



HAL
open science

An Adaptive Quantizer for High Dynamic Range Content: Application to Video Coding

Yi Liu, Naty Sidaty, Wassim Hamidouche, Olivier Deforges, Giuseppe
Valenzise, Emin Zerman

► **To cite this version:**

Yi Liu, Naty Sidaty, Wassim Hamidouche, Olivier Deforges, Giuseppe Valenzise, et al.. An Adaptive Quantizer for High Dynamic Range Content: Application to Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29 (2), pp.531-545. 10.1109/TCSVT.2017.2786746 . hal-01704278

HAL Id: hal-01704278

<https://hal.science/hal-01704278v1>

Submitted on 5 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Adaptive Quantizer for High Dynamic Range Content: Application to Video Coding

Yi Liu, Naty Sidaty, Wassim Hamidouche, Olivier Déforges, Giuseppe Valenzise and Emin Zerman

Abstract—In this paper, we propose an adaptive perceptual quantization method to convert the representation of High Dynamic Range (HDR) content from the floating point data type to integer, which is compatible with the current image/video coding and display systems. The proposed method considers the luminance distribution of the HDR content, as well as the detectable contrast threshold of the Human Visual System (HVS), in order to preserve more contrast information than the Perceptual Quantizer (PQ) in integer representation. Aiming to demonstrate the effectiveness of this quantizer for HDR video compression, we implemented it in a mapping function on the top of the HDR video coding system based on High Efficiency Video Coding (HEVC) standard. Moreover, a comparison function is also introduced to decrease the additional bit-rate of side information, generated by the mapping function. Objective quality measurements and subjective tests have been conducted in order to evaluate the quality of the reconstructed HDR videos. Subjective test results have shown that the proposed method can improve, in a significant manner, the perceived quality of some reconstructed HDR videos. In the objective assessment, the proposed method achieves improvements over PQ in term of the average bit-rate gain for metrics used in the measurement.

Index Terms—High Dynamic Range (HDR), perceptual quantization, HDR video coding, HEVC, objective quality metric, subjective test.

I. INTRODUCTION

AIMING at providing a realistic visual experience, recent displays can reproduce a higher contrast and luminance range, providing a more realistic and immersive visual experience to human observers [1], [2]. In order to support these enhanced capabilities, High Dynamic Range (HDR) image/video technology provides tools for capturing, representing and storing the physical luminance of a scene. Differently from Low Dynamic Range (LDR) image/video represented by 8-bit integer pixels, HDR imaging requires higher bit-depth and floating point values to record the full luminance range perceivable by the human eye [3]. This requirement brings challenges to both storage and distribution of the HDR content. In fact, conventional coding systems are designed for LDR content with limited bit-depth, thus they are not able to process HDR content directly. Therefore, it is desirable to convert an HDR signal from floating point representation to low bit-depth integers that are compatible with the present compression, transmission and display systems.

Yi Liu, Naty Sidaty, Wassim Hamidouche and Olivier Déforges are with the Institute of Electronics and Telecommunications of Rennes (IETR), UMR CNRS 6164, National Institute for applied Sciences, Rennes, France.

Giuseppe Valenzise is with L2S UMR 8506 CNRS, CentraleSupélec, Université Paris-Sud, France.

Emin Zerman is with Télécom Paristech, Université Paris Saclay, France.

In this respect, perceptual transfer functions are used to quantize the original HDR values to integers at a given bit-depth. Mantiuk *et al.* proposed a Perceptual Transfer Function (PTF) for HDR encoding based on the threshold versus intensity (t.v.i) [4]. This PTF requires between 10 and 11 bits to represent the luminance range from 10^{-4}cd/m^2 to 10^8cd/m^2 . In a later work, the same authors improved the quantization error by replacing the t.v.i with the Contrast Sensitivity Function (CSF) of the HDR-Visual Difference Predictor (HDR-VDP) [5]. Miller *et al.* considered the peaks of CSF derived from Barten's model [6] and proposed an Electro-Optic Transfer Function (EOTF), the Perceptual Quantizer (PQ), which was designed to quantize the luminance with peak value up to 10^4cd/m^2 [7]. Besides, Touzé *et al.* defined a transfer function combining Gamma correction and S-log correction to finely quantize dark pixels while avoiding the harsh high light saturation [8]. Boitard *et al.* evaluated these transfer functions with respect to the perceptually uniform of integer-value representation of HDR pixels [9]. The perceptual uniformity property refers to the fact that the difference between two given intensity values results in a perceived difference that has approximately the same perceived magnitude independently from the absolute intensity values of the pixels [10]. They concluded that PQ offered the best uniformity. Meanwhile, more than 10 bits are required by PQ to avoid perceptual loss in their study with the luminance level up to 150cd/m^2 [9].

In 2014, PQ was standardized as SMPTE ST-2084 [11]. With the goal to support end-to-end delivery of HDR video using the most recent video coding standard, the Joint Technical Committee ISO/IEC JTC 1 launched a Call for Evidence (CfE) targeting possible extensions of the High Efficiency Video Coding (HEVC) [12] for efficient coding of HDR and Wide Color Gamut (WCG) video contents [13]. In this CfE, an HDR video coding chain was recommended as Anchor to use HEVC standard to encode videos with 10-bit integers quantized by PQ [13], [14]. The implementation with 10-bit depth is technically compatible with the existing production and display equipments [15]. However, 10-bit PQ encoding may introduce perceptual loss in luminance [7], [9]. Besides, the peak luminance of the current HDR image/video seldom reaches 10^4cd/m^2 , which is the maximum luminance supported by PQ [7], as a result, many codewords are not used in the integer-representation of the HDR content in practice, leading to a waste of resources. Therefore, it is desirable to employ the codewords efficiently based on the actual HDR content in order to preserve better perceptual information in the quantized HDR image/video and increase visual quality.

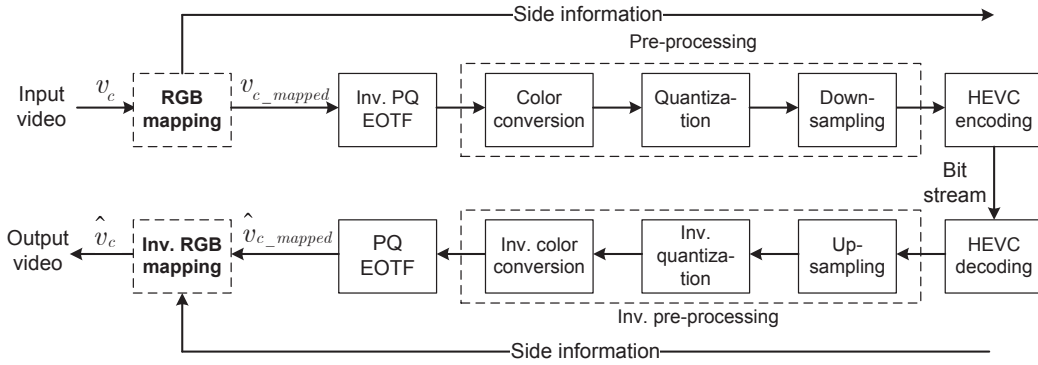


Fig. 1. Diagram of the recommended HDR video coding system with proposed modifications.

There have been several contributions to redistribute codewords for the integer-representation of HDR content. An adaptive reshaping was implemented in the Exploratory Test Model (ETM) for HDR extension of HEVC [16], [17]. In the luma resaper, it divides the range $0-10^4 \text{cd/m}^2$ into a number of pieces of luminance segments and redistributes codewords based on predefined criteria [17]. J. Froehlich *et al.* exploited the local noise and texture of the image. They proposed a Content Aware Quantization (CAQ) method to determine the bit-depth for the quantization of the HDR data [18]. Later, some encoder optimization processes, e.g., perceptual luma quantization and chroma QP offset, are introduced to exploit HDR data characteristics in the video encoder [19]. In [20], an adjustment factor *ratio* was introduced to form an adaptive PQ. This *ratio* is determined by the maximum and minimum luminance of the HDR content. Based on this ratio, the adaptive PQ was reported to show a bit-rate gain compared to the PQ solution in the Cfe.

In this paper, we propose a perceptual quantization method for HDR content and apply it to HDR video coding based on the HEVC standard. Different from the resaper scheme, which halves the codewords for less important luminance intervals and allocates codewords for important intervals by a mathematical power function [17], the proposed method adaptively regulates codewords in luminance intervals based on the minimum detectable contrast in order to avoid contouring artefacts [7], [21] and imperceivable change in contrast [7]. Specifically, this method is conducted in two main steps. The first step consists in the quantization of the HDR content according to the probability distribution of pixels. In this step, a higher number of codewords is allocated to the luminance intervals having a higher density of pixels. It aims to preserve the contrast information, i.e., content details of the original HDR image/frame. However, the over allocation of codewords in one luminance interval leads to a waste of code resource for other intervals [7]. In order to employ the code resource efficiently, the minimum detectable contrast derived from the Barten's model is used in the second step to correct the initial allocation issued from the first step. After that, the luminance value in each interval is quantized uniformly based on the corrected number of codewords. Compared to PQ, the proposed method is able to employ the codewords more efficiently to preserve the contrast information of the HDR

content. Subsequently, we propose a mapping function to apply the quantization method within the HDR video coding. We also present a comparison function to reduce the additional bit-rate cost of the proposed mapping function, and avoid the jitter in luminance of the quantized HDR video. We evaluated the effectiveness of the proposed method on HDR video coding both through subjective tests and using objective HDR quality metrics. The experimental results show that the proposed method improves the quality of the reconstructed HDR videos.

The rest of this paper is organized as follows. Section II describes the HDR video coding system with the proposed modifications, as well as the PQ EOTF solution as background. Section III introduces the proposed adaptive perceptual HDR quantization method. Further details of the application of the proposed method on HDR video coding are provided in Section IV. Section V presents and analyzes the experimental results. Finally, Section VI concludes this paper.

II. DESCRIPTION OF THE HDR VIDEO CODING SYSTEM

The proposed HDR video coding system is based on the coding chain introduced in the Cfe [13]. We add a mapping function at the beginning of the chain in order to adjust the range of the HDR content values. A parameter data stream is generated (side information) in this function. Thanks to the optimized implementation presented in Sections III and IV, the additional bit-rate cost of these parameters is considerably reduced.

