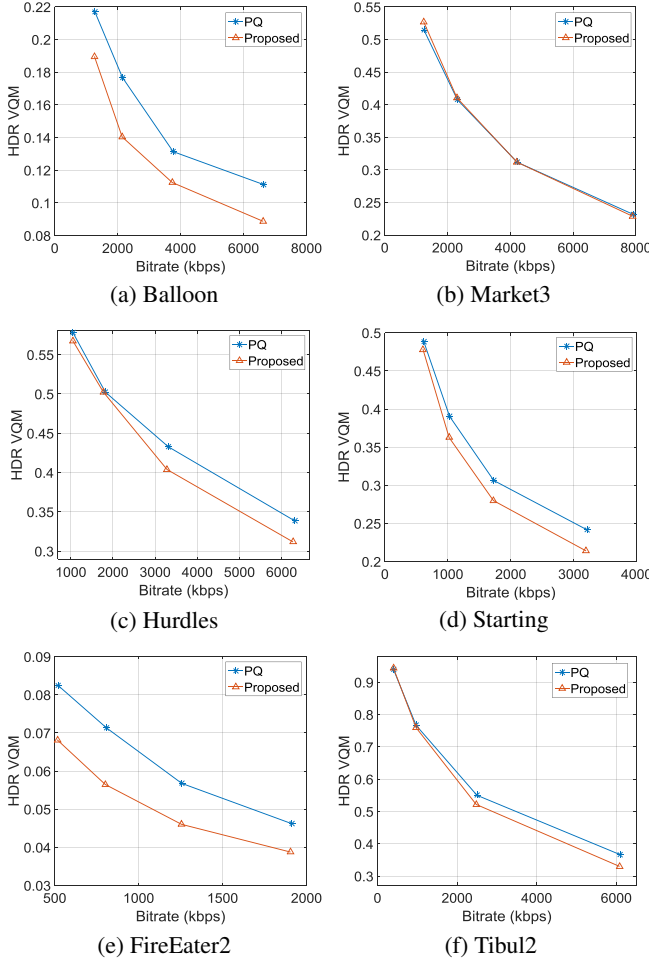

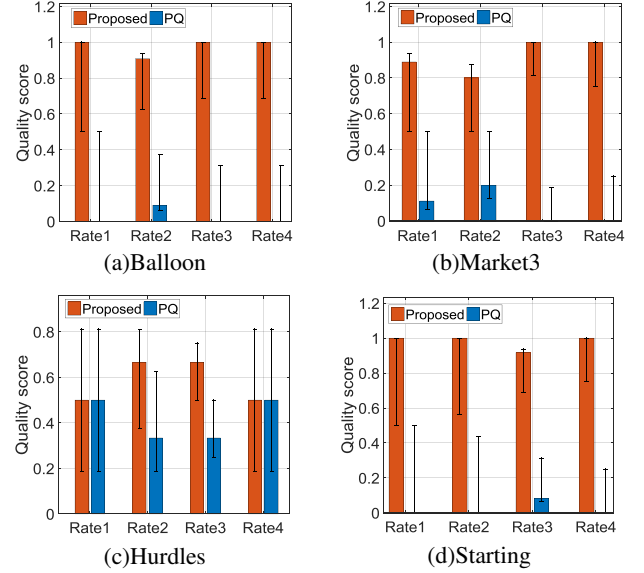**Fig. 6**. Comparison of PQ and the proposed mapping function on the HDR VQM score.

Fig. 7. Comparison of PQ and the proposed mapping function on the subjective score.

This method provides relative quality between stimuli with associated confidence intervals. The comparison on the subjective score is shown in Fig.7 where the score is from 0 to 1. For 'Balloon', 'Market3' and 'Starting', the video coding chain with the mapping function can reconstruct HDR videos with higher perceivable qualities than those without mapping. Particularly at Rate 3 and 4, the confidence intervals of PQ and the proposed have no overlapping and are quite far. It indicates a significant quality improvement contributed by mapping. For 'Hurdles', the quality scores are the same for Rate 1 and 4, while the proposed mapping scheme has better scores for Rate 2 and 3. In summary, the mapping function is beneficial to the HDR video coding for a better perceivable quality.

