IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 19, 2019, accepted April 13, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911673

Weighted Chroma Downsampling and Luma-Referenced Chroma Upsampling for HDR/WCG Video Coding

QINGTAO FU, CHEOLKON JUNG, (Member, IEEE), GE YANG, AND YI LIU
School of Electronic Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Cheolkon Jung (zhengzk@xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872280, and in part by the International S&T Cooperation Program of China under Grant 2014DFG12780.

ABSTRACT High dynamic range (HDR) and wide color gamut (WCG) video coding adopt chroma downsampling and upsampling for coding efficiency, which causes degradation in visual quality. In this paper, we propose weighted chroma downsampling and luma-referenced chroma upsampling for HDR/WCG video coding. We adopt ICtCp color space for HDR/WCG video coding instead of Y'CbCr color space due to its advantage in the chroma data processing. We perform weighted chroma downsampling prior to encoding based on the difference between neighboring pixels. Moreover, we take advantage of the full resolution luma data to generate weights for chroma upsampling by exploring the correlation between luma and chroma data after decoding. The experimental results show that the proposed method reconstructs high quality HDR videos with better image details and less color distortion, and outperforms HDR Anchor in terms of both objective and subjective evaluations.

INDEX TERMS Chroma downsampling, chroma upsampling, HDR/WCG video coding, ICtCp color space, weighted filter.

I. INTRODUCTION

In recent years, the television (TV) industry is facing a wide range of technology development. As one of the most significant feature of high performance displays, high dynamic range (HDR) and wide color gamut (WCG) technologies have received much attention. They provide consumers with an enhanced visual experience by much brighter highlights and finer details in dark areas than conventional displays [1]. The luminance levels and color gamut in the real world scenes have a much wider range than the traditional low dynamic range (LDR) images. HDR contents overcome the shortcoming of the traditional 8-bit representation by using high bit-depth to better represent the real world scenes [2]. However, it needs a huge amount of memory and bits to store and transmit due to the high bit-depth, which presents greater challenges to video data compression. Thus, efficient video compression technology is required for HDR applications.

Since HDR contents and displays have a wide brightness range, the conventional gamma based Electro-Optic Transfer Function (EOTF) is not applicable any more. Miller et al. proposed a new EOTF named Perceptual Quantizer (PQ) based on perception to meet the requirements, which was designed to quantize the luminance with peak value up to 10000 cd/m^2 [3]. Based on the High Efficiency Video Coding (HEVC) standard and PQ, the Moving Pictures Experts Group (MPEG) released a Call for Evidence (CfE) [4] in 2015 to extend the HEVC standard for HDR/WCG video coding applications. Since then, researchers have made significant achievements in HDR/WCG video coding. To improve the color performance and coding efficiency of HDR/WCG content, Dolby Lab proposed the ICtCp color representation [5] to replace the Non-Constant Luminance (NCL) Y'CbCr color space. ICtCp color space significantly improves color uniformity because it is based on the human visual system (HVS). Moreover, ICtCp color space is closer to the constant luminance criteria than Y'CbCr, which makes it friendly in color processing such as chroma subsampling [5], [6]. In 2016, MPEG released an ICtCp Anchor in the Common Test Conditions (CTC) [7], which used the ICtCp color space for HDR/WCG video coding. Fig. 1 shows the coding/decoding pipeline of ICtCp Anchor defined in [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Tian.

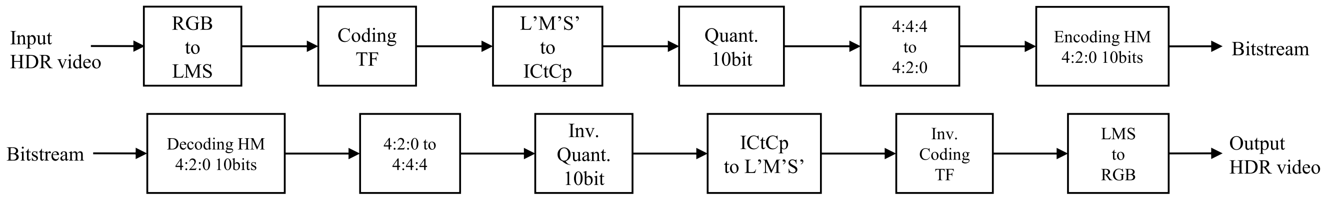


FIGURE 1. **The coding and decoding pipelines of ICtCp Anchor for HDR video compression.**

referred to as Anchor in this work. In the encoding processing chain, the input HDR/WCG video with RGB 4:4:4 format is firstly converted into ICtCp 4:4:4 format, where PQ transfer function (PQ-TF) [3] is utilized for LMS to L'M'S' conversion. Then, a float-to-10bit linear quantization is performed on ICtCp 4:4:4 videos. After that, both chroma components, Ct and Cp, are downsampled from 4:4:4 to 4:2:0, and the 4:2:0 format is used for video compression with the HEVC encoder. In the decoding part, a chain of inverse processes, including 4:2:0 to 4:4:4 chroma upsampling, inverse quantization and inverse color space conversion, is performed on the decoded pictures to reconstruct HDR/WCG videos.

Since HDR video uses a higher bit depth to represent samples, the difference between adjacent pixel values becomes larger, especially in the edge regions. Therefore, compared with traditional LDR videos, HDR videos have higher distortion when performing chroma downsampling, that is, the conventional interpolation filter for downsampling can not obtain good color reproduction for the pixels at edges. In this paper, we propose weighted chroma downsampling for HDR/WCG video coding to improve the color performance in edge areas. We adopt ICtCp color space instead of Y'CbCr due to its advantage in chroma data processing. The proposed chroma downsampling filter is adaptively weighted according to the difference between neighbouring pixels. Moreover, since luma and chroma edges generally have very similar distribution, the luma component with a higher resolution could be utilized to improve the chroma upsampling performance. Thus, we take advantage of the full resolution luma data in 4:2:0 video to generate weights for chroma upsampling. Compared with existing methods, the main contributions of this paper are summarized as follows:

- We introduce adaptive weights to adjust chroma downsampling and upsampling filtering to improve coding efficiency of HDR videos.
- We adopt ICtCp color space for chroma downsampling due to its decorrelation between luma and chroma, and utilize luma information as reference to guide chroma upsampling.
- We explore the correlation between luma and chroma in 4:2:0 HDR video to assign weights for the chroma upsampling filter.

The preliminary results of this paper were presented in [8]. In this extended paper, we provide the details of the proposed HDR/WCG video coding with its motivation and validity. We perform various quantitative measurements to verify the effectiveness and efficiency of the proposed method. Moreover, we conduct subjective evaluation on a commercial HDR TV to demonstrate its performance. Both objective and subjective evaluations verify that the proposed method reconstructs high quality HDR videos with better textures and less color distortion than Anchor.

The remainder of this paper is organized as follows. Section II briefly describes the background and the motivation of the proposed method. Section III presents the details of the proposed method for HDR/WCG video coding. Section IV shows the experimental results and we conclude this work in Section V.

