A 4K/60p HEVC Real-time Encoding System with High Quality HDR Color Representations

Daisuke Kobayashi *Member, IEEE*, Ken Nakamura, Takayuki Onishi *Member, IEEE*, Hiroe Iwasaki, *Member, IEEE*, and Atsushi Shimizu

***Abstract*— This paper describes a 4K/60p High Efficiency Video Coding (HEVC, the latest coding standard) real-time encoding system with high quality high dynamic range (HDR) color representations. The HDR technology provides attractive video contents with a wider dynamic range of luminance, which can make shadows and brighter details appear clearly without any blown out highlights or blocked up shadows. We propose a new rate control method for HDR video coding. The method consists of two algorithms for improving perceptual quality by reducing degradation of HDR-specific color artifacts. The first one, adaptive block size and prediction mode decision, is conducted to suppress perceptual degradation due to prediction error in high chroma areas. The second, local quantization parameter control, is carried out to improve the visual quality with appropriate bit allocation in low chroma areas. The method was implemented in an HEVC real-time encoding system and its performance was assessed by measuring the color difference metrics by 3.86% at the most degraded frame, while PSNR remained almost the same. The subjective quality of the decoded HDR images was improved by reducing color artifacts.**

***Index Terms*—Encoder, High Dynamic Range (HDR), High Efficiency Video Coding (HEVC), Video Coding**

I. INTRODUCTION

THE latest coding standard, H.265/High Efficiency Video Coding (HEVC) [1], has been applied to 4K video broadcasting and distribution services. This standard is now beginning to be adapted to high dynamic range (HDR) images, which are rapidly spreading worldwide to make video services more realistic. To accelerate the spread of HDR video services, more efficient HEVC coding technology that can express high quality HDR is important.

Already HDR videos are providing several services such as satellite broadcasting, IPTV, and Video On Demand (VOD). In the future, the amount of HDR video is expected to increase more and more. To determine HDR video characteristics, a transfer function for mapping between optical signals and electro signals is used. In major HDR video systems, two transfer functions, HLG (Hybrid Log-Gamma) [2] and PQ (Perceptual Quantization) [3] are defined. HLG is designed to be compatible with conventional Standard Dynamic Range (SDR) video systems, mainly used for broadcasting. Although PQ is not compatible with SDR video systems, it can express higher peak luminance levels and so is mainly used for VoD and package media. Both HLG and PQ have been adopted into the parameters for HDR broadcasting service, published as international standard ITU-R Rec. BT.2100 [4]. Therefore, a lot of HEVC encoders will support these transfer functions. However, since encoding HDR video causes great perceptual degradation in video coding [5-6], it is important for a HEVC encoder to reduce the perceptual degradation. Furthermore, improving the quality of encoded HDR video should make it possible to improve the quality of SDR video reproduced from single layer HDR transmission pipelines [7]. Color artifacts are more visible in HDR video coding than in SDR, but there has been insufficient evaluation of perceptual degradation corresponding to the input video signal of a HEVC encoder.

Two issues need to be addressed to achieve efficient HEVC HDR video coding. The first is reducing color artifacts in high chroma areas. The second is reducing HDR-specific perceptual degradation in low chroma areas. Several studies have been proposed to prevent visual degradation of HDR images. Perrin et al. showed that color artifacts can be reduced by using ICtCp color space instead of Y'CbCr color space [8]. However, using different color spaces for HDR makes it necessary to redesign the parameters used to control encoders. Since it is difficult to bring a change of this magnitude to the market promptly, it is desirable to reduce perceptual degradation in Y'CbCr color spaces. Mir et al. proposed a λ-domain rate control method to optimize rate distortion performance in HDR coding; the method achieves high image quality by controlling coding on a picture-by-picture basis [9]. However, visual deterioration occurs locally when HDR video is coded, so coding control in coding units is required. Jacob et al. [10] proposed a method that reduces color artifacts using chroma QP offset, which is controlled on a picture and slice level. However, HDR-specific color artifacts occur locally at the block level and the effect of perceptual degradation depends on the component of the HDR video signal. It is easy to deal with decreasing quantization parameters to suppress quantization errors over the wide-range chroma components; however, excessive reduction of quantization parameters causes increased bitrate. Thus it is necessary to detect which ranges of the input video signal increase perceptual degradation and to deal with only their areas.

Manuscript received...

D. Kobayashi, K. Nakamura, T. Onishi, H. Iwasaki and A. Shimizu are with NTT Media Intelligence Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan (e-mail: {kobayashi.dai, nakamura.ken, onishi.takayuki, iwasaki.hiroe, shimizu.atushi}@lab.ntt.co.jp).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCE.2018.2876677, IEEE Transactions on Consumer Electronics