A. Overview of the Proposed System

Fig. 1 presents the diagram of the proposed HDR video coding system. The input is the HDR video in RGB float values representing the absolute luminance values in cd/m^2 . The RGB values, v_c , are first mapped to different ranges. This mapping function is based on the proposed perceptual quantization method, which is introduced in Section IV. A set of parameters is generated in this step and sent as side information to the decoder. Next, the mapped RGB values, v_{c_mapped} , are converted to PQ domain, R'G'B', by the inverse transform of PQ EOTF. After that, the R'G'B' values enter the Pre-processing block to convert R'G'B' into Y'CbCr color space, and then the resulting float values are quantized to

10-bit integers and finally down-sampled into Y'CbCr 4:2:0 format. The Y'CbCr 4:2:0 10-bit video is encoded by HEVC encoder. The encoded bitstream feeds the inverse coding chain to reconstruct the mapped RGB values, \tilde{v}_{c_mapped} , and RGB values, \tilde{v}_c , with the help of the side information.

B. Perceptual Quantizer

Miller *et al.* introduced the Perceptual Quantizer (PQ) model to describe the relationship between codewords for integer-representation of the HDR signal and the output display luminance [7]:

$$Y(i) = f[V(i)] = L \left(\frac{\max[V(i)^{\frac{1}{m}} - c_1, 0]}{c_2 - c_3 V(i)^{\frac{1}{m}}} \right)^{\frac{1}{n}}, \quad (1)$$

where m , n , c_1 , c_2 and c_3 are given coefficients, Y the display luminance, L the peak luminance, V the luma, indicated by the integer codeword i . Using k -bit codewords, V is expressed by Equation (2).

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

Where V is the normalization of codeword i and $V \in [0, 1]$. When the codeword varies from i to $(i+1)$, the minimum reproducible contrast is defined in Equation (3) [22].

$$m_{c_i} = \frac{\Delta Y}{Y_{ave}} = 2 \frac{Y(i+1) - Y(i)}{Y(i+1) + Y(i)}. \quad (3)$$

Fig. 2 shows the reproducible contrast curves of PQ with codewords with 10-bit, 11-bit and 12-bit depth, as well as the Minimum Detectable Contrast (MDC) thresholds [7]. The MDC refers to the ability to detect discontinuities in tone reproduction of the luminance [21]. The contouring artefact is visible when the reproducible contrast is larger than MDC [7]. Compared to the Schreiber MDC threshold [23], the Barten MDC threshold provides a lower and more strict bound [6], [21]. The curves of PQ have similar shapes with the Barten threshold. Meanwhile, the 12-bit one has the highest precision and is below the Barten threshold. The 11-bit PQ curve is close to the Barten threshold, while the 10-bit one has highest level of quantization that results in perceptual contrast loss [9]. Notice that the recommended PQ model was proposed for a peak luminance of 10^4 cd/m^2 [7], which is often higher than the peak luminance of the actual HDR content. In this work, we propose a perceptual, content-dependent HDR quantization method that makes a better use of the actual dynamic range of the video.

III. PROPOSED PERCEPTUAL QUANTIZATION METHOD

The proposed quantization method enables to efficiently employ the integer codeword i considering both the luminance distribution of pixels in the HDR content and Barten's MDC threshold. Essentially, it is a reallocation procedure of codewords and is carried out in two steps. The first step, named initial codeword allocation, consists in the allocation of codewords based on the luminance distribution. For an HDR image or frame of the video, the majority of pixels often concentrate into several luminance intervals. For example, in

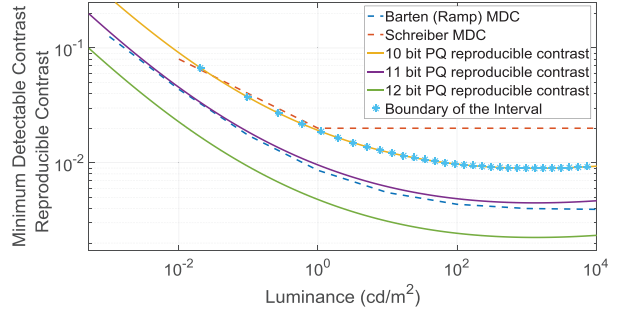


Fig. 2. Minimum detectable contrast thresholds and reproducible contrast curves of 10-bit, 11-bit and 12-bit PQ.

an HDR content describing a bright scene, most pixels have high luminance. The first step allocates the majority of codewords into high luminance intervals to preserve the contrast information in the bright scene, and vice versa. However, increasing the number of codewords allocated to a luminance interval does not necessarily lead to higher visible quality. In fact, if the reproducible contrast is below the Barten's MDC threshold, any change in contrast is imperceivable [7]. Therefore, in the second step of the proposed method, referred as the corrected codeword allocation, the minimum detectable contrast is used to adjust the number of codewords in each luminance interval to avoid the waste of code resource. The details of the procedure are explained in following sections.

A. Initial codeword allocation

In the initial allocation step, the number of codewords used to quantize a certain luminance range is proportional to the number of pixels having the luminance within that range. Specifically, we firstly segment the total luminance range, $0 - 10^4 \text{ cd/m}^2$, into 2^γ intervals, where γ is fixed in this paper to 5 for 32 intervals. The number of intervals is variable and can be considered as an input parameter. However, each interval produces one parameter to transmit to the decoder. Therefore, the number 32 is selected as a trade-off between quality improvement and side information (bit-rate) overhead. This segmentation was processed in quantized PQ data to keep the perceptual uniformity [9]. In one interval, there are $2^k/2^\gamma = 2^{k-\gamma}$ luminance levels, with $0 \leq \gamma \leq k$. For example, each interval has $2^{10-5} = 32$ levels for the 10-bit depth quantization, when $\gamma = 5$. Each level can be denoted by a codeword i , $0 \leq i \leq 1023$. We use Equation (1) to calculate the luminance values of bounds for each interval. Denote the minimum luminance value of the j th interval by $Y_{min}(j)$, the maximum value by $Y_{max}(j)$, the bounds of the j th interval, $[Y_{min}(j), Y_{max}(j)]$, are expressed by Equation (4):

$$\begin{aligned} Y_{min}(j) &= f[V(2^{k-\gamma} \cdot (j-1))] = Y(32 \cdot (j-1)), \\ Y_{max}(j) &= f[V(2^{k-\gamma} \cdot j)] = Y(32 \cdot j) \end{aligned} \quad (4)$$

where $\gamma = 5$, $k = 10$ for codewords in 10-bit depth, $1 \leq j \leq 2^\gamma$, $j \in \mathbb{Z}$. The $Y(\cdot)$ function is given in Equation (1). Fig. 2 presents the bounds by symbol *.

Next, we calculate the probability density of pixels in each interval. Let $N(j)$ denote the number of pixels having the



Fig. 3. One frame of the HDR video ‘Market3’. The original HDR frame is in floating point representation and cannot be displayed directly in a conventional monitor. This figure is displayed by the software Qt4Image with exposure $2^{-10.4}$ [25].

luminance values belong to the interval $[Y_{min}(j), Y_{max}(j)]$, the probability density $p(j)$ is defined in Equation (5):

$$p(j) = \frac{N(j)}{\sum_{s=1}^{2^\gamma} N(s)}, \quad 1 \leq j \leq 2^\gamma. \quad (5)$$

Then, the initial number of relative luminance levels in the j th interval, $n(j)$, is calculated by rounding in Equation (6):

$$n(j) = \lfloor 2^k \cdot p(j) + 0.5 \rfloor, \quad (6)$$

where $\lfloor \cdot \rfloor$ indicates rounding downward. The absolute luminance range in the j th interval can be uniformly quantized into $n(j)$ luminance levels and each level is denoted by an integer codeword i , where $0 \leq i \leq 2^k - 1$, $k = 10$ for 10-bit depth. As expressed in Equations (5) and (6), no codewords are allocated in the j th interval when $[Y_{min}(j), Y_{max}(j)]$ has been out of the luminance range of the HDR content. Meanwhile, 2^k codewords are fully used rather than part of them as employed by PQ. More perceptual information can be preserved by the efficient use of codewords [24].

A large $n(j)$ indicates a small gap between two consecutive levels, and leads to the low reproducible contrast as calculated in Equation (3). According to Equation (6), a high $p(j)$ brings a low reproducible contrast value and vice versa. However, the initial allocation scheme based on the $p(j)$ only could result in an inefficient use of codewords and visible contrast artefacts.

Fig. 3 shows one frame of the HDR video ‘Market3’ with a peak luminance 4000 cd/m^2 . Fig. 4 presents the distribution of $p(j)$ and the reproducible contrast based on $n(j)$. Compared to 10-bit PQ, $n(j)$ contributes lower reproducible contrast values between 17 cd/m^2 and 2300 cd/m^2 . Beyond this range, the contrast derived from $n(j)$ rises up and exceeds the Schreiber threshold when the luminance is lower than 1 cd/m^2 and higher than 3100 cd/m^2 . Between 400 cd/m^2 and 1100 cd/m^2 , $n(j)$ provides a high precision and the contrast is lower than the Barten MDC. However, a better solution is desired to maintain the reproducible contrast close to the Barten MDC without falling below the threshold. Actually, the range below the perceptual MDC does not contribute to a better viewing experience and the number of relative levels in this range can be reduced without noticeable artefacts [7]. In order to efficiently employ the relative levels to preserve the perceptual contrast information, we propose to correct the initial allocation scheme based on the Barten MDC threshold.

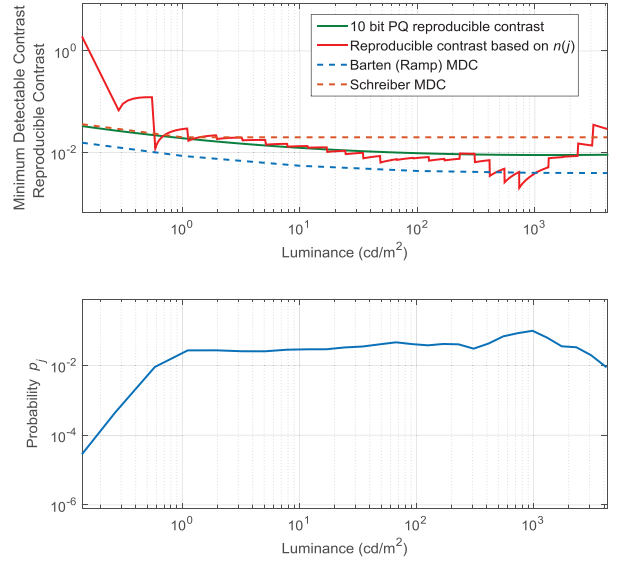


Fig. 4. The probability $p(j)$ and the reproducible contrast based on $n(j)$ with 10-bit depth codewords.

B. Corrected codeword allocation

The first concern in the corrected codeword allocation step is to determine the number of luminance levels in each interval in order to achieve the reproducible contrast as close as the Barten threshold. One solution is to calculate of contrast sensitivity values by the Barten Model, and then apply the peak of these values to compute the minimum detectable contrast (MDC). With MDC, a higher absolute luminance can be obtained and treated as a luminance level as the derivation of PQ EOTF [7]. This step is repeated in the j th interval until the upper bound $Y_{max}(j)$ is achieved. The number of levels is regarded as the number of luminance levels in this interval. However, the Barten Model involves complicated computation [7]. In practical implementation, we use 11-bit PQ to estimate the number of levels for the reproducible contrast close to Barten MDC.