II. BACKGROUND

A. CHROMA DOWNSAMPLING AND UPSAMPLING

Among the processing steps involved in Anchor (Fig. 1), chroma downsampling prior to encoder and upsampling after decoder significantly affects video coding performance because they decrease the fidelity of color reproduction and usually introduce annoying artifacts. The basic idea of using chroma downsampling is that HVS is less sensitive to variations in chrominance than luminance, and therefore it is possible to code chroma components in a lower resolution than luma without visual difference [9]. Previous studies show that chroma downsampling and upsampling improve the rate-distortion (R-D) performance especially at low bit rates [10], [11]. Thoma *et al.* evaluated five different chroma subsampling methods and drew a conclusion that chroma subsampling was suitable for HDR video coding if special preprocessing for handling high contrast edges were taken [12]. However, Boitard *et al.* evaluated the impact of chroma subsampling on the compression efficiency of HDR contents and concluded that chroma subsampling did not improve compression efficiency for them [13]. Despite the different points of view, chroma downsampling and upsampling are still important to video coding efficiency. Dong and Ye proposed to find the optimal downsampling ratio by balancing the downsampling and encoding distortion under a target bit rate, achieving the overall optimal R-D performance [14]. Wang *et al.* proposed luma aware chroma downsampling and upsampling for screen content videos by using the major color of image and index map representation [15]. Yu *et al.* proposed a chroma upsampling scheme for Y'CbCr 420 videos by using the neighboring chroma values to give weights [16]. Korhonen proposed an adaptive upsampling method for chroma data that used luma information to assist chroma upsampling [9]. They exploited mean squared differences of neighbouring luma block pixels to assign weights to chroma pixels. Ström and Wennersten proposed a pre-processing method for HDR videos to remove luminance artifacts occurred in saturated colors after compression when PQ NCL Y'CbCr 4:2:0 was used [17].

B. ICTCP COLOR SPACE

ICtCp color space is designed to replace the NCL Y'CbCr color space for HDR/WCG video coding. It has three main advantages: (1) ICtCp meets the constant luminance criteria better than Y'CbCr; (2) ICtCp maintains better hue linearity than Y'CbCr, i.e. less crosstalk between hue and saturation; (3) ICtCp remarkably improves the just noticeable difference (JND) uniformity, and thus achieves better performance in baseband quantization than Y'CbCr [5]. Because of these advantages, we adopt ICtCp color representation instead of Y'CbCr to perform chroma processing in this work. In the coding part of Anchor, color space conversion from RGB to ICtCp is realized through three steps as follows:

1) Compute LMS response from RGB (ITU-R BT.2020):

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \frac{1}{4096}\begin{pmatrix} 1688 & 2146 & 262 \\ 683 & 2951 & 462 \\ 99 & 309 & 3688 \end{pmatrix}\begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

2) Apply PQ transfer function [3] to LMS:

$$\begin{pmatrix} L' \\ M' \\ S' \end{pmatrix} = EOTF_{PQ}^{-1}\begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{2}$$

with

$$EOTF_{PQ}^{-1}(F) = (\frac{c_1 + c_2 * Y^{m_1}}{1 + c_3 * Y^{m_1}})^{m_2} \tag{3}$$

where $Y = F/10000$, $m_1 = 0.1593017578125$, $m_2 = 78.84375$, $c_1 = 0.8359375$, $c_2 = 18.8515625$, and $c_3 = 18.6875$.

3) Calculate ICtCp from L'M'S':

$$\begin{pmatrix} I \\ Ct \\ Cp \end{pmatrix} = \frac{1}{4096}\begin{pmatrix} 2048 & 2048 & 0 \\ 6610 & -13613 & 7003 \\ 17933 & -17390 & -543 \end{pmatrix}\begin{pmatrix} L' \\ M' \\ S' \end{pmatrix} \tag{4}$$

In the decoding part, the inverse color space conversion from ICtCp to RGB is performed. ICtCp is first converted to L'M'S', then L'M'S' is inverse mapped to LMS using the inverse PQ-TF, and finally LMS is converted into RGB.

C. MOTIVATION

To achieve better highlights in bright regions and better details in dark areas, HDR/WCG videos use 10bit, 12bit or even higher bit depth to represent image data. Due to the higher bit depth used, the sample range in HDR videos is much wider than LDR, and the difference between adjacent pixel values becomes larger especially in the edge areas. Consequently, chroma downsampling filters with fixed coefficients would introduce large color distortion at edges. To reduce the distortion and preserve more textures and details, we propose adaptive weighted chroma downsampling for HDR/WCG video coding. We adaptively adjust the coefficient values of the chroma downsampling filter by assigning weights according to the difference between neighboring pixels. A larger weight is assigned to a closer neighboring pixel, and vice versa. Thus, the color performance in edge areas would be improved. In addition, we propose luma-referenced weighted chroma upsampling for HDR/WCG video coding in this work. Luma and chroma components usually have very similar edge information. Fig. 2 shows a small region of the luma component (I) and its corresponding chroma ones (Ct and Cp). It can be observed that the area with edges in luma shows similar edges in chroma, which implies that there is a strong correlation in edge areas between luma and chroma. Since luma data are not downsampled when coding 4:2:0 format video, the full resolution luma data could be useful for improving chroma upsampling performance. Therefore, we take advantage of the full resolution luma data to generate weights for upsampling filter based on the correlation between luma and chroma pixels, i.e. luma-referenced chroma upsampling. The proposed chroma downsampling and upsampling method is simple yet effective so that it does not require much additional computational complexity over Anchor.

FIGURE 2. Luma and chroma components. Left: I component. Middle: Ct component. Right: Cp component.

III. PROPOSED METHOD

A. WEIGHTED CHROMA DOWNSAMPLING

Anchor performs encoding using ICtCp 4:2:0 format videos. Chroma downsampling from 4:4:4 to 4:2:0 is realized by a 2-D filter, which first performs 4:4:4 to 4:2:2 filtering in the horizontal direction and then conducts 4:2:2 to 4:2:0 filtering in the vertical direction. The same 3-tap filter is applied to both the horizontal and vertical downsampling, which can be formulated as:

$$f = c_{-1} \cdot x_{-1} + c_0 \cdot x_0 + c_1 \cdot x_1 \tag{5}$$

where x is the chroma data, c is the filter coefficient values, and f is the downsampled chroma pixel value. The subscript $\{-1, 0, 1\}$ refers to the left, the center (the current), and the right pixels, respectively. Three fixed coefficient values $[c_{-1}, c_0, c_1] = [1, 6, 1]/8$ are used for both horizontal and vertical filtering. It should be noted that this downsampling filter only utilizes the distance between adjacent pixels for chroma downsampling. However, we have found that the difference information between their values is very useful to preserve textures and details in HDR/WCG video reconstruction. Thus, based on the difference between the adjacent pixels, we introduce two weights w_1^{down} and w_2^{down} to the downsampling filter to adjust the filter coefficients adaptively. To be specific, we assign a larger weight to the adjacent pixel with a closer pixel value, and vice versa. w_1^{down} and w_2^{down} are formulated as follows:

$$w_1^{down} = 1 + \theta^{down} \cdot \frac{\Delta x}{x_{max} - x_{min}} \tag{6}$$

$$w_2^{down} = 1 - \theta^{down} \cdot \frac{\Delta x}{x_{max} - x_{min}} \tag{7}$$

with

$$\Delta x = |x_1 - x_0| - |x_{-1} - x_0| \tag{8}$$

where θ^{down} is a parameter to adjust the weights, x_{max} and x_{min} are the maximum and minimum pixel values of chroma components.