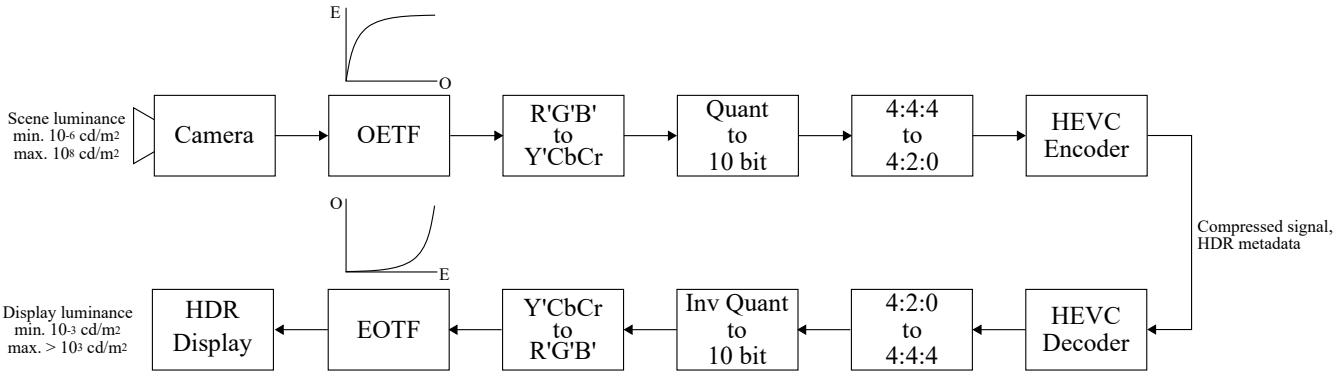


Fig. 1. Block diagram of HDR video transmission system.

To solve these problems, we have developed a new 4K/60p HEVC real-time encoder that enables high-quality HDR video coding through the use of two encoder control methods that reduce perceptual degradation. Experimental results show that improved subjective quality was obtained in both high and low chroma areas of decoded HDR images. They also show that color difference measurement metrics were suppressed by 3.86% in the most degraded frame, while the PSNR remained almost the same.

The rest of this paper is organized as follows. Section II explains the background to our research and the motivation it provided to us. Section III presents our proposed method and Section IV describes how we implemented it in our real-time 4K/60p HEVC encoder. An evaluation of HDR encoding efficiency is given in Section V and Section VI concludes with a summary of key points.

II. BACKGROUND AND MOTIVATIONS

A. HDR video transmission system

Fig. 1 shows the block diagram of a typical HDR video transmission system incorporating an HEVC encoder. Such systems include an OETF (Optical-Electro Transfer Function) and EOTF (Electro-Optical Transfer Function) to reproduce HDR images. The former maps scene luminance to video signals and the latter maps the video signals to display luminance within a display.

The input HDR video data captured by camera is in the 4:4:4 linear domain with 4:4:4 RGB color space. The input HDR video data is first converted from a linear RGB signal to a non-linear R'G'B' signal that is mapped using OETF. The obtained R'G'B' signal is converted to 4:4:4 Y'CbCr color space by ITU-R Rec. BT.2020 [11], quantized to 10-bit representation and finally subsampled to 4:2:0.

The 4:2:0 10-bit Y'CbCr signal can be encoded using the HEVC encoder and then the bitstream is transmitted. At the receiver side, the transmitted bitstream is decoded using the HEVC decoder and finally converted back to RGB linear-light by mapping EOTF before displaying the content.

The HEVC encodes a 10-bit Y'CbCr signal as an input video in the same way as the SDR image even when encoding the HDR image, as shown from Fig. 1. Therefore, encoding can proceed by dealing with the syntax to be controlled for displaying the HDR image on the display. However, the OETF and EOTF for HDR have a strong non-linearity to express widely dynamic ranges using the same resources as SDR. Assignment of video signal level and luminance are different between HDR and SDR. With 10 bit video signals, SDR and HDR respectively allocate peak brightness of 100 cd/m^2 and 10,000 cd/m^2 (cd/m^2 is the SI unit of luminance). Because of this difference in OETF and EOTF characteristics between SDR and HDR, if the same amount of pixels is varied in HDR and SDR, the variation in the amount of screen luminance is larger in HDR than in SDR, especially in high dynamic range areas. For this reason, perceptual degradation in HDR video coding is more easily seen than in SDR.

Since changes in luminance differ with the basic video signal level, higher video signal levels result in larger changes in luminance. Therefore, to reduce perceptual degradation due to coding, it is necessary to carry out adaptive video coding for HDR videos. However, since encoders for HDR videos are used in the same way as conventional encoders for SDR videos, the coding control method for HDR videos is preferably implemented within the framework of the SDR encoder architecture.

B. Color artifacts in high chroma areas

One cause of image quality deterioration in video coding is distortion due to quantization errors caused by large motion prediction errors. In the case of SDR, even when the pixel value of the decoded image is slightly different due to encoding, the difference in pixel values on the display is difficult to perceive and a good subjective quality image may be obtained. However, in HDR, since the difference in lightness given by the pixel difference is significantly larger than in SDR, perceptual degradation that cannot be perceived by SDR may be perceived. Especially in HDR, differences in lightness are not only caused by differences in luminance Y', but also by differences in chrominance CbCr. Therefore, to improve coding quality it is necessary to clarify the range of pixel values in which the lightness difference becomes large. It is also necessary to perform coding control so that degradation is not noticeable.

We examined the effects of lightness differences produced by chroma components (with constant luminance Y' value) by using the CIE 1976 L*a*b* [12] color space model. L*, a*, b* values can be derived as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCE.2018.2876677, IEEE Transactions on Consumer Electronics

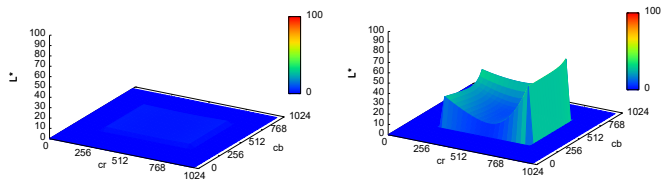


Fig. 2. Relationship of lightness L^* value depending on Cb, Cr. Left column: SDR. Right column: HDR