As shown in Fig. 2, the 11-bit PQ has a reproducible contrast curve that is greater than but very close to the Barten threshold. As reported in [7], according to the results of the Just Noticeable Difference (JND) test, 11-bit is adequate to prevent visible artefacts using PQ. Besides, Boitard *et al.* also observed that the minimum bit-depth required to encode HDR luminance is no more than 10.5 bits for PQ to avoid perceptual loss [9]. In practical applications, the number of bits needs to be round up. Therefore, it is reasonable to regard 11-bit PQ as a reference of the perceptual threshold. With the 11-bit PQ and $\gamma = 5$, the number of luminance levels in each interval is determined to be 64, due to 1 bit increase from 10-bit PQ. Then, in 10-bit depth, the corrected number of codewords in the j th interval is within $[32, 64]$ as indicated in (7):

$$n_{adp}(j) = \begin{cases} 32, & 0 < n(j) < 32; \\ n(j), & n(j) = 0 \text{ or } 32 \leq n(j) \leq 64; \\ 64, & n(j) > 64. \end{cases} \quad (7)$$

If there are codewords unused after the allocation based on

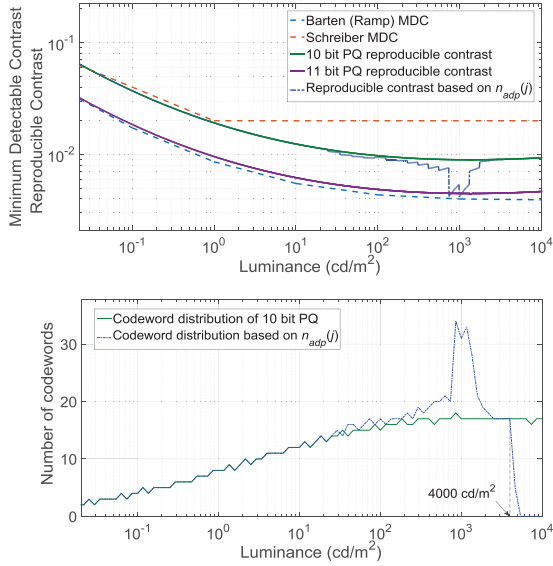


Fig. 5. Reproducible contrast and codeword distribution based on $n_{adp}(j)$ with 10-bit depth.

Equation (7), the available codewords are allocated to intervals according to probability densities, $p(j)$, from high to low, under the condition that $n_{adp}(j) \leq 64$, and until all codewords are used:

$$\sum_{j=1}^{2^\gamma} n_{adp}(j) = 2^{10}.$$

The luminance value in the j th interval $[Y_{min}(j), Y_{max}(j)]$ can be uniformly quantized into $n_{adp}(j)$ levels. Each level is denoted by an integer codeword i . We define a cumulative function of $n_{adp}(j)$ as

$$F(0) = 0, \quad F(j) = \sum_{s=1}^j n_{adp}(s), \quad 1 \leq j \leq 2^\gamma. \quad (8)$$

The range of the integer codewords in the j th interval is within $[F(j-1), F(j)]$. Each codeword i corresponds to an absolute luminance $Y_{adp}(i)$:

$$Y_{adp}(i) = Y_{min}(j) + \frac{Y_{max}(j) - Y_{min}(j)}{n_{adp}(j)} \cdot (i - F(j-1)),$$

$$F(j-1) \leq i < F(j), \quad 1 \leq j \leq 2^\gamma \quad (9)$$

Substituting $Y_{adp}(i)$ into Equation (3) yields the reproducible contrast based on $n_{adp}(j)$. Fig. 5 shows the reproducible contrast and the codeword distribution derived from $n_{adp}(j)$ for 'Market3'. Because the peak luminance of 'Market3' is 4000 cd/m^2 , the codewords, which are originally allocated to the intervals over 4000 cd/m^2 by PQ, have been moved to intervals less than 4000 cd/m^2 to reduce the reproducible contrast.

Compared to the reproducible contrast of $n(j)$ in Fig. 4, $n_{adp}(j)$ prevents the reproducible contrast from being over the 10-bit PQ to avoid producing perceivable artefact more than PQ. Besides, the contrast of $n_{adp}(j)$ does not go lower than the Barten threshold, indicating that no codewords are spent for imperceptible improvements in contrast.

In the next section, we introduce the application of the proposed perceptual quantization method to HDR video coding based on HEVC.

IV. APPLICATION TO THE HDR VIDEO CODING

In the HDR video coding chain recommended in CfE, the inverse PQ EOTF is applied to the HDR R, G and B components independently [13]. In this case, we propose an additional function, RGB mapping, before the inverse EOTF as depicted in Fig. 1. In the RGB mapping function, the value of R, G and B components is extended or narrowed based on the proposed perceptual quantization method described in Section III. Meanwhile, a set of parameters is generated for the decoder to reconstruct the HDR video. We also involve a decision step to compare the codewords of allocation for sequential frames to reduce the parameters to transmit.

A. Mapping function

In the mapping function, we firstly calculate $n(j)$ for the j th interval with regard to the probability distribution of the value of R, G and B components, and then determine the corrected number of codewords, $n_{adp}(j)$. Next, the $n_{adp}(j)$ is applied into Equation (1) to obtain an adapted luminance interval $[\hat{Y}_{min}(j), \hat{Y}_{max}(j)]$:

$$\hat{Y}_{min}(j) = f[V(F(j-1))] = Y(F(j-1)),$$

$$\hat{Y}_{max}(j) = f[V(F(j))] = Y(F(j)) \quad (10)$$

where $1 \leq j \leq 2^\gamma$, $F(\cdot)$ refers to Equation (8), $Y(\cdot)$ represents Equation (1), $k = 10$ for 10-bit depth. As depicted in Fig. 1, v_c is the original value of the R, G and B component, the mapped value v_{c_mapped} is determined by Equation (11):

$$v_{c_mapped} = \left(\hat{Y}_{max}(j) - \hat{Y}_{min}(j) \right) \cdot \frac{v_c - Y_{min}(j)}{Y_{max}(j) - Y_{min}(j)} + \hat{Y}_{min}(j) \quad (11)$$

The obtained v_{c_mapped} is sent to inverse PQ EOTF function and processed in the following steps. After down-sampling, the HDR video is represented by 10-bit depth integers and in Y'CbCr 4:2:0 format. Fig. 6 presents the Y'CbCr video frames generated by PQ only (without mapping) and the proposed quantization method (with mapping). The 10-bit video is displayed by the PYUV player [26]. From Fig. 6, we can observe that the mapped frame shows more contrast details than the one of PQ, thanks to the efficient use of codewords.

The corrected number of luminance levels in the j th interval, $n_{adp}(j)$, is transmitted as side information to the decoder side. The inverse RGB mapping function calculates the initial interval $[Y_{min}(j), Y_{max}(j)]$ and corrected interval $[\hat{Y}_{min}(j), \hat{Y}_{max}(j)]$ with the help of $n_{adp}(j)$. Let \tilde{v}_{c_mapped} denote the decoded mapped value, the reconstructed R, G and B component value \tilde{v}_c is expressed in Equation (12).

$$\tilde{v}_c = (Y_{max}(j) - Y_{min}(j)) \cdot \frac{\tilde{v}_{c_mapped} - \hat{Y}_{min}(j)}{\hat{Y}_{max}(j) - \hat{Y}_{min}(j)} + Y_{min}(j) \quad (12)$$

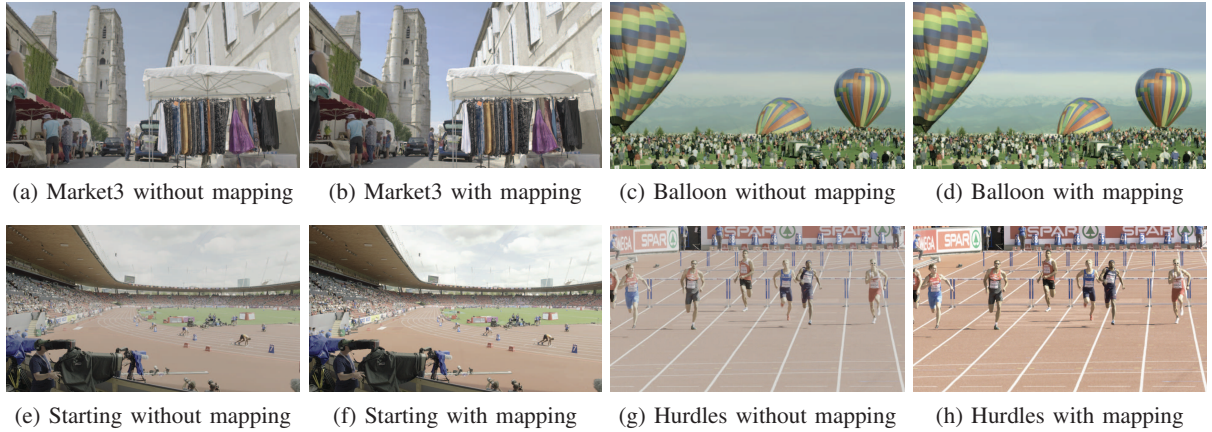


Fig. 6. HDR video frames represented by 10-bit integers. The HDR frame is desired to be displayed on HDR monitor. The images in this figure are screenshots from the conventional monitor to illustrate the difference in contrast. The display software is PYUV player [26].

In Section III-B, $n_{adp}(j)$ is within the set $\{n_{adp}(j) \in \mathbb{Z} \mid 32 \leq n_{adp}(j) \leq 64\}$, whose cardinality is 33 and requires 6 bits to represent: $2^6 = 64$. Thus, the side information of one frame requires, $6 \times 31 = 186$, bits to represent. The number of luminance levels in the 32th interval ($\gamma = 5$), $n_{adp}(32)$, can be computed as follows:

$$n_{adp}(32) = 2^k - F(31), \quad (13)$$

where $k = 10$ for the 10-bit depth. However, the adjacent frames often have similar or even the same allocation of codewords. In this case, it is not necessary to transmit $n_{adp}(j)$, but 1 bit to signal the decoder to apply the set of $n_{adp}(j)$ of the previous frame. In order to compare the values of $n_{adp}(j)$ between adjacent frames, we add a similarity comparison function at the RGB mapping block.