The coefficient values of the proposed weighted chroma downsampling filter are as follows:

$$[\hat{c}_{-1}, \hat{c}_0, \hat{c}_1] = [w_1^{down}, 6, w_2^{down}]/8 \tag{9}$$

In (9), we obtain w_1^{down} and w_2^{down} based on the difference between the center and the left/right pixels which adaptively adjust the filter coefficients according to the video content.

B. LUMA-REFERENCED CHROMA UPSAMPLING

Chroma upsampling from ICtCp 4:2:0 to 4:4:4 is performed in the decoding part. Anchor first performs 4:2:0 to 4:2:2 filter in the vertical direction, and then performs 4:2:2 to 4:4:4 filtering in the horizontal direction. A 4-tap filter with the fixed coefficient values is used for both the vertical and horizontal chroma upsampling. The chroma upsampling filter in Anchor is formulated as follows:

$$x^* = c_{-2} \cdot x_{-2} + c_{-1} \cdot x_{-1} + c_0 \cdot x_0 + c_1 \cdot x_1 \tag{10}$$

where x is the chroma data and c the filter coefficient values. x^* denotes the pixel to be determined by upsampling, and the subscripts refer to the positions of adjacent pixels in chroma components as illustrated in Fig. 3. The coefficient values are defined as follows:

$$[c_{-2}, c_{-1}, c_0, c_1] = [-4, 36, 36, -4]/64 \tag{11}$$

For 4:2:0 videos, chroma downsampling causes chroma information loss, while luma keeps the full resolution as shown in Fig. 3. Since there is a strong correlation between chroma and luma data, we take advantage of the correlation between luma and chroma data to improve chroma upsampling performance, especially for edge areas. We propose luma-referenced chroma upsampling to assign different weights for the upsampling filter by exploring the full resolution luma values and the correlation with chroma data. More specifically, we first take four consecutive pixels in the chroma 4:2:0 data $\{x_{-2}, x_{-1}, x_0, x_1\}$ and four associated luma pixels at the corresponding positions in the luma data $\{y_{-2}, y_{-1}, y_0, y_1\}$ as illustrated in Fig. 3. Then, we evaluate the correlation between these chroma and luma pixels. The correlation coefficient r between luma and chroma pixels is calculated by:

$$r^2 = \frac{(\sum_{i=-2}^{1}(x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=-2}^{1}(x_i - \bar{x})\sum_{i=-2}^{1}(y_i - \bar{y})} \tag{12}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of four pixels of x_i and y_i, respectively. A threshold TH determines whether the luma pixels are highly correlated with chroma ones. When $r^2 > TH$, two weights w_1^{up} and w_2^{up} are introduced into the chroma upsampling filter as follows:

$$w_1^{up} = 1 + \theta^{up} \cdot \frac{\Delta y}{y_{max} - y_{min}} \tag{13}$$

$$w_2^{up} = 1 - \theta^{up} \cdot \frac{\Delta y}{y_{max} - y_{min}} \tag{14}$$

with

$$\Delta y = |y_0 - y^*| - |y_{-1} - y^*| \tag{15}$$

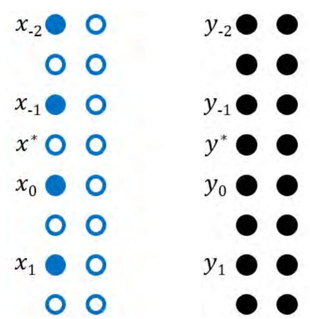


FIGURE 3. Conceptual diagram for chroma upsampling. Left: Chroma pixels in 4:2:0 video. The solid dots represent the samples before upsampling, while the empty dots denote the chroma samples to be determined by upsampling. Right: Luma pixels in 4:2:0 video.

where y^* is the pixel between y_{-1} and y_0 as shown in Fig. 3, and θ^{up} is a parameter to adjust the weights for chroma pixel values. The weighted coefficients for the proposed upsampling filter are obtained as follows:

$$[\hat{c}_{-2}, \hat{c}_{-1}, \hat{c}_0, \hat{c}_1] = [-4, 36w_1^{up}, 36w_2^{up}, -4]/64 \tag{16}$$

We perform weighted chroma downsampling and luma-referenced chroma upsampling on the encoding and decoding sides for HDR/WCG video coding, respectively. In areas with sharp edges, the proposed method improves chroma performance by assigning large weights. In flat areas where there is no big difference between pixel values, the proposed method hardly changes the original filter coefficients, so we maintain the original sampling results there. Moreover, due to the low computational complexity, the proposed luma-referenced chroma upsampling can be easily implemented in HDR TVs to improve the visual quality of HDR content.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method to verify its superiority for HDR/WCG video coding. The eight HDR sequences in Table 1 are used for tests, and Fig. 4 shows their thumbnails. We perform extensive experiments according to the test conditions defined in CTC [7], using BT.2020 color primaries, ICtCp color space and HEVC Main 10 Profile with Random Access configuration for encoding. For each sequence, we generate the bitstreams at four quantization parameter (QP) levels. Considering the coding efficiency and computational complexity, we only perform the proposed weighted downsampling on the horizontal direction, i.e. from ICtCp 4:4:4 to 4:2:2, using the filter coefficients in (9), and the luma referenced upsampling on the vertical direction, i.e. from ICtCp 4:2:0 to 4:2:2, using the filter coefficient values in (16). For the tests, we set θ^{down} in (6) and (7) to 2, θ^{up} in (13) and (14) to 4, the threshold TH for upsampling to 0.36. For all the test videos, the bit-depth is 10, and we set $x_{min} = y_{min} = 0$ and $x_{max} = y_{max} = 1023$.

TABLE 1. Test sequences for HDR/WCG video coding.

Sequence	fps	Resolution	Gamut	Frames
FireEater	25	1920x1080	BT2020	200
Market	50	1920x1080	BT2020	400
BalloonFestival	24	1920x1080	BT2020	240
GarageExit	24	1920x1080	P3D65	288
Sunrise	25	1920x1080	BT2020	200
ShowGirl	25	1920x1080	P3D65	339
Hurdles	100	1920x1080	BT2020	500
Starting	100	1920x1080	BT2020	500

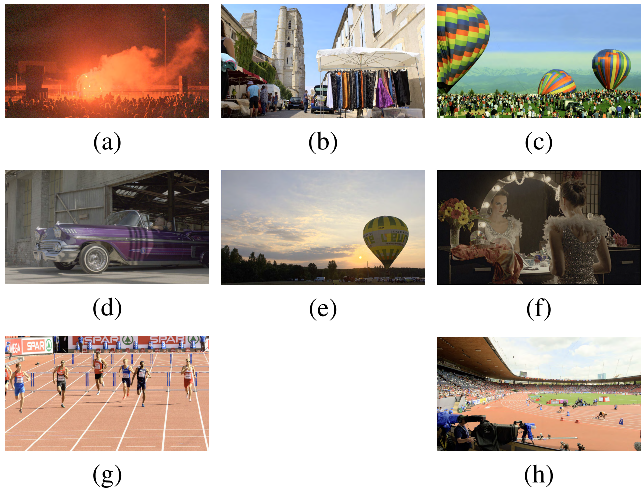


FIGURE 4. HDR video sequences for experiments. (a) *FireEater*. (b) *Market*. (c) *BalloonFestival*. (d) *GarageExit*. (e) *SunRise*. (f) *ShowGirl*. (g) *Hurdles*. (h) *Starting*.