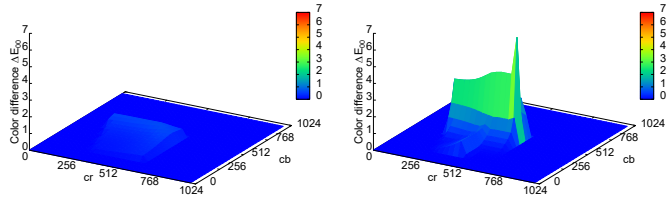


Fig. 3. Relationship of ΔE_{00} value depending on Cb, Cr. Left column: SDR. Right column: HDR.

$$L^* = 116\, f\left(\frac{Y}{Y_n}\right) - 16 \tag{1}$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \tag{2}$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \tag{3}$$

$$f(t) = \begin{cases} t^{1/3}, & t > (6/29)^3 \\ [(29/3)^3 t + 16]\,/\,116, & otherwise \end{cases} \tag{4}$$

Here, X, Y, Z, are the tristimulus values on CIE XYZ color space, and X_n, Y_n, Z_n are the reference white points.

The CIE XYZ color space can be derived as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5}$$

Here, R, G, B are the linear RGB values, which can be derived by applying a gamma function (for SDR) or EOTF (for HDR) to the non-linear $R'G'B'$ values.

Fig. 2 shows the relationship of the lightness L^* value depending on the chroma (Cb, Cr) components for both SDR and HDR formats. In the figure, L^*=100 corresponds to 10,000 cd/m^2 because both SDR and HDR signals are treated. In the case shown, the maximum value of L^* greatly differs between SDR and HDR. Although luminance Y' is the same, lightness L^* is very different depending on the value of the (Cb, Cr) components. The figure also shows that changes in the Cr value cause more rapid changes in L^* than changes in the Cb value. This is because the L^* curve changes more sharply when the Cr value changes than when the Cb value changes. Rapid changes in L^* emphasize color differences when users watch HDR videos on an HDR display. It can be seen that color artifacts, which are a particular cause of changes in the Cr value, are more visible. This is because perceptual degradation caused by prediction errors becomes larger in HDR.

C. Perceptual degradation caused by slight changes in low chroma areas

Image quality degradation due to quantization errors also occurs in images with no motion and few prediction errors. For example, neighboring pixel values gradually change in intra-frames such as those in gradation images. Color banding is a representative example of degradation. Particularly in HDR, both the dynamic range and the color gamut are extended more than in SDR, so slight differences in pixel values affect color differences on the display.

In general, to reduce quantization errors, it is sufficient to allocate more bits by performing quantization finely. However, assigning many bits results in an increased bitrate. To prevent this, it is necessary to allocate more bits to areas where degradation is conspicuous in the encoding. Therefore, it is important to identify where the degradation is noticeable. We conducted the preliminary experiment described below to identify the areas where perceptual degradation occurred.

The perceptual degradation caused by slight changes in the chroma (Cb, Cr) components is shown by calculating the ΔE_{00} value on the basis of the CIEDE2000 color difference model [13]. Since the model is based on a CIE 1976 L*a*b* color space that has perceptual uniformity, changing the same amount of color differences should produce about the same changes in visual perception. For comparing ΔE, the distance between two points in L*a*b* color space, the CIEDE2000 color difference model is designed to make it possible to obtain evaluation results with scores more similar to those for human visual perception.

The ΔE_{00} value is calculated as follows:

$$\Delta E_{00} = \left\{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)\right\}^{1/2} \tag{6}$$

where $\Delta L'$, $\Delta C'$, $\Delta H'$ are respectively the lightness difference, the chroma difference and the hue difference in the CIEDE2000 model. S_L, S_C, S_H are scaling functions and k_L, k_C, k_H are respectively weighting factors for luminance, chroma, and hue.

Fig. 3 shows that ΔE_{00} increases in the low chroma area. The maximum values of ΔE_{00} are 0.7 (SDR) and 6.4 (HDR). In SDR, ΔE_{00} is a value of 1 or less regardless of the chroma (Cb, Cr) components, and the color differences are similar for any color. On the other hand, in HDR, ΔE_{00} increases rapidly in the low chroma area. Although ΔE_{00} becomes smaller as luminance Y becomes smaller, the maximum value of ΔE_{00} is still large in the low chroma area. Therefore, the color artifacts in the low chroma area (especially the gray gradation area) are more visible in HDR video coding.

III. PROPOSED METHOD IN HDR VIDEO CODING

A. Block size control in high chroma areas

Our proposed method includes a coding block size and prediction mode decision method that reduces perceptual degradation in HDR video coding. The point of the method is to identify areas where visual degradation occurs due to perceptual lightness differences and to select a small coding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCE.2018.2876677, IEEE Transactions on Consumer Electronics

block size at these areas.

First, to detect visible color artifacts areas causing perceptual luminance differences, the method we propose calculates the inter-frame pixel difference ΔF_{diff} between the current picture and the previous picture and then counts the number of pixels N_1 including specific color ranges (Y', Cb, Cr) = (y_1 to y_2, cb_1 to cb_2, cr_1 to cr_2) in the target block. The inter-frame pixel difference ΔF_{diff} is calculated to determine whether the motion amount of the picture is large. The specific color ranges are determined in advance so as to include a range in which the perceptual lightness difference becomes large such as the high chroma areas described in Section II.