B. Comparison of $n_{adp}(j)$

We consider the cumulative function $F(\cdot)$ of $n_{adp}(j)$ for each frame to compare the similarity of two sequential frames. For the t th frame, we search for an interval index j_{t_min} , which is the minimum value to make $F(j)$ larger or equal to a threshold, $\alpha \cdot 2^k$:

$$j_{t_min} = \min\{j \in [1, 2^\gamma] \mid F(j) \geq \alpha \cdot 2^k\} \quad (14)$$

where α is a real value with $\alpha \in [0, 1]$, $k = 10$ for the 10-bit depth. Equation (14) indicates that there have been $\alpha \cdot 2^k$ codewords allocated from 1st to j_{t_min} th interval. We also calculate $j_{(t+1)_min}$ for the $(t+1)$ th frame. If $j_{t_min} = j_{(t+1)_min}$, the encoder side applies $n_{t,adp}(j)$ of t th frame to map the RGB value of the $(t+1)$ th frame and transmits 1 bit to the decoder side. Otherwise, the $(t+1)$ th frame is mapped based on $n_{(t+1),adp}(j)$ that need to be transmitted.

The comparison function reduces the additional bit-rate cost of the side information. Meanwhile, it can prevent the 10-bit videos from the jitter in luminance. Fig. 7 shows one example of the jitter. Without the comparison function, the 91st frame in (b) seems to be darker than the 90th frame in (a), due to the different codeword allocation results. With the application of the comparison function when $\alpha = 0.85$, the sequential frames

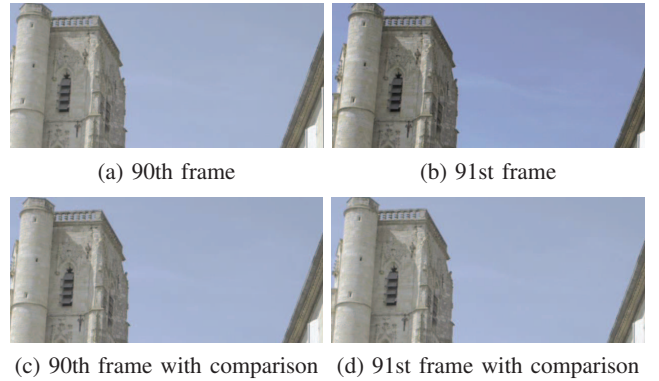


Fig. 7. Generated 10-bit Y'CbCr frames. (a) and (b) are two sequential frames without process of the comparison function; (c) and (d) are frames with the comparison function. The frames are displayed by PYUV player [26].

keep the same luminance level as shown in (c) and (d) to avoid the flickering that results in uncomfortable visual experience and disturbs the inter-prediction of HEVC [12], [27].

In order to evaluate the effectiveness of the proposed mapping function on HDR video coding, we perform an objective evaluation and conduct a subjective test to compare the qualities of HDR videos reconstructed by the PQ solution and that by the proposed mapping function.

V. EXPERIMENTAL RESULTS

We implemented the mapping function in the HDR video coding chain of CfE [13], as shown in Fig. 1. The encoding and decoding is performed by two solutions:

- The first one is based on the HM 16.2 model [28], which performs the conventional HEVC video coding. This solution is used to evaluate the coding improvement contributed only from the proposed mapping function. The HM 16.2 model is configured with the Main10 profile Random Access (RA) mode, which is used in CfE [13]. The HDR video coding chain composed of PQ and the HM 16.2 is named "Anchor 1.0" in this section.
- The second one is an optimized HM 16.7 model [28], [29]. This solution has some enhanced processing steps: percep-



Fig. 8. Examples of the sequences used in the experiment. These images are screenshots of HDR frames displayed with the software Qt4Image [25].

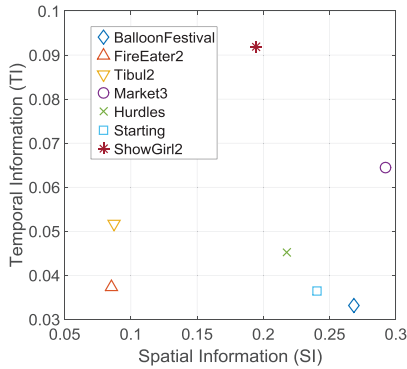


Fig. 9. Spatial information (SI) versus temporal information (TI) indexes of the test video sequences.

tual luma quantization, chroma QP (quantization parameter) offset, etc. These steps are introduced to improve the HEVC encoder for the HDR coding [19]. This solution is used to evaluate compatibility of the proposed mapping function to the recommended processing steps. The HM 16.7 also follows the RA coding conditions [29]. The HDR video coding chain composed of PQ and the optimized HM 16.7 is named “Anchor 3.2” [30].

The experimental data set consists of seven HD resolution (1920×1080) HDR video sequences. Their names and peak luminance values are: *BalloonFestival* (5000 cd/m^2), *Market3* (4000 cd/m^2), *Hurdles* (3000 cd/m^2), *Starting* (3000 cd/m^2), *FireEater2* (4000 cd/m^2), *Tibul2* (4000 cd/m^2) and *ShowGirl2* (4000 cd/m^2). The color container of *ShowGirl2* is P3D65, while BT.2020 is used in other sequences [13], [29]. Fig. 8 shows the frame examples of each content. The test video sequences have distinct spatial and temporal complexity as depicted in Fig. 9. Fig. 9 shows the spatial information (SI) and temporal information (TI) indexes computed for the luminance component of each content [31]. Before computing the SI and TI indexes, the HDR signal of each content has been converted to perceptual uniform data by PQ [7] and normalized into the range $[0, 1]$. *FireEater2* and *Tibul2* have relatively small SI and TI values, *BalloonFestival*, *Starting* and *Hurdles* have larger SI values, *Market3* shows higher complexity for both measures, while *ShowGirl2* has the highest TI index.

TABLE I
BIT-RATES OF THE SIDE INFORMATION FOR EACH VIDEO CONTENT.

| Name | Frames per second | N_{key} | Bit-rate (kbps) |
|-----------------|-------------------|-----------|-----------------|
| BalloonFestival | 24 | 1 | 0.043 |
| Market3 | 50 | 4 | 0.166 |
| Hurdles | 50 | 1 | 0.069 |
| Starting | 50 | 10 | 0.235 |
| FireEater2 | 25 | 1 | 0.048 |
| Tibul2 | 30 | 1 | 0.053 |
| ShowGirl2 | 25 | 21 | 0.312 |

Each video is processed by two types of the HDR coding chains, one is the HDR coding system with the PQ only, the other one is the coding system with the proposed mapping function. The proposed mapping function needs to send the information of the codeword allocation for each frame. Thanks to the comparison function introduced in Section IV-B, the additional bit-rate of this side information is quite restricted. Table I provides the bit-rates of the side information for each video content. The value of N_{key} represents the number of frames, which send the information of codeword allocation, $n_{adp}(j)$. The other frames only transmit 1 bit for each to indicate that they use the allocation method of the previous frame. The additional bit-rates of the side information have been added into the bit-rate of the proposed mapping function.

Table II shows the bit-rate cost of each coding chain when the HEVC is performed by HM 16.2. The bit-rate of the proposed solution is lower than that of PQ. Table III provides the bit-rate cost by the optimized HM 16.7 encoding.

A. Objective Evaluation

In the objective evaluation, the original HDR content is used as the reference video to evaluate the reconstructed HDR video by objective metrics. Hanhart *et al.* benchmarked objective quality metrics for HDR content and demonstrated that HDR-VDP2 [32] and HDR-VQM [33] were reliable predictors of perceived quality [34]. Azimi *et al.* investigated the performance of existing quality metrics in evaluating the quality of HDR content [35]. They concluded that HDR-VDP2 and VIF [36] with PU encoding [37] (PU-VIF) gave results well correlated with the subjective evaluation. Recently, Zerman *et al.* conducted an extensive evaluation of full-reference HDR

TABLE II
BIT-RATES OF TEST VIDEOS WHEN HEVC IS PERFORMED BY HM 16.2. $\Delta R = (Rate(Proposed) - Rate(PQ))/Rate(PQ)$.

| Video | Rate1 (kbps) | | | Rate2 (kbps) | | | Rate3 (kbps) | | | Rate4 (kbps) | | |
|------------|--------------|----------|--------------|--------------|----------|--------------|--------------|----------|--------------|--------------|----------|--------------|
| | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % |
| Balloon | 6639.6 | 6631.3 | -0.125 | 3763.4 | 3737.0 | -0.701 | 2154.5 | 2147.0 | -0.348 | 1275.4 | 1273.6 | -0.141 |
| Market3 | 7914.5 | 7892.3 | -0.280 | 4219.6 | 4209.8 | -0.232 | 2313.0 | 2302.6 | -0.450 | 1248.1 | 1246.1 | -0.160 |
| Starting | 3213.2 | 3205.0 | -0.255 | 1730.0 | 1727.3 | -0.156 | 1028.5 | 1016.8 | -1.138 | 618.5 | 615.2 | -0.534 |
| Hurdles | 6311.8 | 6293.6 | -0.288 | 3312.5 | 3306.0 | -0.196 | 1818.9 | 1791.5 | -1.506 | 1052.2 | 1043.6 | -0.817 |
| FireEater2 | 1912.3 | 1909.2 | -0.162 | 1257.2 | 1219.5 | -2.999 | 808.6 | 803.6 | -0.618 | 519.8 | 517.7 | -0.404 |
| Tibul2 | 6090.6 | 5853.4 | -3.895 | 2499.8 | 2379.9 | -4.796 | 970.1 | 929.0 | -4.237 | 402.4 | 387.0 | -3.827 |
| ShowGirl2 | 3317.1 | 3188.2 | -3.886 | 1652.0 | 1626.1 | -1.568 | 969.4 | 948.3 | -2.177 | 573.9 | 566.6 | -1.272 |

TABLE III
BIT-RATES OF TEST VIDEOS WHEN HEVC IS PERFORMED BY OPTIMIZED HM 16.7. $\Delta R = (Rate(Proposed) - Rate(PQ))/Rate(PQ)$.

| Video | Rate1 (kbps) | | | Rate2 (kbps) | | | Rate3 (kbps) | | | Rate4 (kbps) | | |
|------------|--------------|----------|--------------|--------------|----------|--------------|--------------|----------|--------------|--------------|----------|--------------|
| | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % | PQ | Proposed | ΔR % |
| Balloon | 3838.4 | 3825.16 | -0.346 | 2271.3 | 2269.5 | -0.080 | 1571.9 | 1569.4 | -0.159 | 1238.2 | 1236.4 | -0.145 |
| Market3 | 7897.8 | 7979.7 | 1.037 | 2723.0 | 2788.9 | 2.421 | 1754.4 | 1752.4 | -0.115 | 1301.1 | 1307.4 | 0.489 |
| Starting | 2685.1 | 2537.3 | -3.367 | 1579.2 | 1562.6 | -1.051 | 881.6 | 924.0 | 4.809 | 556.0 | 554.0 | 0.360 |
| Hurdles | 6349.9 | 6132.8 | -5.504 | 3526.3 | 3409.8 | -3.304 | 1770.6 | 1723.6 | -2.665 | 1090.0 | 1062.9 | -2.486 |
| FireEater2 | 1927.9 | 1942.8 | 0.772 | 1291.3 | 1300.9 | 0.738 | 852.0 | 856.9 | 0.577 | 560.1 | 563.4 | 0.596 |
| Tibul2 | 6758.6 | 6671.7 | -1.285 | 2776.4 | 2572.3 | -7.354 | 1093.4 | 1062.1 | -2.856 | 474.5 | 470.2 | -0.920 |
| ShowGirl2 | 3179.4 | 3102.5 | -2.417 | 1626.3 | 1640.4 | 0.865 | 969.6 | 990.2 | 2.126 | 589.6 | 612.8 | 3.930 |

quality metrics [38]. Their analysis shows that HDR-VQM and HDR-VDP2 provide accurate predictions of subjective scores for compression-like distortion. Therefore, we apply HDR-VDP2, HDR-VQM and PU-VIF in this test.