A. SUBJECTIVE EVALUATION

Figs. 5 and 6 show visual comparisons of the reconstruction results by Anchor, the proposed method, and the original HDR images. To show the difference, we take a part of each image and zoom in it for visual comparisons. As shown in Fig. 5, the reconstructed image of *FireEater* by Anchor contains obvious blocking artifacts while the proposed method removes them successfully. For *Market* and *Sunrise*, it can be observed that the proposed method preserves more details than Anchor. For *GarageExit*, the proposed method reconstructs more natural-looking results than Anchor. As shown in Fig. 6, the reconstructed frames of *ShowGirl* and *BalloonFestival* by Anchor contain color distortions along boundaries compared with the results of the proposed method. For *Hurdles*, the proposed method preserves better details than Anchor. The result for *Starting* also shows that the proposed method reconstructs HDR videos with better visual quality and better details. The experimental results show that the proposed method effectively improves the visual quality of the reconstruction results than Anchor in terms of both detail preservation and color reproduction. That is because we employ the difference information of samples and take advantage of the correlation between luma and chroma components.

Moreover, we conduct subjective tests on a commercial HDR TV with a peak brightness of 1500 cd/m^2 [18]. The luminance higher than 1500 cd/m^2 is clipped in this TV. Since most content in the test HDR image is less than 1000 nits, the brightness of our HDR TV is sufficiently enough to display most content of the HDR videos. We set the viewing conditions according to the ITU-R BT.500-13 Recommendation [19]. Total eighteen observers who have passed the normal visual acuity and normal color vision tests participate in the subjective evaluation. We decode the bitstream and then convert the reconstructed HDR frames to YUV444 format 10-bit videos using PQ-TF, and generate the video format required by the HDR TV for display. For each sequence, we perform subjective comparisons at two quality levels: one is high quality video (high bitrate) and the other one is low quality video (low bitrate). The high quality video is obtained at a small QP when encoding, while the low quality video is obtained at a large QP when encoding. Each test video contains 200 frames. Following the instructions in [19], two sequences, *FireEater* and *Market*, are used as training sequences to demonstrate the range and the type of video impairments to observers. Two sequences, *GarageExit* and *Hurdles* are used to stabilize their opinion at the beginning of the test session. The other four sequences, *BalloonFestival*, *Sunrise*, *Starting*, and *Showgirl*, are used for tests. In the test session, referring to [20], the pair of videos generated by Anchor and the proposed method are randomly arranged side by side to be simultaneously displayed to observers. Each video pair is played twice at the rate of 25 fps. The observers are asked to choose which video in the pair ('left' or 'right') has a better overall quality based on their impression. Fig. 7 shows the subjective evaluation results for the four test sequences in terms of preference. The results show that the proposed method gets more preference than Anchor, which indicates that the proposed method produces HDR videos with higher visual quality than Anchor. It can be inferred from the evaluations that the proposed method is effective in improving visual quality of HDR videos.

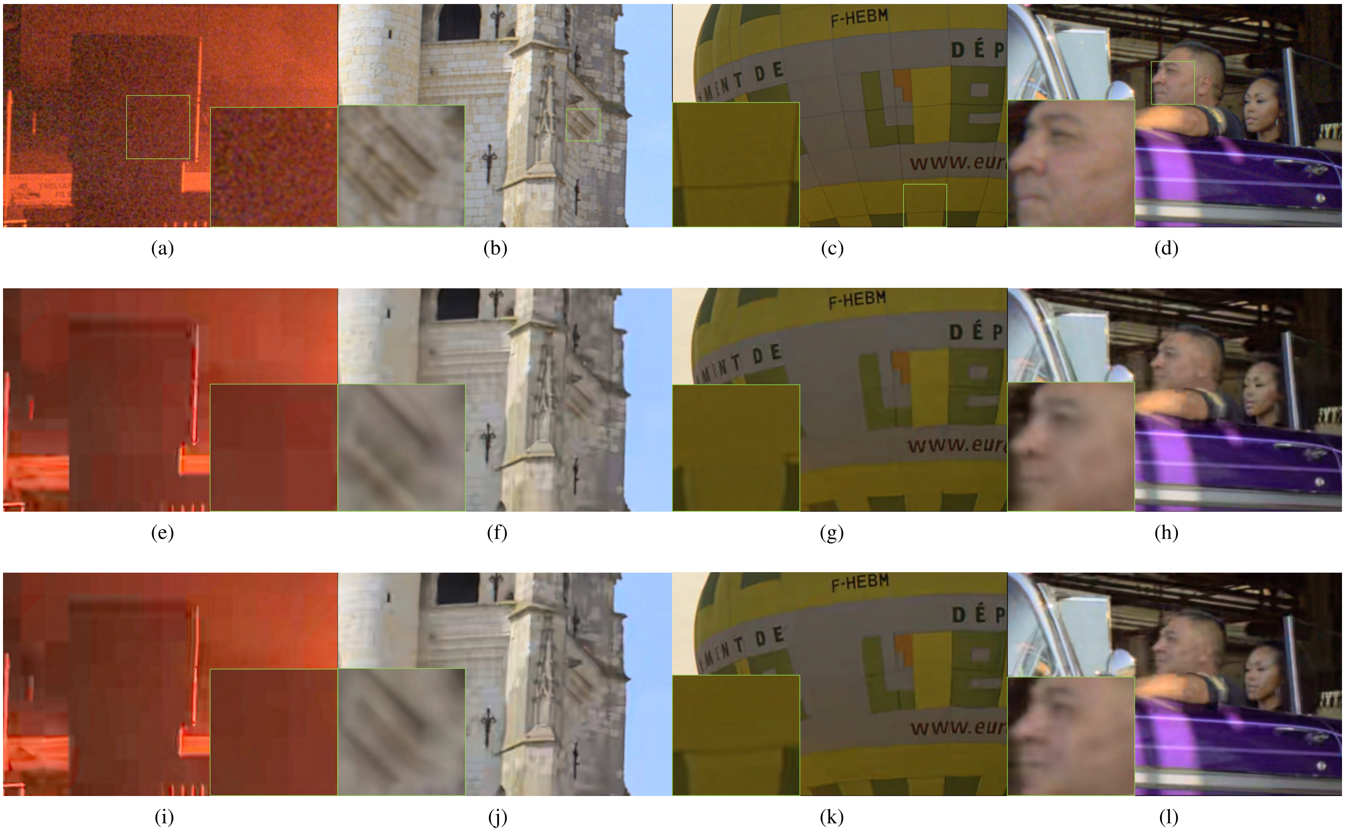


FIGURE 5. Visual comparison of HDR reconstruction results. From left to right: ***FireEater***, ***Market***, ***Sunrise***, and ***GarageExit***. (a) Original. (b) Original. (c) Original. (d) Original. (e) Anchor. (f) Anchor. (g) Anchor. (h) Anchor. (i) Proposed. (j) Proposed. (k) Proposed. (l) Proposed.