5. CONCLUSION

We present a new adaptive perceptual quantization method for HDR content to efficiently employ the codewords to represent the luminance. With this method, the quantized HDR image can present more contrast details. We also apply it to HDR video coding and introduce a new mapping function. Experimental results show that the HDR video coding with the proposed mapping function can achieve reconstructed videos with better qualities than those of the video coding with PQ only in both objective and subjective evaluations.

video	tPSNR XYZ (in %)		PSNR_DE (in %)	
	[12]	proposed	[12]	proposed
Market3	-3.37	-2.1	-7.44	-10.5
Tibul2	-3.17	0.9	-5.04	-2.6
Balloon	-1.60	-2.1	-3.43	-10.5
FireEater2	-5.88	-15.2	-6.12	-21.5
Average	-3.51	-4.63	-5.51	-11.28

Table 2. Comparison of [12] and proposed mapping function on the BD-rate gain

6. REFERENCES

[1] Consumer Technology Association, "Next generation of innovation debuts as CES 2016 kicks off," in *CES Press Release*. Consumer Electronics Show, 2016.

[2] E. Reinhard, G. Ward, S. N. Pattanaik, P. E. Debevec, and W. Heidrich, *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting (2. ed.).*, Academic Press, 2010.

[3] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H. Seidel, "Perception-motivated high dynamic range video encoding," in *ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, SIGGRAPH '04, pp. 733–741, ACM.

[4] R. Mantiuk, K. Myszkowski, and H. Seidel, "Lossy compression of high dynamic range images and video," in *SPIE 6057, Human Vision and Electronic Imaging XI*. SPIE, 2006, pp. 60570V–60570V–10.

[5] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," in *The 2012 Annual Technical Conference Exhibition*, Oct. 2012, pp. 1–9.

[6] P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," in *Proc. SPIE*, 2003, vol. 5294, pp. 231–238.

[7] R. Boitard, R. Mantiuk, and T. Pouli, "Evaluation of color encodings for high dynamic range pixels," in *Proc. SPIE*, 2015, vol. 9394, pp. 93941K–93941K–9.

[8] G. J. Sullivan, J. R. 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, Dec. 2012.

[9] A. Luthra, E. François, and W. Husak, "Call for Evidence (CfE) for HDR and WCG Video Coding," WG11 N15083, International Organization for Standardization, Feb. 2015.

[10] E. François, C. Fogg, Y. He, X. Li, A. Luthra, and A. Segall, "High dynamic range and wide color gamut video coding in HEVC: Status and potential future enhancements," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 63–75, Jan. 2016.

[11] SMPTE Technical Commitee 10E SG, "Study Group Repout High-Dynamic-Range (HDR) Imaging Ecosystem," Tech. Rep., SMPTE, Sep. 2015.

[12] S. Yu, C. Jung, and P. Ke, "Adaptive PQ: Adaptive perceptual quantizer for hevc main 10 profile-based hdr video coding," in *2016 Visual Communications and Image Processing (VCIP)*, Nov. 2016, pp. 1–4.

[13] William F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer, 1993.

[14] Radiocommunication Sector of ITU, "The present state of ultra-high definition television," Tech. Rep., International Telecommunication Union, Jul. 2015.

[15] B. Bross, W. J. Han, J. R. Ohm, G. J. Sullivan, Y. K. Wang, and T. Wiegand, "HM, the reference software for HEVC," https://hevc.hhi.fraunhofer.de/.

[16] M. Narwaria, M. Perreira, 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.

[17] P. Hanhart, M. V. Bernardo, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP Journal on Image and Video Processing*, p. 39, 2015.

[18] SIM2 Corporation, "SIM2 HDR47 Monitors," http://hdr.sim2.it/productslist, 2016.

[19] J. S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, Oct. 2011.