To improve video coding quality, in general, block size and prediction mode decisions are generally made by comparing rate-distortion (RD) cost evaluations [14]. The RD cost J_i for mode i is given by

$$J_i = D_i + \lambda \cdot R_i \tag{7}$$

where λ is a Lagrange parameter that corresponds to the tangent slope of the RD curve, D_i is coding distortion for mode i, and R_i is the number of bits for mode i. The encoder selects the prediction mode and block size for which RD cost has a minimum value. In our proposed method, the prediction mode and block size decisions are conducted by comparing RD cost functions calculated as

$$J_i = D_i + \alpha \cdot \lambda \cdot R_i \tag{8}$$

where α is a scaling parameter. For $\alpha < 1$, coding distortion is given priority over the number of bits in (2), and a smaller mode of coding distortion is finally selected. In the proposed method, α is set as follows:

$$\begin{cases} 0 < \alpha < 1, & \Delta F_{diff} > TH_1 \text{ and } N_1 > TH_2 \\ \alpha = 1, & otherwise \end{cases} \tag{9}$$

When the ΔF_{diff} surpasses threshold TH_1 and N_1 surpasses threshold TH_2, α is set to less than 1 in (3) so that smaller size coding blocks are selected to reduce prediction errors. In other cases, α is set to equal 1. This is the same as the usual RD cost evaluation formula. Since a smaller size coding block is selected to reduce prediction errors, perceptual degradation by chroma components due to video coding can be reduced. The proposed method operates in the same way as the existing block size and prediction mode decision framework, so it is able to predict and determine block sizes for HDR videos with only slight modifications.

B. QP control in low chroma areas

To keep the given bitrate and assign bits for the best image quality, a rate control method is conducted. In the method, the quantization parameter (QP) is basically controlled to keep the target bitrate. Local QP control is performed to improve subjective quality after determining the base QP (QP_{base}), which is the QP for each frame. This determines the QP of each coding unit (CU). In MPEG-2 Test Model 5 [15], the major rate control algorithm, the QP is decreased if the activity of the target block is low and increased if the activity is high.

With our proposed method, local QP adjustment reduces perceptual degradation in low chroma areas. The method counts the number of pixels N_2 in the target coding block, which includes specific color ranges: (Y', Cb, Cr) = (y_3 to y_4, cb_3 to cb_4, cr_3 to cr_4). These specific color ranges are determined so as to include ranges in which the perceptual color difference ΔE_{00} becomes large, such as the low chroma area shown in Section II.

The proposed method sets a small QP for the target block in which perceptual degradation may occur due to quantization errors. If the QP_{base}, which is calculated by any rate control algorithm in the target coding picture, surpasses the pre-defined threshold TH_3 and N_2 surpasses the pre-defined threshold TH_4, a new quantization parameter QP_{new} in this target block is calculated as

$$QP_{new} = QP_{base} + \Delta QP \tag{10}$$

where ΔQP values are calculated on the basis of the QP_{base} and N_2 in the target block. Perceptual degradation increases as QP_{base} becomes higher and N_2 becomes larger, so a larger negative value is set to ΔQP as the QP_{base} and N_2 for block increases.

Our proposed method controls bit allocation in order to adaptively change the QP between the blocks with noticeable perceptual degradation and unnoticeable blocks. Therefore, this method can improve subjective quality while maintaining the overall bitrate.

IV. IMPLEMENTATION

A. Encoder architecture

The encoder we describe in this paper was developed using the high-performance 4K/60p HEVC encoder LSI (codenamed "NARA") [16] we previously developed for 4K/8K video services. Fig. 4 shows a summarized block diagram of NARA relevant to the proposed method.

The structure of the image feature extraction (IFE) block in Fig. 4 is described in Fig. 5. In the IFE block, input image data is read and stored in line buffer every 32 lines, and several types of image feature parameters (e.g., the inter-frame difference value, the number of pixels with the specific color, the average, the variance) are extracted and calculated in the unit of 16×16 pixels. These feature parameters are aggregated for each coding block size (i.e., 16×16, 32×32, and 64×64) and stored in an external memory.

The local parameter control (LPC) block in Fig. 4 decreases QP in low chroma areas to reduce color artifacts. The structure of the LPC block is shown in Fig. 6. In the LPC block, a new QP is calculated on the basis of QP_{base}, ΔQP value, and the feature value in the target coding block. The LPC block obtains the future values from an external memory for each coding block and then determines which QP control tables (ΔQP tables) are used on the basis of the QP_{base}, which is calculated by in-picture rate control.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCE.2018.2876677, IEEE Transactions on Consumer Electronics