TABLE 2. BD-rate performance (%) of the proposed chroma down and up sampling compared with Anchor.

Sequence	X	Y	Z	XYZ	tO-XYZ	DE100	L100	Y	U	V	YUV
FireEater	0.0	0.1	-0.4	-0.1	0.1	0.1	0.0	0.2	-22.8	0.4	0.2
Market	0.1	0.1	-0.1	0.0	0.0	0.3	0.1	0.0	-5.0	-13.9	-0.1
Sunrise	0.1	0.0	0.0	0.0	0.2	0.4	0.0	0.0	0.6	-3.5	-0.1
GarageExit	0.0	0.0	-0.4	-0.2	-0.2	-0.2	0.1	0.0	-3.9	-1.4	-0.2
ShowGirl	0.2	0.0	-0.4	-0.1	-1.0	-0.2	0.0	-0.1	-12.0	-2.6	-0.4
BalloonFestival	0.2	0.4	-0.6	-0.1	-0.4	-0.7	0.4	0.4	-14.5	-21.7	-0.3
Hurdles	0.1	0.2	-0.5	-0.1	-0.2	-0.2	0.2	0.1	-11.6	-11.7	-0.4
Starting	0.2	0.3	-0.5	0.0	0.0	-0.3	0.4	0.2	-10.1	-12.6	-0.3
Overall	0.1	0.1	-0.4	-0.1	-0.2	-0.1	0.1	0.1	-9.9	-8.4	-0.2

B. OBJECTIVE EVALUATION

For objective evaluation, we first calculate the overall BD-rate performance of the proposed chroma downsampling and upsampling method compared with Anchor. In the tests, we evaluate the BD-rate performance in terms of tPSNR-X, tPSNR-Y, tPSNR-Z, tPSNR-XYZ, and tOSNR-XYZ, DE100, PSNRL100, tPSNR-Y, tPSNR-U, tPSNR-V, and tPSNR-YUV. tPSNR measures distortions between the original and reconstructed images after PQ-TF. Table 2 shows the BD-rate gains (%) of the proposed method over Anchor. The results show that the proposed method achieves better BD-rate performance than Anchor in terms of tPSNR-U and tPSNR-V by 9.9% and 8.4% for U and V components, respectively, with a slight decrease in tPSNR-Y of 0.1%. For *BalloonFestival*, *Hurdles* and *Starting*, the BD-rate gains in tPSNR-U and tPSNR-V are more than 10%, sometimes, even 20%. Note that both tPSNR-U and tPSNR-V are related to chroma components which indicates that the proposed method improves chroma performance for HDR/WCG video coding. The overall BD-rate performance in terms of tPSNR-XYZ and tOSNR-XYZ is also improved by 0.1% and 0.2%, respectively. The improvement is not significant because the proposed method focuses on chroma performance and mainly processes pixels with large neighbouring difference. For pixels that are close to neighbouring ones, the proposed method provides a very small change in

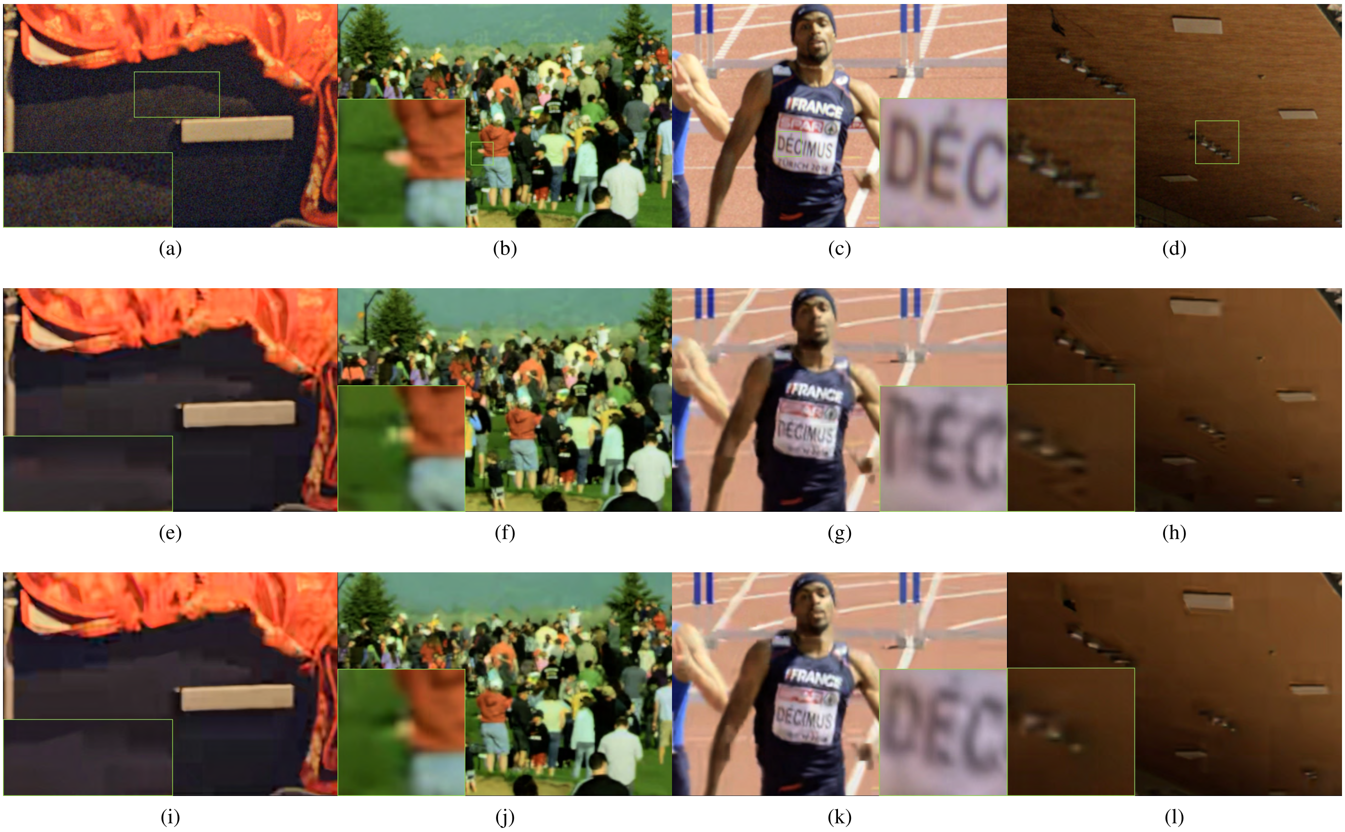


FIGURE 6. Visual comparison of HDR reconstruction results. From left to right: *ShowGirl*, *BalloonFestival*, *Hurdles*, and *Starting*. (a) Original. (b) Original. (c) Original. (d) Original. (e) Anchor. (f) Anchor. (g) Anchor. (h) Anchor. (i) Proposed. (j) Proposed. (k) Proposed. (l) Proposed.

TABLE 3. BD-rate performance (%) of the proposed method, [6], and [21] compared with NCL Y'CbCr Anchor.