1) *HDR-VDP2*: Fig. 10 shows the quality scores measured by HDR-VDP2. Because HDR-VDP2 is designed to evaluate HDR images [32], we applied it to calculate the quality score of each frame and used the average value of all frames as the quality score of the video.

As depicted in Fig. 10, the proposed method outperforms PQ to reproduce HDR videos with higher quality for *BalloonFestival*, *Market3*, *Hurdles* and *Starting* contents, by the HEVC encoding of both HM 16.2 and optimized HM 16.7. For *FireEater2* and *ShowGirl2*, the quality scores between PQ and the proposed method are close at low bit-rates, while PQ gains better scores at high bit-rates. PQ also obtains higher scores for *Tibul2*.

2) *HDR-VQM*: HDR-VQM considers the temporal coherence and gives the overall quality estimation of the HDR video. Fig. 11 presents the quality assessment by the HDR-VQM and shows that the proposed method is preferred for *BalloonFestival*, *Hurdles*, *Starting*, *FireEater2* and *Tibul2* videos. HDR-VQM does not provide the significant quality difference for *Market3*. For *ShowGirl2*, the proposed method gains desirable evaluation at low bit-rates, while PQ becomes better for high bit-rates.

3) *PU-VIF*: The VIF is designed for the LDR image quality assessment [36]. Therefore, the Perceptual Uniform (PU) encoding [37] is used to transform luminance values into approximately perceptually uniform LDR values [35]. After that, the transformed image is measured by the VIF metric.

Fig. 12 shows the results of VIF. With the conventional HM 16.2 encoding, the proposed method obtains higher scores for contents: *BalloonFestival*, *Market3*, *Hurdles* and *FireEater2*. The difference between the performance of PQ (Anchor 1.0)

and the proposed method is not obvious for *Starting* and *Tibul2*. For *ShowGirl2*, PQ gives higher scores at high bit-rates.

In terms of the optimized HM 16.7 encoding, The proposed method also has higher scores for *Hurdles* and *FireEater2*, while slight advantage for *BalloonFestival*, and *Market3*. PQ (Anchor 3.2) provides slightly higher scores for *Starting* and *Tibul2*. PQ also shows better performance at high bit-rates for *ShowGirl2*.

Besides, we calculated the BD-Rate [39] gain of the proposed method. The BD-Rate is used to express the average bit-rate difference in percentage over a range of qualities. Regarding the bit-rates of PQ as reference, a positive BD-Rate value indicates an additional bit-rate required to achieve one quality range in comparison with PQ, while a negative value refers to a reduction in bit-rate. For HDR-VQM, we used the Equation (15) [40],

$$HDRVQM_{normalized} = \frac{4}{1 + exp(HDRVQM)} - 1 \quad (15)$$

to get the normalized HDR-VQM score, which equals to 1 for perfect quality and decrease for degraded quality. The normalized score was used in the computation of BD-Rate.

Table IV shows the BD-Rate results with the HM 16.2 encoding. Compared to PQ, the BD-Rate gained by the proposed method, in average, are 5.7% measured by HDR-VDP2, 15.1% by HDR-VQM and 4.6% by PU-VIF.

Table V shows the BD-Rate results with the optimized HM 16.7 encoding. The BD-Rate gains of the proposed method are 2.4% by HDR-VDP2, 15.3% by HDR-VQM and 0.9% by PU-VIF.

With regard to Fig. 10, Fig. 11 and Fig. 12, the comparison between the proposed method and PQ is similar with both HM 16.2 and optimized HM 16.7 encoding. Comparing the average BD-Rate gains of HDR-VDP2 and PU-VIF, the proposed method behaves better in Table IV than V. The

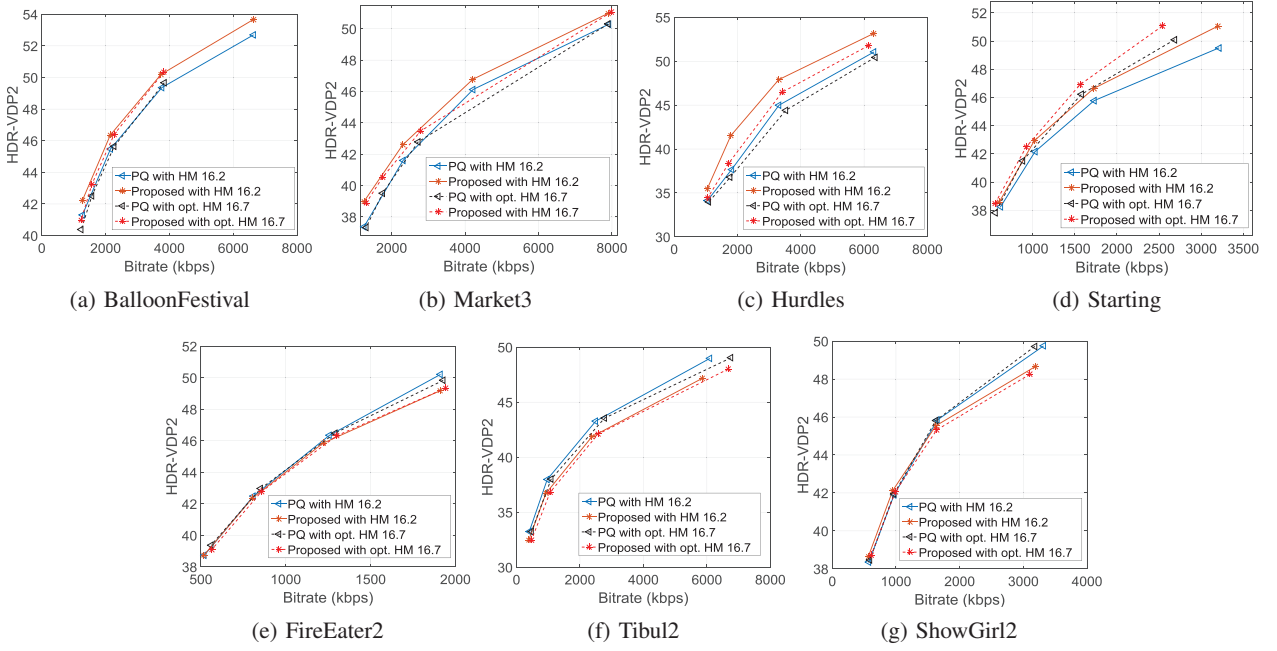


Fig. 10. Quality assessment by HDR-VDP2. A higher VDP score indicates a better quality.

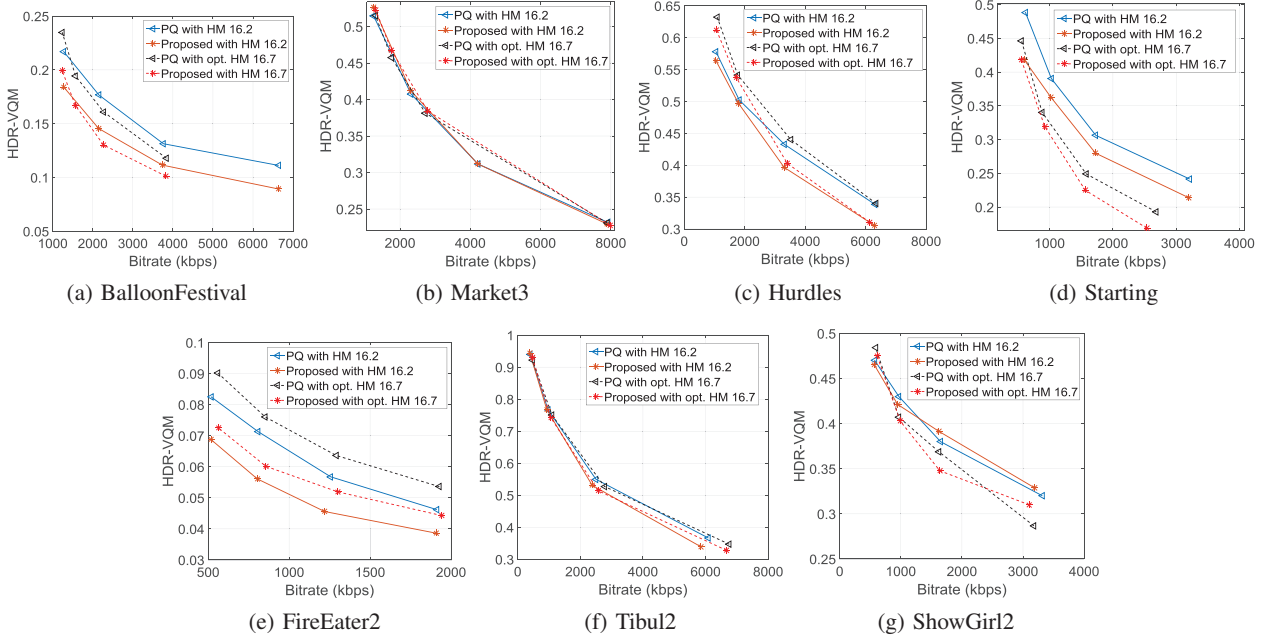


Fig. 11. Quality assessment by HDR-VQM. A lower VQM value indicates a better quality.

reason probably relies on the luma QP or chroma QP offset modifications used in the optimized HM 16.7 [19]. These modifications have shifted bits from dark regions to bright regions with the purpose of ameliorating luma and chroma artefacts, which decreases the relative improvement contributed by our proposed method for bright region. Meanwhile, our proposed method achieves higher BD-Rate gain evaluated by HDR-VQM, which considers the temporal coherence in the video, with both the HM 16.2 and the optimized HM 16.7 encoding solutions.