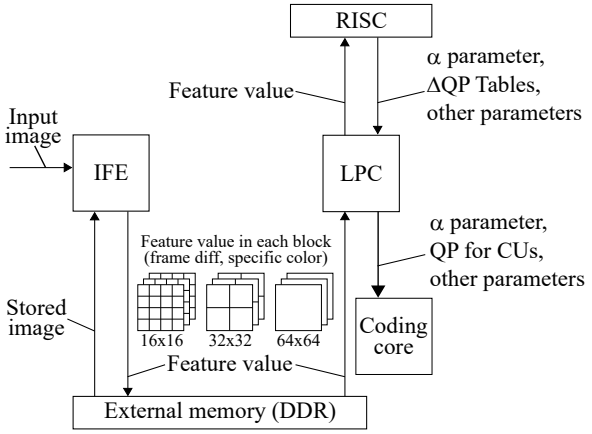


Fig. 4. Summarized block diagram of NARA LSI

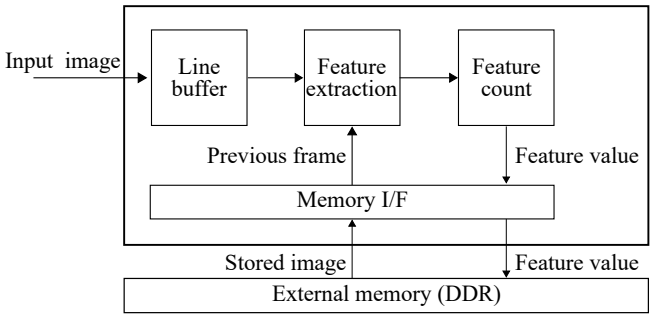


Fig. 5. Image feature extraction (IFE) block

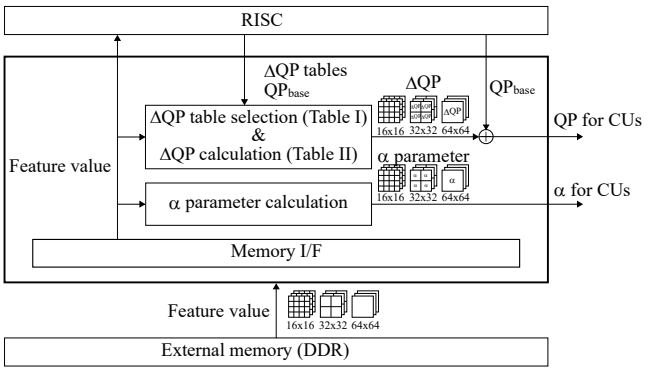


Fig. 6. Local parameter control (LPC) block

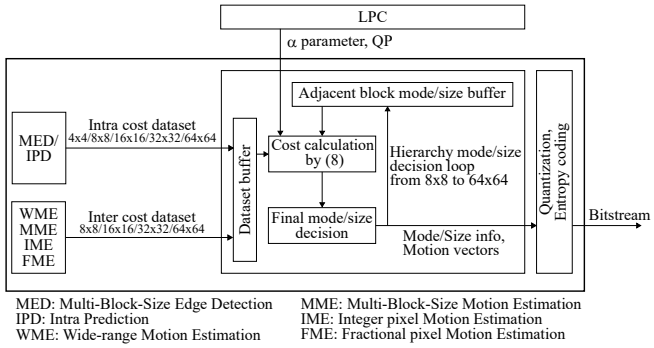


Fig. 7. Coding core block

The ΔQP tables are designed to have a large negative *ΔQP* value when the QP_{base} value increases. The *ΔQP* value depends on the number of specific color pixels given by the IFE block. It is set to a large negative value when an increased number of specific color pixels is included in each block. In this proposed method, a two-stage look-up table is used to determine *ΔQP* in accordance with the QP_{base} value and the specific color number. Table I lists the ΔQP values used in accordance with the QP_{base}, and Table II lists the ΔQP values used in accordance with the number of specific colors in the target block. The ΔQP value is set for each coding block size and the coding core block quantizes by using (10), whose *ΔQP* value is obtained from the LPC block. Furthermore, α parameter for (8) is also calculated by counting the specific colors for each coding block size.

The coding core block in Fig. 4 determines the RD-cost-effective prediction mode and block size. The structure of the coding core block is described in Fig. 7. To obtain the most suitable prediction mode and block size, intra prediction (MED/IPD in Fig. 7) and inter prediction (WME, MME, IME, FME in Fig. 7) are performed for each block size. These intra/inter prediction costs are transferred to the cost calculation block and compared. The mode/size determination is executed as a hierarchy loop from the 8×8 to 64×64 block size. Prediction mode and block size are determined by (8) so that smaller blocks will be selected if the number of pixels having considerable perceptual luminance differences exceeds the threshold, using feature values given by the IFE block.

Controlling these processes by using RISC-based firmware enabled us to speedily develop the encoder by using our previous resources. The firmware also implements syntaxes for HDR video formats.

TABLE I
ΔQP TABLE SELECTION BASED ON BASE QP

Base QP range	ΔQP table
QP_{base} <= 42	ΔQP table (0)
42 < QP_{base} <= 48	ΔQP table (1)
48 < QP_{base}	ΔQP table (2)

TABLE II
ΔQP TABLES

	Number of specific colors (N_2) in 16x16 pixels					
	< 160	< 184	< 208	< 232	< 255	>= 255
ΔQP table (0)	0	0	0	0	-1	-1
ΔQP table (1)	0	0	-1	-2	-2	-2
ΔQP table (2)	-1	-2	-3	-4	-5	-5

TABLE III
ENCODER SPECIFICATIONS

Input Video Format	4K/60p (3840x2160 / 59.94, 60 fps / progressive)
Input I/F	SMPTE 424M, Y'CbCr 4:2:2, 10 bit x4
Output I/F	DVB-ASI, IP (MPEG-2 transport stream)
Supported HDR type	PQ (ST2084)
Size	1U

B. Specifications

Table III lists the encoder specifications. This encoder has 3G-SDI x4 interfaces for input video and DVB-ASI and IP interfaces for HEVC streams output. The coding units in the HEVC encoder are high-performance encoder LSIs we previously developed. We were able to fabricate the encoder with a 19-inch rack 1U size.