	Proposed				[6]				[21]			
Sequence	XYZ	tO-XYZ	DE100	L100	XYZ	tO-XYZ	DE100	L100	XYZ	tO-XYZ	DE100	L100
FireEater	2.7	-3.6	-22.1	-7.1	6.4	-0.7	-21.6	-7.2	2.1	0.5	-0.9	-0.1
Market	-1.2	-1.6	-14.2	-0.2	-1.0	-1.3	-13.4	0.0	-3.5	-5.5	-8.3	-7.2
SunRise	-1.9	-4.5	-41.0	-0.9	-1.7	-4.4	-41.7	-0.9	-3.0	-6.3	-8.0	-6.7
GarageExit	-1.4	-1.1	-1.8	-1.4	-1.8	-1.6	-2.4	-2.0	0.6	-1.3	-2.4	-1.3
BalloonFestival	0.7	1.1	2.5	0.0	1.1	1.6	4.5	-0.2	-0.9	-7.5	-11.5	-6.6
Hurdles	-1.7	-2.8	-19.4	0.9	-1.6	-2.7	-18.3	0.9	-2.8	-2.8	-3.8	-3.2
Starting	-3.7	-5.0	-15.7	1.4	-3.6	-4.8	-14.9	1.6	-2.3	-4.1	-6.2	-4.6
Overall	-0.9	-2.5	-15.9	-1.0	-0.3	-2.0	-15.4	-1.1	-1.4	-3.9	-5.9	-4.2

weights so that it does not change the results significantly. Moreover, we compare the objective performance of the proposed method with those of [6] and [21]. As a basis for this work, [6] introduced the ICtCp color space and evaluated the compression performance of ICtCp relative to Y'CbCr. [21] is one of our previous work that adopts irregularity concealment effect for HEVC encoder optimization to improve HDR video coding efficiency. Table 3 shows BD-rate performance of them compared with NCL Y'CbCr Anchor. The results indicate that the proposed method achieves a performance improvement in coding efficiency over [6]. Besides, [21] achieves better coding performance than the proposed method in terms of tPSNR-XYZ, tOSNR-XYZ, and PSNRL100, while the proposed method shows much better DE100 performance.

To clearly show the performance of the proposed chroma upsampling method, we only perform the proposed luma-referenced chroma upsampling in Anchor and compare the compression performance. Table 4 lists the BD-rate evaluation results. It can be observed that the proposed chroma upsampling method improves the BD-rate performance over Anchor. Especially for tPSNR-U and tPSNR-V, the proposed method achieves 1.8% and 4.0% improvements, respectively. Besides, we perform the proposed luma-referenced chroma upsampling method on test videos without HEVC compression, i.e. without the blocks of "Encoding HM" and

TABLE 4. BD-rate performance (%) of the proposed chroma upsampling compared with Anchor.

Sequence	X	Y	Z	XYZ	tO-XYZ	DE100	L100	Y	U	V	YUV
FireEater	0.0	0.0	-0.4	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.6	-0.1
Market	0.0	0.0	-0.1	0.0	-0.1	-0.2	0.0	0.0	-1.4	-0.8	0.0
SunRise	0.0	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.0	0.3	-0.2	0.0
GarageExit	-0.1	0.0	-0.4	-0.2	-0.2	-0.5	0.0	0.0	-3.0	-1.2	-0.2
ShowGirl	0.0	0.0	-0.5	-0.2	-0.2	-0.4	0.0	0.0	-0.6	-1.4	-0.1
BalloonFestival	-0.1	0.0	-0.5	-0.2	-0.3	-1.1	0.0	0.0	-3.3	-7.4	-0.2
Hurdles	-0.2	0.0	-0.4	-0.2	-0.3	-2.3	0.1	0.0	-3.3	-10.5	-0.4
Starting	-0.2	0.0	-0.4	-0.2	-0.3	-1.8	0.0	-0.1	-3.0	-9.8	-0.4
Overall	-0.1	0.0	-0.4	-0.2	-0.2	-0.8	-0.0	-0.0	-1.8	-4.0	-0.2

TABLE 5. PSNR gains (dB) on uncompressed videos using luma-referenced chroma upsampling.

Sequence	X	Y	Z	XYZ	tO-XYZ	DE100	L100	Y	U	V	YUV
FireEater	0.10	0.00	0.09	0.09	0.04	0.03	0.00	0.12	0.02	0.14	0.13
Market	0.00	0.01	-0.01	-0.01	0.03	0.01	0.00	0.08	0.02	0.01	0.02
SunRise	-0.01	0.00	0.00	0.00	0.02	0.03	0.01	0.00	-0.07	-0.01	-0.02
GarageExit	0.12	0.06	0.08	0.09	0.05	0.06	0.01	0.20	0.19	0.09	0.15
ShowGirl	0.05	0.03	0.08	0.08	0.05	0.08	0.02	0.03	0.05	0.02	0.05
BalloonFestival	0.13	0.03	0.10	0.10	0.10	0.07	0.02	0.19	0.15	0.18	0.17
Hurdles	0.34	0.04	0.10	0.13	0.09	0.01	0.00	0.26	0.14	0.40	0.33
Starting	0.31	0.08	0.07	0.10	0.08	0.02	0.02	0.24	0.12	0.34	0.29
Overall	0.13	0.03	0.07	0.07	0.06	0.04	0.01	0.14	0.08	0.15	0.14

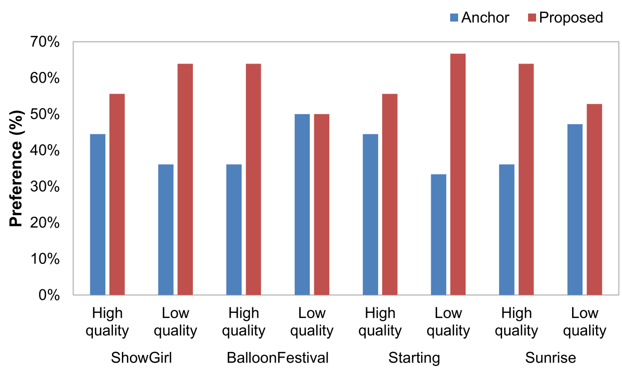


FIGURE 7. Subjective evaluation results of four test sequences at two quality levels.

"Decoding HM" in Fig. 1, then evaluate the tPSNR performance on the reconstructed videos. The tPSNR gains (dB) on uncompressed videos are listed in Table 5. It shows that the proposed luma-referenced chroma upsampling method achieves tPSNR performance improvements over Anchor, especially for *BalloonFestival*, *Hurdles* and *Starting*. This is because these three videos have a lot of sharp edges with large inter-pixel differences.

We also evaluate and compare the HDR-VQM [22] and HDR-VDP2 [23] performance of the proposed method with Anchor. We provide HDR-VQM curves in Fig. 8 and HDR-VDP2 curves in Fig. 9. A lower HDR-VQM value or a larger HDR-VDP2 value indicates a higher visual quality. As shown in these figures, the proposed method achieves similar HDR-VQM and HDR-VDP curves to Anchor. This is because the proposed method mainly considers the chroma

TABLE 6. BD-rate performance (%) in terms of MS-SSIM compared with Anchor.