In order to compare the performance of the proposed method with the Reshaper [17] and adaptive PQ [20], we also calculated the BD-Rate gains measured by a set of PSNR metrics: tPSNR, tOSNR, PSNR_DE100, PSNR_MD100 and PSNR_L100 as used in the document [17]. tPSNR calculates the square error (SE) between the reconstructed and original HDR image/frame for each color component. The average of SE of color components is used to calculate the PSNR value for the tPSNR score [13]. tOSNR computes the root of SE for each pixel and sums all root values to measure the overall

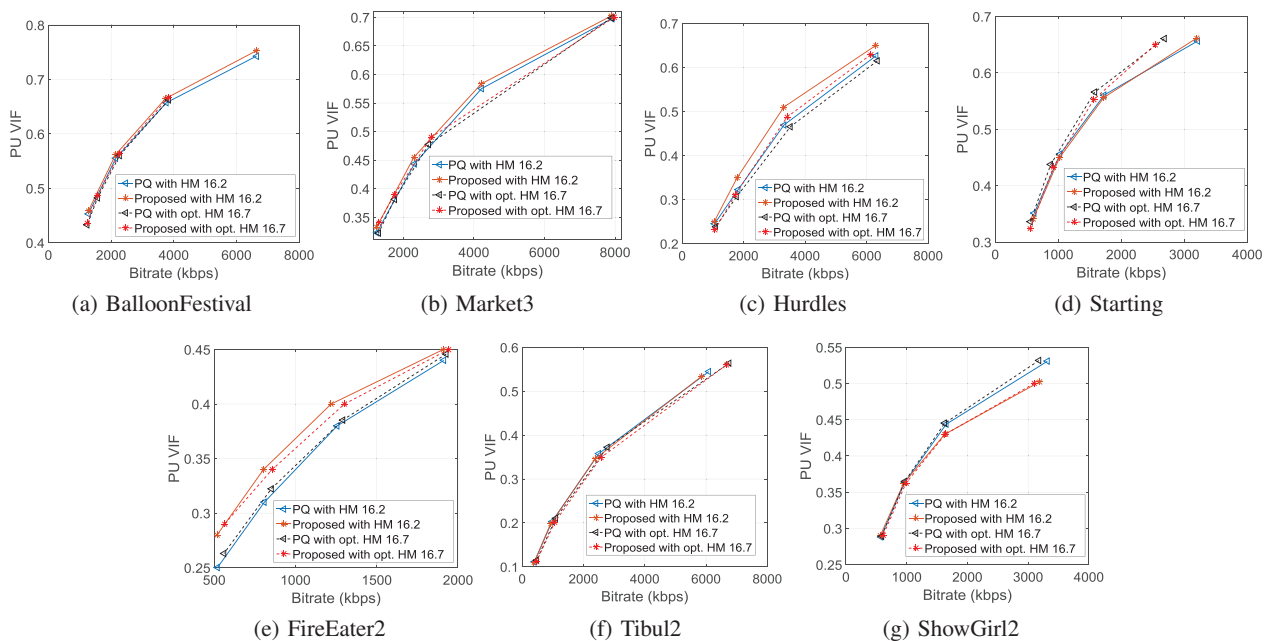


Fig. 12. Quality assessment by PU-VIF. A higher PU-VIF value indicates a better quality.

TABLE V

BD-RATE GAINS OF THE PROPOSED METHOD WITH HM 16.2 AS REFERENCE, MEASURED BY HDR-VDP2, HDR-VQM AND PU-VIF.

| video | HDR-VDP2 (%) | HDR-VQM (%) | PU-VIF (%) |
|----------------|--------------|--------------|-------------|
| Balloon | -12.2 | -33.2 | -4.4 |
| Market3 | -12.2 | 1.5 | -4.8 |
| Hurdles | -27.2 | -15.3 | -13.2 |
| Starting | -11.7 | -15.6 | 1.5 |
| FireEater2 | 2.4 | -38.7 | -16.6 |
| Tibul2 | 21.0 | -8.7 | 0.5 |
| ShowGirl2 | 0.2 | 4.6 | 5.2 |
| Average | -5.7 | -15.1 | -4.6 |

TABLE V

BD-RATE GAINS OF THE PROPOSED METHOD WITH HM 16.7 AS REFERENCE, MEASURED BY HDR-VDP2, HDR-VQM AND PU-VIF.

| video | HDR-VDP2 (%) | HDR-VQM (%) | PU-VIF (%) |
|----------------|--------------|--------------|-------------|
| Balloon | -8.6 | -26.4 | -2.0 |
| Market3 | -8.3 | 3.3 | -3.3 |
| Hurdles | -18.5 | -15.6 | -8.1 |
| Starting | -8.9 | -11.7 | 5.7 |
| FireEater2 | 3.4 | -41.5 | -11.0 |
| Tibul2 | 17.9 | -8.7 | 3.5 |
| ShowGirl2 | 6.1 | -6.2 | 9.1 |
| Average | -2.4 | -15.3 | -0.9 |

difference between two images/frames. The overall difference value is used to generate a PSNR value for the tOSNR score. PSNR_DE100 is applied to measure the color difference. It is defined as a PSNR value and calculated by the deltaE2000 metric [41]. It is derived with 100 as reference luminance value [13]. PSNR_MD100 uses the maximum deltaE2000 distance of each frame to calculate PSNR. PSNR_L100 is the PSNR of mean square error of L component of the CIELAB [42] color space that is also used for the deltaE2000 metric. These

TABLE VI

COMPARISON BETWEEN THE ADAPTIVE PQ [20] AND THE PROPOSED METHOD ON THE BD-RATE GAIN.

| Video / Method | tPSNR-XYZ (%) | | PSNR_DE (%) | |
|-----------------|---------------|--------------|--------------|--------------|
| | [20] | proposed | [20] | proposed |
| Market3 | -3.37 | -3.2 | -7.44 | -9.6 |
| Tibul2 | -3.17 | -0.9 | -5.04 | -4.0 |
| BalloonFestival | -1.60 | -2.2 | -3.43 | -10.2 |
| FireEater2 | -5.88 | -14.3 | -6.12 | -19.8 |
| Average | -3.51 | -5.15 | -5.51 | -10.9 |

metrics are computed using HDRTools (v0.11) [13], [43].

With the HM 16.2 encoding, Yu *et al.* reported a BD-Rate gain to PQ measured by tPSNR and PSNR_DE as shown in Table VI [20]. It can be noticed that the proposed method has better average results than the adaptive PQ [20].

The report [17] provides the available BD-Rate gain of the Reshaper applied in the Y'CbCr color space as used in our proposed method. Table VII shows the results of the Reshaper and proposed method in comparison with Anchor 3.2. Both Reshaper and the proposed method are joint with optimization of encoder. The proposed method achieves the average BD-Rate gain for tOSNR, PSNR_DE100 and PSNR_MD100, while the Reshaper was reported to need more bit-rate resource in average measured by tOSNR and PSNR_DE100 [17]. The proposed method increases the BD-Rate by 2.5% in terms of tPSNR, where the Reshaper needs 6.7% in addition. For the metric, PSNR_L100, the Reshaper gains the bit-rate by 4.9%, which is better than that of the proposed method. According to PSNR_DE100, the proposed method behaves well in color reproduction. This advantage is probably due to the fact that the proposed method applies the mapping adjustment in the RGB components. This action also preserves the contrast information in the original color space.

TABLE VII
BD-RATE GAIN OF THE RESHAPER [17] AND THE PROPOSED METHOD IN COMPARISON WITH ANCHOR 3.2.

| video | tPSNR-XYZ (%) | | tOSNR-XYZ (%) | | PSNR_DE100 (%) | | PSNR_MD100 (%) | | PSNR_L100 (%) | |
|-----------------|---------------|----------|---------------|----------|----------------|----------|----------------|----------|---------------|----------|
| | [17] | proposed | [17] | proposed | [17] | proposed | [17] | proposed | [17] | proposed |
| FireEater2 | 1.0 | -11.5 | 0.0 | -33.1 | -6.3 | -21.9 | -17.7 | -17.9 | 3.4 | 9.7 |
| Market3 | 2.8 | -2.1 | -1.0 | -4.2 | 12.8 | -21.6 | -80.3 | -59.7 | -10.8 | -9.9 |
| ShowGirl2 | 11.7 | 13.7 | 16.1 | 10.0 | -4.9 | 17.7 | -2.8 | -47.7 | -0.3 | 14.5 |
| BalloonFestival | 13.7 | -1.7 | 9.2 | -5.2 | 19.0 | -17.7 | -21.9 | 11.5 | -5.1 | -8.1 |
| Hurdles | 0.9 | 8.4 | -2.6 | -1.5 | 28.5 | -13.5 | 64.9 | 68.9 | -11.0 | -1.4 |
| Starting | 9.9 | 8.2 | 6.2 | 1.3 | 50.3 | -30.9 | -17.0 | -37.6 | -5.7 | -2.9 |
| Average | 6.7 | 2.5 | 4.7 | -5.5 | 16.6 | -14.7 | -12.5 | -13.8 | -4.9 | 0.3 |

TABLE VIII
QUALITY SCORES OF HDR VIDEOS WITHOUT COMPRESSION, MEASURED BY PSNR METRICS.

| video | tPSNR-XYZ | | tOSNR-XYZ | | PSNR_DE100 | | PSNR_MD100 | | PSNR_L100 | |
|-----------------|-----------|----------|-----------|----------|------------|----------|------------|----------|-----------|----------|
| | PQ | proposed | PQ | proposed | PQ | proposed | PQ | proposed | PQ | proposed |
| FireEater2 | 56.601 | 57.643 | 55.014 | 56.154 | 49.155 | 49.227 | 24.725 | 25.230 | 58.381 | 58.539 |
| Market3 | 50.267 | 50.486 | 49.886 | 50.044 | 36.788 | 36.765 | 21.874 | 22.089 | 47.892 | 48.540 |
| Tibul2 | 54.433 | 55.919 | 53.726 | 55.890 | 46.557 | 47.562 | 29.548 | 29.667 | 53.072 | 53.457 |
| ShowGirl2 | 49.051 | 48.875 | 48.011 | 47.953 | 39.411 | 39.374 | 22.505 | 22.513 | 53.010 | 53.109 |
| BalloonFestival | 49.266 | 49.544 | 51.725 | 52.218 | 40.679 | 40.904 | 21.151 | 21.043 | 48.243 | 48.843 |
| Hurdles | 50.707 | 50.925 | 49.175 | 49.276 | 36.865 | 36.923 | 22.234 | 22.320 | 47.341 | 47.575 |
| Starting | 48.068 | 48.120 | 47.745 | 47.753 | 36.115 | 36.195 | 21.276 | 21.300 | 47.707 | 47.925 |
| Average | 51.199 | 51.644 | 50.755 | 51.327 | 40.796 | 40.993 | 23.330 | 23.452 | 50.807 | 51.141 |

Besides the comparison based on the video coding, the distortion involved by PQ and the proposed method is also evaluated by objective metrics. In this case, the quantized HDR data, which has the 10-bit depth precision, is de-quantized back to floating data directly. The reconstructed HDR video is compared with the reference one to give a quality score. The optimized modifications for the pre-processing are not used in this comparison. We aim to evaluate the improvement contributed by the proposed method.