V. EVALUATION

We conducted an objective quality evaluation to ascertain the validity of the proposed method and the implemented encoder's coding quality performance. We calculated the color difference ΔE_{00} and the peak signal-to-noise ratio (PSNR) as evaluation metrics and used them to evaluate the quality of decoded images. The ΔE_{00} values indicate that the differences from the original image become less as the values become smaller, and the PSNR values indicate that the differences become less as the values become larger. Table IV lists the evaluation conditions and Fig. 8 shows the test sequences. Each test sequence is included the following viewpoints: 1) an image with large motion ("(a) Birds"), 2) an image including lots of colors ("(b) Garden"), 3) an image with lots of low chroma color ("(c) Women"), and 4) an image in which block noise and banding easily occurs ("(d) Sunset"). In Table IV, GOP (group of pictures) specifies the order in which intra- and inter-frames are arranged. M is the distance between two anchor B/P-pictures and N is the distance between two I-pictures. The N_1, N_2, TH_1, TH_2, TH_3, TH_4 and α values were defined in preliminary experiments. To determine the suitable α value to minimize ΔE_{00} on average, we conducted encoding simulations with several α and QP values. Table V lists the ΔE_{00} obtained in the simulations. The simulations were performed with α={1.0 (conventional), 0.3, 0.5, and 0.8} and QP={32, 37, 42, and 47}. From Table V, since ΔE_{00} is the minimum in the case of $\alpha = 0.5$ on average, α were determined to be 0.5. The values of N_1 and N_2 were determined from the analysis described in Sec. II and test images confirmed whether their ranges covered perceptual degraded colors. The TH_1, TH_2, TH_3, and TH_4 values were also determined by performing several encoding simulations.

Table VI lists a comparison of ΔE_{00} results obtained with the proposed and conventional methods. The improvement ratio for the entire sequence gain and maximum gain per frame in each test sequence is shown. The entire sequence gain G_{seq} is calculated by

$$G_{seq} = \left(1 - \frac{\underset{i}{\text{ave}}\left(E_{00}\left(F_o(i), F_p(i)\right)\right)}{\underset{i}{\text{ave}}\left(E_{00}\left(F_o(i), F_c(i)\right)\right)}\right) \times 100 \qquad (11)$$

where ave(...) is a function to calculate average values of the argument. $E_{00}(a, b)$ is a function to calculate ΔE_{00} between the i-th original frame and the decoded frame. The $F_o(i)$, $F_p(i)$, and $F_c(i)$ values are respectively those for the i-th original frame, the decoded image obtained with the proposed method, and the decoded image obtained with the conventional method. The maximum gain per frame G_{frame} is calculated by

TABLE IV
EVALUATION CONDITIONS

Encoder	HEVC Encoder LSI "NARA"
Profile, Level	HEVC Main 4:2:2 10, Level 5.1
Input video format	3840x2160 / 59.94 fps / progressive
Colorimetry	ITU-R Rec. BT.2020
HDR format type	PQ (ST2084)
GOP structure	M = 8, N = 32
Bitrate	15 Mbps
Test sequences	4 images from LUCORE [17]
Number of frames	192 frames per sequence
Color range for N_1	Y': 280 to 550, Cb: 400 to 490, Cr: 560 to 650
TH_1	16% in frame
TH_2	75×10^6
α	0.5
Color range for N_2	Y': 495 to 640, Cb: 496 to 527, Cr: 496 to 527
TH_3	same as the base QP range in Table I
TH_4	same as the number of specific colors in Table II

(a) Birds (b) Garden

(c) Women (d) Sunset

Fig. 8. Test sequences [17]

$$G_{frame} = \max_i \left(1 - \frac{E_{00}\left(F_o(i), F_p(i)\right)}{E_{00}\left(F_o(i), F_c(i)\right)}\right) \times 100 \qquad (12)$$

where max(...) is a function to calculate the highest value of the argument. With the proposed method, ΔE_{00} is reduced more than with the conventional method for all images. The entire sequence gain improvement in ΔE_{00} is small since the proposed method only processes in cases where the number of specific color pixels exceeds the thresholds described in Section II. On the other hand, G_{frame} increased by at most 3.86 %. Fig. 9 shows the ΔE_{00} per frame in the test sequence. The proposed method shows that ΔE_{00} is decreased among some sequential frames in each sequence. Table VII lists the gain G_{seq} tallied for each picture type. The gain is large for the I-picture for all sequences, so the gain increased for the other picture types, which refer to the I-picture. In particular, the gain for "Birds" improved at

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCE.2018.2876677, IEEE Transactions on Consumer Electronics

TABLE V
ΔE_{00} OF ENCODED IMAGES WITH DIFFERENT QP AND α VALUES

	α=1.0 (conventional)	α=0.3	α=0.5	α=0.8
QP=32	2.267	2.292	2.29	2.29
QP=37	2.597	2.647	2.641	2.638
QP=42	3.162	3.183	3.193	3.176
QP=47	4.502	4.364	4.279	4.487
Average	3.132	3.122	3.101	3.148

TABLE VI
EVALUATION RESULTS (ΔE_{00})

Sequence	Conventional	Proposed	Gain G_{seq} (%)	Gain G_{frame} (%)
Birds	2.287	2.265	0.96	3.86
Garden	1.383	1.380	0.21	1.99
Women	1.470	1.462	0.54	1.37
Sunset	2.158	2.157	0.05	2.19

TABLE VII
EVALUATION RESULTS (GAIN G_{SEQ} TALLIED FOR EACH PICTURE TYPE)

Sequence	Gain G_{seq} tallied for each picture type (%)				
	I	P/B0	B1	B2	B3
Birds	0.837	0.308	0.544	0.865	1.211
Garden	0.797	0.469	0.521	0.58	0.564
Women	0.867	0.357	0.083	0.204	0.222
Sunset	0.723	-0.074	-0.064	0.025	0.027