Sequence	R	G	B	RGB
FireEater	1.4	3.3	-1.3	1.6
Market	0.3	0.4	-0.2	0.1
Sunrise	-0.5	-0.5	-0.4	-0.4
GarageExit	-0.4	0.1	-0.9	-0.4
ShowGirl	-0.5	-0.1	0.4	-0.1
BalloonFestival	-2.2	0.8	-5.8	-2.7
Hurdles	-0.4	0.4	-0.7	-0.2
Starting	-0.1	0.4	-0.4	0.0
Overall	-0.3	0.6	-1.2	-0.3

pixels with large difference to improve the chroma performance. Furthermore, we evaluate Multiscale Structural Similarity (MS-SSIM) performance of the proposed chroma downsampling and upsampling method. MS-SSIM [24] metric is an extension of SSIM for image quality assessment which has been integrated in HDRTools software. We evaluate MS-SSIM performance of R, G, and B components using HDRTools, and calculate the overall BD-rate performance in terms of MS-SSIM. The BD-rate performance is listed in Table 6. The results show that the proposed method achieves 0.3% overall BD-rate improvement, which indicates that the proposed method obtains better perceptual quality than Anchor. More specifically, we achieve 2.7% BD-rate improvement on *BalloonFestival*. This is because *BalloonFestival* contains a lot of edges and areas with strong color changes.

In addition, we measure time costs for the forward conversion (prior to encoder) and inverse conversion (after decoder)

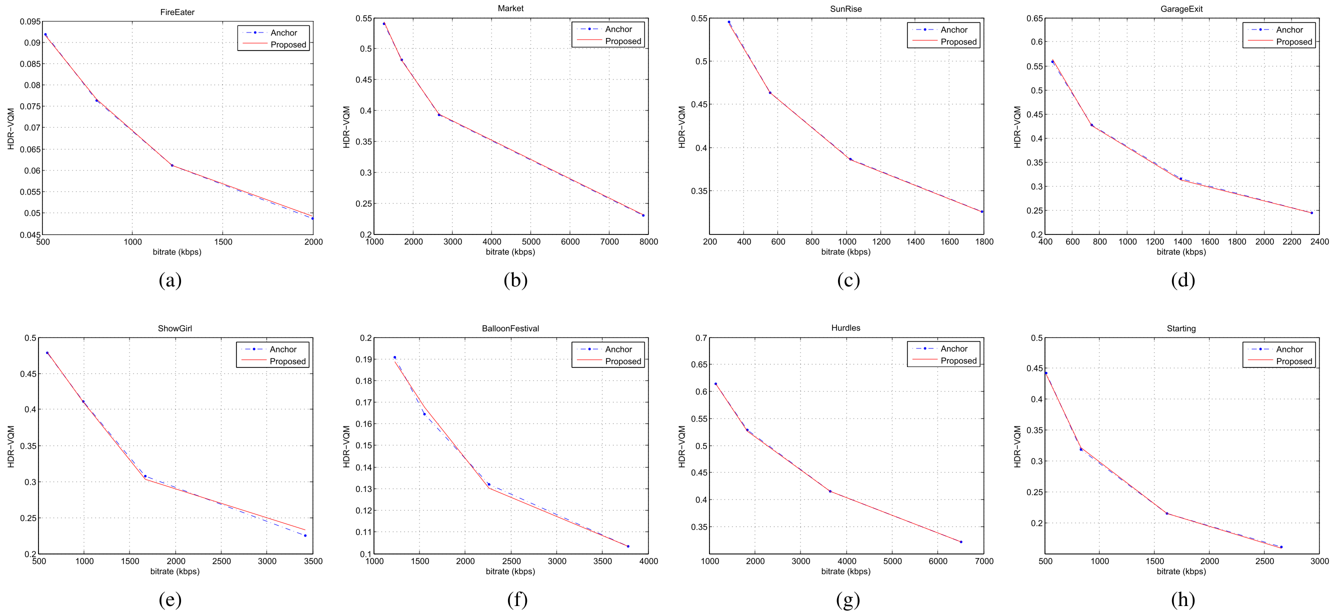


FIGURE 8. **HDR-VQM performance comparison using rate distortion curves. A lower HDR-VQM value indicates a better quality. (a) *FireEater*. (b) *Market*. (c) *SunRise*. (d) *GarageExit*. (e) *ShowGirl*. (f) *BalloonFestival*. (g) *Hurdles*. (h) *Starting*.**

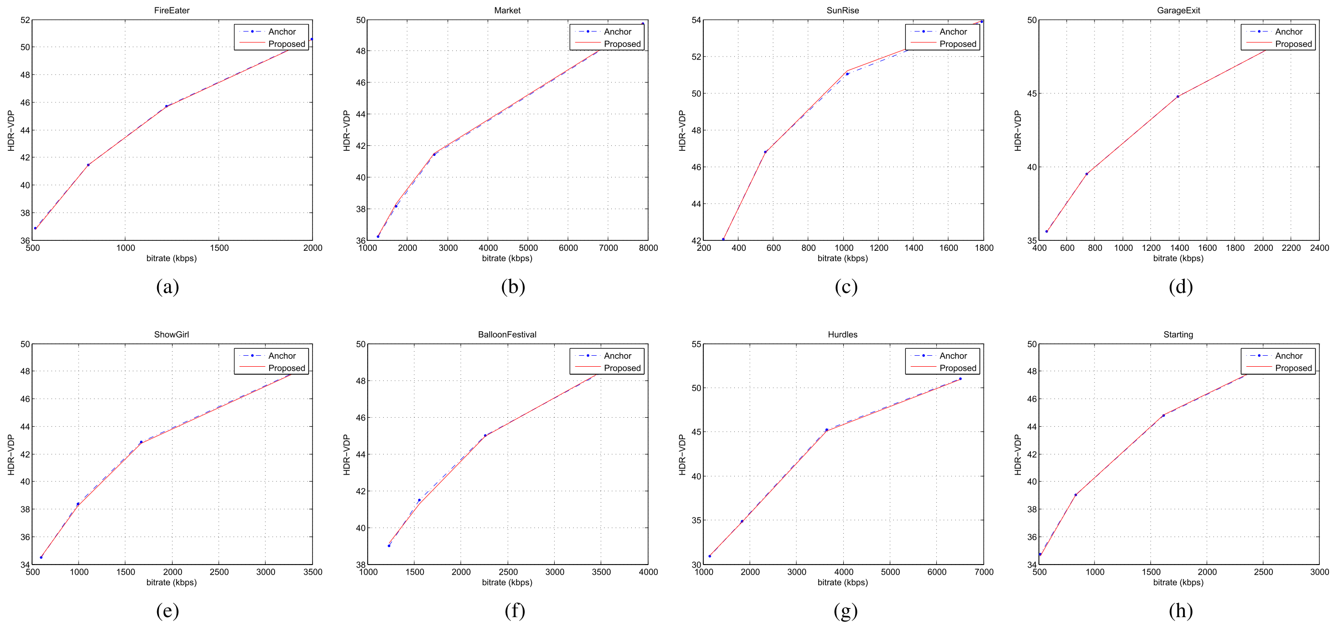


FIGURE 9. **HDR-VDP2 performance comparison using rate distortion curves. A larger HDR-VDP2 value indicates a better quality. (a) *FireEater*. (b) *Market*. (c) *SunRise*. (d) *GarageExit*. (e) *ShowGirl*. (f) *BalloonFestival*. (g) *Hurdles*. (h) *Starting*.**

to evaluate computational efficiency. This is because the extra computational complexity of the proposed method is mainly from the chroma downsampling and upsampling processes. For the forward conversion, we test 100 HDR frames for each test sequence, i.e. total 800 frames for eight sequences. For the inverse conversion, a total of 3200 frames are tested for eight sequences including four QPs. We perform the tests on a workstation with dual Intel Xeon CPU E5-2640 (2.4 GHz) and 32GB RAM. The total time consumption of both Anchor and the proposed method is listed in Table 7.