Table VIII shows the quality scores measured by PSNR metrics. The proposed method achieves higher scores in most comparisons, while it shows slightly lower scores than PQ for *Market3* in PSNR_DE100, for *BalloonFestival* in PSNR_MD100. For the video *ShowGirl2*, PQ has better quality measured by tPSNR-XYZ, tOSNR-XYZ and PSNR-DE100. The source data of the video *ShowGirl2* are not provided in linear domain to record the absolute luminance, but has been processed by PQ with 12-bit precision [13], [29]. It has preserved contrast information desired by PQ and probably leads to a contribution for the PQ 10-bit depth quantization. Except *ShowGirl2*, the source data of other videos are provided in linear domain. In average, the proposed method achieves higher scores for all metrics in Table VIII.

Table IX provides the evaluation performed by HDR-VDP2, HDR-VQM and PU-VIF. The proposed method also shows better average quality in this table. Meanwhile, PQ obtains higher quality assessment for the video *ShowGirl2*. The metric PU-VIF gives values larger than unity to the proposed method for the videos *FireEater2* and *Tibul2*. It indicates a contrast enhancement of the reference image in terms of VIF [36].

The processing time consumed by PQ and the proposed method is also provided on a computer with Windows 7 operating system, Inter(R) i7-6700 CPU (3.4GHz) and 32GB RAM. The processing time is shown in Table X. Compared

TABLE IX
QUALITY ASSESSMENT BY HDR-VDP2, HDR-VQM AND PU-VIF.

| video | VDP2 | | VQM ($\cdot 10^{-3}$) | | PU-VIF | |
|----------------|-------|-------|-------------------------|-------|--------|--------|
| | PQ | Prop. | PQ | Prop. | PQ | Prop. |
| FireEater2 | 69.37 | 74.98 | 0.287 | 0.118 | 0.9993 | 1.0027 |
| Market3 | 84.86 | 85.05 | 1.414 | 1.161 | 0.9993 | 0.9991 |
| Tibul2 | 63.62 | 67.44 | 1.979 | 1.572 | 0.9994 | 1.0005 |
| ShowGirl2 | 74.87 | 74.50 | 2.088 | 2.259 | 0.9994 | 0.9987 |
| Balloon | 73.78 | 74.17 | 2.889 | 2.675 | 0.9995 | 0.9998 |
| Hurdles | 88.31 | 88.70 | 2.622 | 2.423 | 0.9992 | 0.9994 |
| Starting | 81.42 | 81.61 | 2.214 | 2.112 | 0.9993 | 0.9995 |
| Average | 76.60 | 78.06 | 1.928 | 1.760 | 0.9993 | 0.9999 |

TABLE X
THE PROCESSING TIME CONSUMED BY PQ AND THE PROPOSED METHOD FOR THE QUANTIZATION OF HDR VIDEOS. UNIT: MILLISECOND (MS).
 $\Delta T = (time(Proposed) - time(PQ)) / time(PQ)$

| video | time (ms) | | ΔT % |
|----------------|-----------|----------|--------------|
| | PQ | Proposed | |
| FireEater2 | 816077 | 886137 | 8.58 |
| Market3 | 1637859 | 1790833 | 9.34 |
| Tibul2 | 1180530 | 1227439 | 3.97 |
| ShowGirl2 | 1338932 | 1399460 | 4.52 |
| Balloon | 995342 | 1086462 | 9.15 |
| Hurdles | 2208726 | 2538400 | 14.93 |
| Starting | 1915952 | 2102513 | 9.74 |
| Average | | | 8.61 |

with PQ, the proposed mapping function slightly increases the processing time by 8.61% in average. This additional cost of time can be decreased by the optimization of the code and the using of look-up tables.

Although objective evaluation methods predict the visual quality of the HDR videos from different ways, it is still desirable to observe the performance of the proposed method in the HDR display. Due to the high quality of the reconstructed

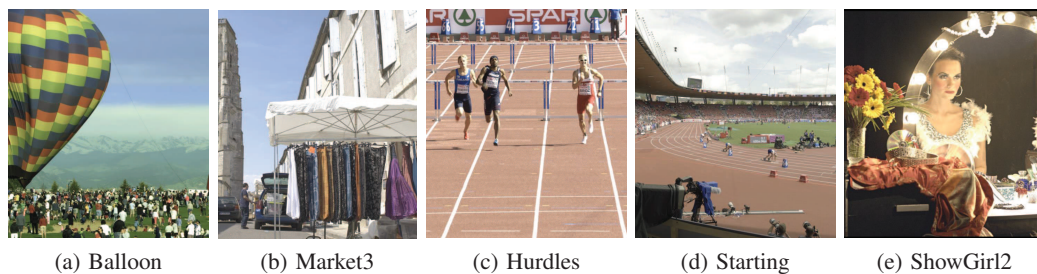


Fig. 13. Representative frames of the cropped sequences used in the subjective test. These images are screenshots of the conventional monitor. The HDR frames are displayed with the software Qt4Image [25].

HDR content without the HM coding, it is not easy to detect the distortion in the video. Therefore, the HDR videos decoded from the HEVC bit-stream are used to evaluate the subjective quality. The proposed work aims to provide contributions to the CfE [13]. Meanwhile, the proposed method shows similar performance in comparison with PQ after the encoding by HM 16.2 and optimized HM 16.7, in terms of HDR-VDP2, HDR-VQM and PU-VIF. The decoded HDR videos by the use of HM 16.2 are used for comparison in the following subjective evaluation section.

B. Subjective Evaluation

In order to evaluate, subjectively, the efficiency of the proposed solution, we have performed a set of subjective quality experiment, involving human viewers assessing the perceived quality. Below, test environment and evaluation procedure are described in detail.

1) *Test Environment*: Following the ITU-R BT.500-13 Recommendation [44], the subjective tests have been conducted in a dark, quiet room, with the ambient illumination of the room at 2.154 lux and 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 [45] with 1920×1080 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.

2) *Test Methodology*: We used the Paired Comparison (PC) methodology to evaluate the stimuli. Two videos of the same content were presented simultaneously in side-by-side fashion. Each video was clipped into 8 seconds duration and cropped to 952×1080 pixels with 16 pixels of black border to separate two videos. The cropped frame examples of the videos are shown in Fig. 13.

In the paired comparison, one of the two videos was always the reconstructed sequence by PQ, the other one was the proposed solution at the closest bit-rate. Each pair of videos was presented twice, having each video displayed on both sides (PQ vs. Proposed and Proposed vs. PQ). In each trial, the observer was asked to judge which video in pair ('left' or 'right') has a better overall quality in consideration of the textured areas and color. The option 'same' was not included to force users to choose one of the stimuli. As a complement, if the observer made two opposite choices, we considered this

pair of videos as a tie judged by this observer. Otherwise, the video selected twice was regarded to have better visual quality.

3) *Test Procedure*: The first section is a training part to familiarize the observers with the evaluation procedure. In the training, example videos from the *FireEater2* and *Tibul2* sequences, which have relatively low SI and TI indexes, were used to show the visible distortion. The original videos were presented to illustrate the high quality HDR content. Next, the original video and the degraded video with high perceivable distortion were presented side by side to show the loss in texture and color artefacts. For the same content, two videos, one with high distortion and the other with less distortion, were also compared to train the observers to select the video with better quality. The training videos were not used for test.

In the test, videos from *BalloonFestival*, *Market3*, *Hurdles*, *Starting* and *ShowGirl2* sequences were compared. The orders of displayed videos were randomized and designed to prevent the same content from being shown consecutively in order to reduce the contextual effects as suggested in [46]. 16 observers (8 females and 8 males), aged between 23 to 39 years old, took part in the test. They were also checked for both visual acuity and color blindness.

4) *Statistical Analysis*: The results of the subjective experiment have been analyzed for the detection of the outliers. we used *pwcmp* software's outlier detection mechanism [47], which is based on the relative likelihood of each observer's selection. According to the analysis, there is not any outlier found. We apply the statistical analysis proposed in [46] to analyze the results of the subjective test and determine whether the visual quality difference between two coding conditions is statistically significant.

Let w_{PQ} denote the winning frequency of PQ (observers selected PQ for better quality), w_{pro} the winning frequency of the proposed solution (observers selected the proposed solution for better quality), t the tie frequency, and $w_{PQ} + w_{pro} + t = N_s$, where N_s is the number of observers. The preference probability of the proposed solution is defined in Equation (16):

$$p_{pro} = \frac{w_{pro}}{N_s} + \frac{t}{2N_s}. \quad (16)$$

The initial hypothesis is that PQ and the proposed method have the equal probability to be preferred, i.e., they are the outcome of a Bernoulli process $B(N, p)$, where $N = N_s$ and $p = 0.5$. Thus, p_{pro} is the winning probability in a sequence of Bernoulli trials, and follows a Binomial distribution. In

order to accept or reject the initial hypothesis, we perform a Binomial test on the collected data. To this end, the Cumulative Distribution Function (CDF) of the Binomial distribution is used to find the critical thresholds for the determination of significance.

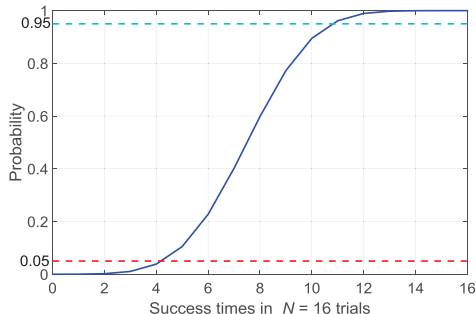


Fig. 14. Cumulative distribution function for Binomial distribution with $N = 16$ and $p = 0.5$.

Let $B(n, N, p)$ denote the CDF of the Binomial distribution with $N = 16$ and $p = 0.5$, where n is the number of preferences of a method. Fig. 14 gives the probability curve of $B(n, N, p)$. We select 95% as the level of significance of the test. Since $B(11, 16, 0.5) = 0.9616$, if there are 11 or more votes in favor of video from the proposed method, we consider this video giving significant better quality than PQ. Similarly, 5% is used to decide whether the quality of the proposed method is worse than PQ. Due to $B(4, 16, 0.5) = 0.0384$ and $B(5, 16, 0.5) = 0.1051$, if there are 4.5 or less votes for the video, we consider it having worse quality than PQ.