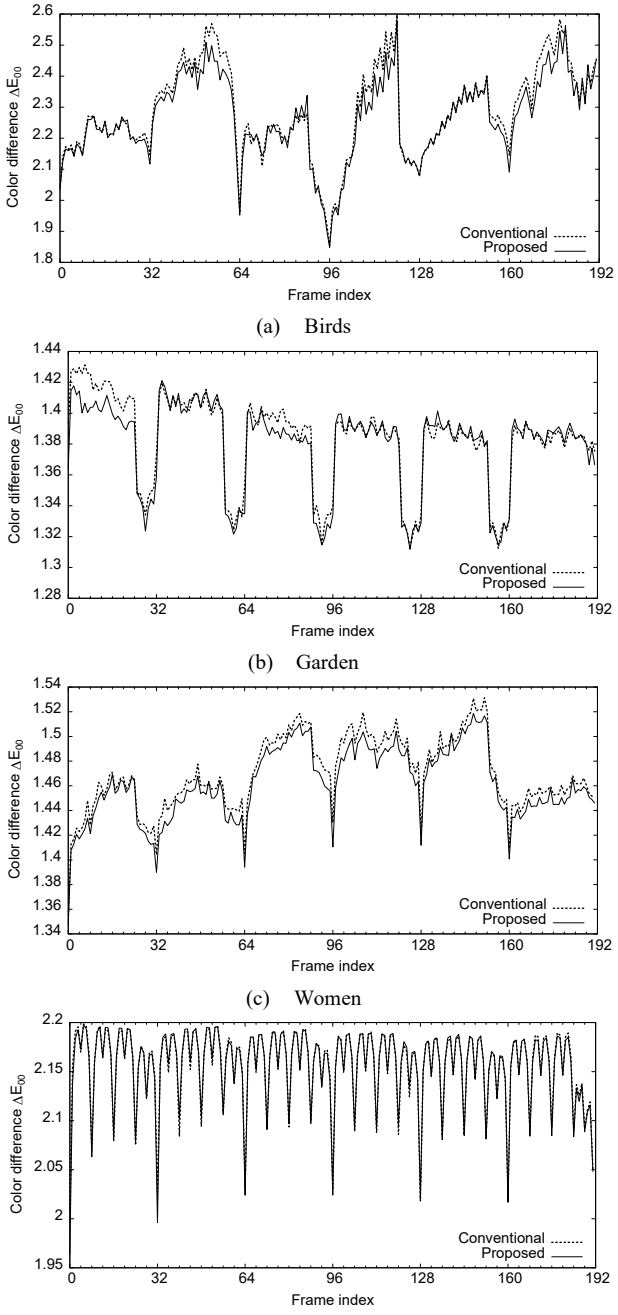
TABLE VIII
EVALUATION RESULTS (PSNR [DB])

	Conventional		Proposed	
	PSNR-Y	PSNR-C	PSNR-Y	PSNR-C
Birds	31.09	35.82	31.11	35.83
Women	41.27	46.23	41.26	46.21
Garden	41.66	43.29	41.67	43.29
Sunset	38.03	43.95	38.00	42.39

higher coding layers. This shows that reducing perceptual degradation that occurs when coding a reference picture helps to improve the quality of the entire sequence, since the proposed method can reduce prediction errors due to large motion.

Table VIII shows a comparison of the PSNR results. The PSNR is slightly lower with the proposed method than with the conventional method. This is because the proposed method uses an in-frame quantization control method whose aim is to improve perceptual quality.

Fig. 10 shows decoded image examples. (Note: the images are tone mapped for publication.) With the conventional



(a) Birds

(b) Garden

(c) Women

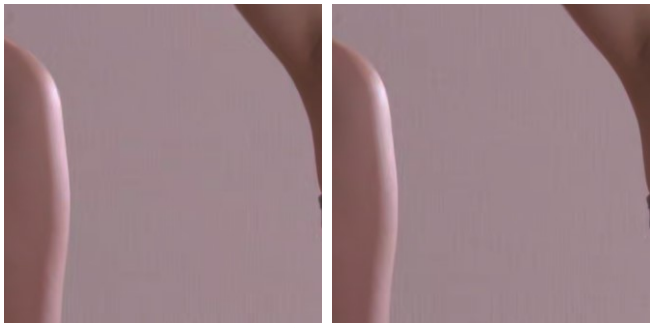
(d) Sunset

Fig. 9. ΔE_{00} results

method, the color distortion of cyan and magenta can be seen in the gray part of the wall, but with the proposed method, these color distortions are reduced (see the left side of (a)). With the proposed method the artifacts are not visible because of quantization parameter control in low chroma areas (see the right side of (a)). The proposed method also removed color artifacts near the boundary of the moving bird and the background since smaller coding blocks are selected during the mode size decision control process. Furthermore, with the conventional method, the color distortion of cyan and magenta can be seen in the white part of the clothes, but with the proposed method, these color distortions are reduced (see the left side of (b) and the right side of (b)). In other cases, the proposed method reduces the color artifacts.

(a) Birds

(b) Garden



(c) Women

(d) Sunset

Fig. 10. Comparison of decoded images
Left: Conventional method processing. Right: Proposed method processing.
Color artifacts on the left are removed on the right.

VI. Conclusion

In this paper, we proposed two methods that take perceptual degradation in high quality HDR video encoding into account to achieve more realistic video transmission. The methods were implemented in our 4K/60p real-time HEVC encoder and found to reduce color artifacts of decoded HDR videos. The implementation results obtained with a very high-quality encoder promise to assist in the rapid spreading of HDR video services.

References

[1] G. J. Sullivan, J-R. Ohm, W.-J. Han, and T. Weigand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[2] ARIB STD-B67, “Essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production,” 2015.

[3] SMPTE ST 2084:2014 “High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays,” SMPTE, 2014.

[4] ITU-R, "Image parameter values for high dynamic range television for use in production and international programme exchange," Recommendation ITU-R BT.2100-0, 2016.