TABLE 7. **Time consumption for the forward and inverse conversions.**

Method	Forward Time (s)	Inverse Time (s)
Anchor	650.102	2410.513
Proposed	682.591	2481.240
Overall (%)	105.00%	102.93%

The proposed method needs extra time costs of 5.00% and 2.93% over Anchor for the forward and inverse conversions, respectively.

V. CONCLUSION

In this paper, we have proposed weighted chroma downsampling and luma-referenced chroma upsampling for HDR/WCG video coding. We have adopted ICtCp color space to prevent color artifacts since ICtCp has better decorrelation between luma and chroma. We have used the difference between adjacent pixels for chroma downsampling to preserve more image textures and details. Moreover, we have utilized luma data for chroma upsampling by exploring the correlation between luma and chroma to reduce the distortions introduced by chroma downsampling. The objective evaluation results demonstrate that the proposed method improves coding efficiency for HDR videos over Anchor. Visual comparison results indicate that the proposed method reconstructs higher quality HDR videos with less color distortions and better image details than Anchor. The subjective tests on a commercial HDR TV also verify that the proposed method produces high quality HDR videos.

ACKNOWLEDGMENT

The authors would like to thank Novatek Microelectronics Corp. in Xi'an for providing the HDR TV for the subjective evaluations.

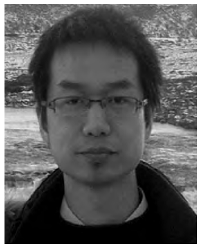
REFERENCES

[1] *Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange*, Standard ITU-R BT.2100, Jun. 2017.
[2] 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 Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 63–75, Jan. 2016.
[3] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imag. J.*, vol. 122, no. 4, pp. 52–59, Oct. 2012.
[4] A. Luthra, E. Francois, and W. W. Husak, *Call for Evidence (CFE) for HDR and WCG Video Coding*, Standard ISO/IEC JTC1/SC29/WG11 (MPEG), Geneva, Switzerland, Feb. 2015.
[5] Dolby. *ICtCp White Paper*. Accessed: Jun. 19, 2017. [Online]. Available: https://www.dolby.com/us/en/technologies/dolby-vision/ICtCp-white-paper.pdf
[6] T. Lu, F. Pu, P. Yin, T. Chen, and W. Husak, "Conversion and coding practices for HDR/WCG ICTCP 4:2:0 video," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, pp. 13–22.
[7] E. François, J. Sole, J. Strom, and P. Yin, *Common Test Conditions for HDR/WCG Video Coding Experiments*, Standard ISO/IEC JTC1/SC29/WG11 JCTVC-X1020, Geneva, Switzerland, May 2016.
[8] Q. Fu, C. Jung, and G. Yang, "Weighted chroma downsampling and luma-referenced chroma upsampling for HDR/WCG video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2565–2569.
[9] J. Korhonen, "Improving image fidelity by luma-assisted chroma subsampling," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.
[10] W. Lin and L. Dong, "Adaptive downsampling to improve image compression at low bit rates," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2513–2521, Sep. 2006.
[11] V.-A. Nguyen, Y.-P. Tan, and W. Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2008, pp. 1624–1627.
[12] H. Thoma, M. de-Frutos-López, and J. Auer, "Chroma subsampling for HDR video with improved subjective quality," in *Proc. Picture Coding Symp. (PCS)*, Dec. 2013, pp. 345–348.
[13] R. Boitard, M. T. Pourazad, and P. Nasiopoulos, "Evaluation of chroma subsampling for high dynamic range video compression," in *Proc. IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, Dec. 2016, pp. 696–699.
[14] J. Dong and Y. Ye, "Adaptive downsampling for high-definition video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 480–488, Mar. 2014.
[15] S. Wang, K. Gu, S. Ma, and W. Gao, "Joint chroma downsampling and upsampling for screen content image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1595–1609, Sep. 2016.
[16] Y.-C. Yu *et al.*, "Chroma upsampling for YCbCr 420 videos," in *Proc. IEEE Int. Conf. Consum. Electron.*, Taipei, Taiwan, Jun. 2017, pp. 163–164.
[17] J. Ström and P. Wennersten, "Chroma adjustment for HDR video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 11–15.
[18] *Sony XBR-55X930E–Specifications*. Accessed: Oct. 15, 2018. [Online]. Available: https://www.displayspecifications.com/en/model/6e73a09
[19] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Standard ITU-R BT.500-13, Jan. 2012.
[20] Y. Liu, W. Hamidouche, O. Déforges, and F. Pescador, "A multi-modeling electro-optical transfer function for display and transmission of high dynamic range content," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 350–358, Nov. 2017.
[21] S. Yu, C. Jung, and Q. Lin, "HEVC encoder optimization for HDR video coding based on irregularity concealment effect," *Signal Process., Image Commun.*, vol. 64, pp. 68–77, May 2018.
[22] M. Narwaria, M. P. D. Silva, and P. L. Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Process., Image Commun.*, vol. 35, pp. 46–60, Jul. 2015.
[23] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imag.*, vol. 24, no. 1, 2015, Art. no. 010501.
[24] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.

QINGTAO FU received the B.S. degree in telecommunication engineering and the M.S. degree in information and communication engineering from Xidian University, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree. His research interests include image processing and video coding.

CHEOLKON JUNG (M'08) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Sungkyunkwan University, South Korea, in 1995, 1997, and 2002, respectively. He was with the Samsung Advanced Institute of Technology (Samsung Electronics), South Korea, as a Research Staff Member, from 2002 to 2007. He was a Research Professor with the School of Information and Communication Engineering, Sungkyunkwan University, South Korea, from 2007 to 2009. Since 2009, he has been with the School of Electronic Engineering, Xidian University, China, where he is currently a Full Professor and the Director of the Xidian Media Lab. His main research interests include image and video processing, computer vision, pattern recognition, machine learning, computational photography, video coding, virtual reality, information fusion, multimedia content analysis and management, and 3DTV.

GE YANG received the B.S. and M.S. degrees in electronic engineering from Xidian University, China, in 2014 and 2017, respectively. She is currently with ZTE, China. Her research interests include image processing and video coding.



YI LIU received the B.S. degree in electronic information engineering from Northwestern Polytechnical University, China, in 2008, the M.S. degree in telecommunications and information system from Xidian University, China, in 2011, and the Ph.D. degree from the National Institute for Applied Sciences of Rennes (INSA), France, in 2015. He was with the Department of Electronics and Computer Engineering, National Institute for Applied Sciences of Rennes (INSA), in France, from 2015 to 2017, as a Research Scientist. In 2018, he joined the School of Electronic Engineering, Xidian University, where he is currently a Lecturer. His research interests include image and video processing, image and video coding, image and video quality assessment, and audio signal processing.