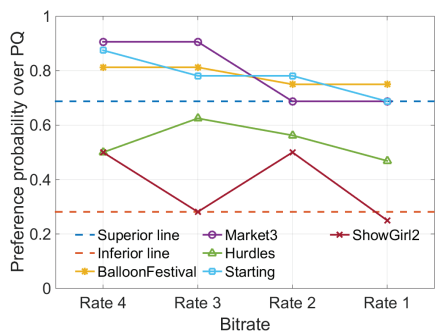


Fig. 15. Preference probability of selecting the proposed version over PQ, when HM 16.2 is used. Rate 4 < Rate 3 < Rate 2 < Rate 1.

5) *Results*: Fig. 15 presents the preference probability of selecting videos generated by the proposed method over PQ for each HDR content. The superior line refers to the probability of high quality: $11/16 = 0.6875$, while the inferior refers to the low quality: $4.5/16 = 0.2812$. Values on or above the superior line indicate statistically significant visual quality better than PQ. Meanwhile, values on or below the inferior line show statistically significant lower quality compared with PQ. Values between the two lines imply that there is no statistically significant difference in the quality of the two methods.

For *BalloonFestival*, *Market3* and *Starting* video contents, the proposed method provides statistically significant superior visual quality than PQ at all bit-rates. The advantage is



(a) The proposed method (b) PQ

Fig. 16. Parts of the reconstructed HDR frames from *Market3* at Rate 4. These screenshots are from conventional monitor with low color gamut. The texture information is better represented by the proposed method than PQ.

more obvious at low bit-rates. For *Hurdles*, the proposed method gets more votes at Rate 2 and 3 than PQ. However, its advantage is not significant enough to be distinguished. Regarding the *ShowGirl2* content, the proposed method does not show improvements on visual quality compared with PQ, and its videos at Rate 1 and 3 seems to bring more distortion.

6) *Results Analysis*: The proposed method outperforms PQ for *BalloonFestival*, *Market3* and *Starting*. The reconstructed HDR videos from the proposed method represents better texture information than that of PQ as illustrated in Fig. 16. However, the mapping function does not show advantage for the video *ShowGirl2*, where the proposed method performs equivalently to PQ or worse (at highest bit-rate). This merits some additional considerations. Indeed, in the preparation of the test material for the experiments we tried to compare the methods at the same (or as similar as possible) bit-rate, by changing the sequence-level quantization parameters (QP) accordingly. For the sake of fairness, we coded all the sequences with the proposed method at a bit-rate not greater than that used for the same condition coded with PQ. In the case of *ShowGirl2*, due to the high TI of this sequence (Fig.9), it was not possible to accurately align the bit-rates of the PQ and the proposed method, and the closest bit-rate obtained for the proposed method is 1.27% to 3.89% (for the highest bit-rate) lower than PQ. This is enough to visually affect the result of coding with the proposed scheme. Nevertheless, the results of the subjective experiment show that, despite the lower bit-rate of the proposed adaptive quantization scheme, there is no statistically significant loss of quality.

VI. CONCLUSION

We presented a new perceptual quantization method to efficiently employ the codewords to preserve the contrast information in an HDR content with integer representation. This method adapts well to the practical HDR content due to the consideration of both luminance distribution and Minimum Detectable Contrast. We proposed a mapping function to apply this method to an HDR video coding solution based on HEVC. Thanks to a comparison scheme embedded in the mapping function, the similarity of sequential frames is used to decrease the additional bit-rate cost of the side information and to avoid flickering. The experimental results show that the proposed

method can significantly improve the visible quality of some reconstructed HDR videos. Meanwhile, average bit-rate gains are also achieved for the objective evaluation metrics used in the measurement.

REFERENCES

- [1] P. Ledda, "Product review: High dynamic range displays," *Presence*, vol. 16, no. 1, pp. 119–122, Feb. 2007.
- [2] Consumer Technology Association, "Next generation of innovation debuts as CES 2016 kicks off," in *CES Press Release*. Consumer Electronics Show, 2016.
- [3] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010.
- [4] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H. Seidel, "Perception-motivated high dynamic range video encoding," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: ACM, 2004, pp. 733–741.
- [5] R. Mantiuk, K. Myszkowski, and H. Seidel, "Lossy compression of high dynamic range images and video," in *Human Vision and Electronic Imaging XI*, Feb. 2006, pp. 60570V–10.
- [6] P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," vol. 5294, 2003, pp. 231–238.
- [7] 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.
- [8] D. Touz , S. Lasserre, Y. Olivier, R. Boitard, and E. Fran ois, "HDR video coding based on local LDR quantization," in *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, Sarajevo, 2014.
- [9] R. Boitard, R. Mantiuk, and T. Pouli, "Evaluation of color encodings for high dynamic range pixels," in *Human Vision and Electronic Imaging*, vol. 9394, 2015, pp. 93 941K–9.
- [10] C. Poynton and B. Funt, "Perceptual uniformity in digital image representation and display," *Color Research & Application*, vol. 39, no. 1, pp. 6–15, 2014.
- [11] *High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*, Society of Motion Picture and Television Engineers Std. SMPTE ST 2084:2014, Aug. 2014.
- [12] 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.
- [13] A. Luthra, E. Fran ois, and W. Husak, "Call for Evidence (CfE) for HDR and WCG Video Coding," International Organization for Standardization, WG11 N15083, Feb. 2015.
- [14] E. Fran ois, 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.
- [15] SMPTE Technical Committee 10E SG, "Study Group Report High-Dynamic-Range (HDR) Imaging Ecosystem," Society of Motion Picture and Television Engineers, Tech. Rep., Sep. 2015.
- [16] K. Minoo, T. Lu, P. Yin, L. Kerofsky, D. Rusanovskyy, and E. Fran ois, "Description of the reshapener parameters derivation process in ETM reference software," Joint Collaborative Team on Video Coding (JCT-VC), Tech. Rep. JCTVC-W0031, Feb. 2016.
- [17] T. Lu, F. Pu, P. Yin, T. Chen, and W. Husak, "HDR CE2: CE2.a-2, CE2.c, CE2.d and CE2.e-3," Joint Collaborative Team on Video Coding (JCT-VC), Tech. Rep. JCTVC-W0084, Feb. 2016.
- [18] J. Froehlich, G. M. Su, S. Daly, A. Schilling, and B. Eberhardt, "Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture," in *IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 884–888.
- [19] J. Samuelsson, C. Fogg, A. Norkin, A. Segall, J. Str m, G. J. Sullivan, P. Topiwala, and A. Tourapis, "Conversion and coding practices for HDR/WCG Y'CbCr 4:2:0 video with PQ transfer characteristics," Joint Collaborative Team on Video Coding (JCT-VC), Tech. Rep. JCTVC-Z1017, Feb. 2017.
- [20] S. Yu, C. Jung, and P. Ke, "Adaptive PQ: Adaptive perceptual quantizer for HEVC main 10 profile-based HDR video coding," in *Visual Communications and Image Processing (VCIP)*, Nov. 2016, pp. 1–4.
- [21] *The present state of ultra-high definition television*, Radiocommunication Sector of ITU Std. BT.2246-5, Jul. 2015.
- [22] P. Candry and B. Maximus, "Projection displays: New technologies, challenges, and applications," *Journal of the Society for Information Display*, vol. 23, no. 8, pp. 347–357, 2015.
- [23] W. F. Schreiber, *Fundamentals of Electronic Imaging Systems*. Springer, 1993.
- [24] C. Mota, J. Gomes, and M. I. A. Cavalcante, "Optimal image quantization, perception and the median cut algorithm," *Anais da Academia Brasileira de Ciencias*, vol. 73, pp. 303 – 317, 09 2001.
- [25] E. Velazquez-Armendariz. qt4Image: HDR file viewer. Cornell University. [Online]. Available: <https://bitbucket.org/edgarv/hdritools>
- [26] PYUV: raw video sequence player. DSP Laboratory UniPG. [Online]. Available: http://dsplab.diei.unipg.it/software/pyuv_raw_video_sequence_player
- [27] J. Vanne, M. Viitanen, and T. D. Hmlinen, "Efficient mode decision schemes for HEVC inter prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 9, pp. 1579–1593, Sept 2014.
- [28] B. Bross, W. J. Han, J. R. Ohm, G. J. Sullivan, Y. K. Wang, and T. Wiegand. HM, the reference software for HEVC. Fraunhofer HHI. [Online]. Available: <https://hevc.hhi.fraunhofer.de>
- [29] E. Fran ois, P. Yin, and J. Sole, "Common test conditions for HDR/WCG video coding experiments," International Organization for Standardization, WG11 N15793, Oct. 2015.
- [30] J. Str m, J. Sole, and Y. He, "Report of HDR CE1," Joint Collaborative Team on Video Coding (JCT-VC), JCT-VC m37605, Feb. 2016.
- [31] *Subjective video quality assessment methods for multimedia applications*, Telecommunication Standardization Sector of ITU Std. ITU-T P.910, 2008.
- [32] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 40:1–40:14, Jul. 2011.
- [33] M. Narwaria, M. P. D. Silva, and P. L. Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, no. C, pp. 46–60, Jul. 2015.
- [34] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 39, 2015.
- [35] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Evaluating the performance of existing full-reference quality metrics on High Dynamic Range (HDR) video content," vol. 1, no. 11, p. 788, 2014.
- [36] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb 2006.
- [37] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full dynamic range images," in *Human Vision and Electronic Imaging XIII*, San Jose, USA, Jan. 2008, pp. 6806–10.
- [38] E. Zerman, G. Valenzise, and F. Dufaux, "An extensive performance evaluation of full-reference HDR image quality metrics," *Quality and User Experience*, vol. 2, no. 1, p. 5, 2017.
- [39] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," VCEG-M33, Mar. 2001.
- [40] N. Manish. (2015, Oct.) HDR-VQM software. [Online]. Available: <https://sites.google.com/site/narwariam/hdr-vqm>
- [41] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.
- [42] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.
- [43] HDRTools package. ITU-T and ISO/IEC. [Online]. Available: <https://gitlab.com/standards/HDRTools>
- [44] *Methodology for the subjective assessment of the quality of television pictures*, Radiocommunication Sector of ITU Std. BT.500-13, 2012.
- [45] SIM2 Corporation, "SIM2 HDR47 Monitors," <http://hdr.sim2.it/productslist>, 2016.
- [46] P. Hanhart, M. Rerabek, and T. Ebrahimi, "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies," in *Applications of Digital Image Processing XXXVIII*, 2015, p. 95990G.
- [47] R. Mantiuk. (2017, Mar.) Matlab functions for scaling of pairwise comparison experiment results. [Online]. Available: <https://github.com/mantiuk/pwcmp>