[5] L. Kerofsky, Y. Ye, Y. He, Recent developments from MPEG in HDR video compression", *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 879-883, Sep. 2016.

[6] S. Iwamura, A. Ichigaya, Y. Nishida, "Comparison of Coding Performance between Hybrid Log-Gamma system and Perceptual Quantization system: Objective quality evaluation between different HDR formats," *ITE technical report*, Vol.40, No.14, pp. 9-12, Mar. 2016.

[7] M. Azimi, R. Boitard, M.T. Pourazad, P. Nasiopoulos, “Performance evaluation of single layer HDR video transmission pipelines”, *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 267-276, Aug. 2017.

[8] A. F. Perrin, M. Rerabek, W. Husak, T. Ebrahimi, “ICtCp Versus Y'CbCr: Evaluation of ICtCp Color Space and an Adaptive Reshaper for HDR and WCG”, *IEEE Consum.Electron. Mag.*, vol. 7, no. 3, pp. 38-47, May 2018.

[9] J. Mir, D. S. Talagala, and A. Fernando, “Optimization of HEVC λ-domain Rate Control Algorithm for HDR Video,” *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, pp.825-826, 2018.

[10] J. Ström, K. Andersson, M. Pettersson, P. Hermansson, J. Samuelsson, A. Segall, J. Zhao, S. Kim, K. Misra, A. Tourapis, Y. Su, and D. Singer, "High Quality HDR Video Compression using HEVC Main 10 Profile," *Proc. Picture Coding Symp. (PCS)*, pp. 1-5, Dec. 2016.

[11] ITU-R, "Parameter values for ultra-high definition television systems for production and international programme exchange," Recommendation ITU-R BT.2020-2, 2015.

[12] Colorimetry – Part 4: CIE 1976 L*a*b* Colour Spaces, ISO 11664-4/CIE 014-4, 1976.

[13] Colorimetry – Part 6: CIEDE2000 Colour-Difference Formula, ISO 11664-6/CIE 014-6, 2013.

[14] G. J. Sullivan, R. L. Baker, "Rate-distortion optimization for video compression,", *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[15] Test Model 5 ISO/IEC JTC1/ SC29/ WG11/ N0400 MPEG93/457, Apr. 1993.

[16] T. Onishi, T. Sano, Y. Nishida, K. Yokohari, J. Su, K. Nakamura, K. Nitta, K. Kawashima, J. Okamoto, N. Ono, R. Kusaba, A. Sagata, H. Iwasaki, M. Ikeda, and A. Shimizu, "Single-chip 4K 60fps 4:2:2 HEVC Video Encoder LSI with 8K Scalability," *Dig. Symp. VLSI Circuits,* pp. C54-C55, Jun 2015.

[17] UHD/HDR Test Sequences 'LUCORE', IMAGICA Corp., [Online]. Available: http://www.imagica.com/news/lucore/

Daisuke Kobayashi (M’18**)** received the B.E. and M.E. degrees in Information and Communication Engineering from the University of Electro-Communications, Japan, in 2007 and 2009. He joined NTT Cyber Space Laboratories, Nippon Telegraph and Telephone Corporation (NTT), Kanagawa, Japan in 2009 and since then has been engaged in research and development work on high quality image coding and transmission. He is

currently a Research Engineer at Visual Media Project of NTT Media Intelligence Laboratories, Kanagawa, Japan. He is a member of IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Ken Nakamura received the master's degree in Science and Technology from Keio University, Japan, in 1997. In 1997, he joined NTT Human Interface Laboratories, Nippon Telegraph and Telephone Corporation (NTT), Kanagawa, Japan, and since then has been engaged in research and development work on video coding and processing. He is currently a Senior Research Engineer doing work for the Visual Media Project of NTT Media Intelligence Laboratories, Kanagawa, Japan. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the Institute of Image Information and Television Engineers.

Takayuki Onishi (M'17) received the master's degree in Information and Communication Engineering from the University of Tokyo, Japan, in 1999. He joined NTT Cyber Space Laboratories, Kanagawa, Japan in 1999 and since then has been engaged in research and development on high quality image coding and transmission. He is now a Senior Research Engineer in NTT Media Intelligence Laboratories, Kanagawa, Japan and his research interests include hardware and software implementation of high performance video encoders and network video distribution technologies. He received the 2016 Achievement Award from the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan. He is a member of IEEE, IEICE and Information Processing Society of Japan (IPSJ).

Hiroe Iwasaki (M'09) received the B.S. and Ph.D. degrees in Information Science from Tsukuba University, Ibaraki, in 1991 and 2006. She joined NTT Switching System Laboratories in 1991 and since then has been engaged in R&D work on an architecture and real-time operating system for multi-media embedded system LSIs. She is currently a Senior Research Engineer and a Supervisor doing work for the Visual Media Project of NTT Media Intelligence Laboratories, Kanagawa, Japan. She is a member of IEEE, the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the Information Processing Society of Japan (IPSJ).

Atsushi Shimizu received the B.E. and M.E. degrees in Electronic Engineering from Nihon University, Tokyo, in 1990 and 1992. Since joining NTT in 1992, he has been engaged in work on video compression algorithms and software development. He is currently a Senior Research Engineer and a Supervisor doing work for the Visual Media Project of NTT Media Intelligence Laboratories, Kanagawa, Japan. He is a member of the Institute of Electronics, Information and Communications Engineers (IEICE) of Japan, the Institute of Image Electronics Engineers of Japan (IIEEJ) and the Institute of Image Information and Television Engineers.