Weighted-prediction-based color gamut scalability extension for the H.265/HEVC video codec

Alireza Aminlou[1,2], Kemal Ugur[1], Miska M. Hannuksela[1] , Moncef Gabbouj[2]
[1] Nokia Research Center, Tampere, Finland, [2] Tampere University of Technology, Tampere, Finland

ABSTRACT

Color gamut scalability refers to coding a video in a layered manner where the base and enhancement layers are coded in different color gamut spaces. Color gamut scalability and its relationship with spatial scalability are currently being studied for the scalable extension of HEVC (SHVC) to enable coding of ultra-high definition content having BT.2020 color gamut with 10-bit precision as an enhancement layer and high definition content having BT.709 color gamut with 8-bit precision as the base layer. In this paper, we propose to use the weighted prediction tool of the SHVC standard to map the color gamut of the base layer to the enhancement layer. In addition, we also propose a high-precision bit-depth mapping of the base layer to the enhancement layer that jointly performs upsampling with a bit-depth increase. Simulation results show that these two schemes improve the coding efficiency of the All Intra and Random Access configurations by about 6.8% and 3.6% on average, respectively, compared to a basic scheme where the bit-depth of the base layer is increased by simple bit-shifting. These gains are achieved by imposing no changes to the SHVC standard; hence make the proposed method very useful for practical use-cases as well.

***Index Terms*—**Color gamut scalability, Scalable video coding, Scalable extension of H.265/HEVC (SHVC)

1 INTRODUCTION

The emerging 4K ultra-high definition (UHD) as the new standard for the recent multimedia services has several advantages compared to conventional standard definition (SD) and high definition (HD) production by providing higher resolution picture format and higher dynamic range of real world colors. The latter one is achieved by using different color gamut than HD format. Particularly, UHD is using BT.2020 recommendation [1] that has wider color range and higher bit depth than BT.709 [2], used in HD.

The recently developed and finalized High Efficiency Video Coding (H.265/HEVC) standard by the Joint Collaborative Team on Video Coding (JCT-VC) supports the coding of videos with UHD picture format. However, there is an interest to support both HD and UHD format in a single bitstream format using the scalable extension of HEVC standard (SHVC) [3]. With this approach, the devices with non-UHD decoder or limited network bandwidth can benefit the base layer (BL) bitstream with HD picture format, while the other devices with UHD decoding capability can receive the UHD bitstream. Furthermore, the support of the non-HEVC BL in current SHVC allows transmission of the content to devices with H.264/AVC decoder, which allows introduction of UHD services in a backwards compatible manner.

In order to make sure the resulting SHVC standard serves the industry widely, the enhancement layer (EL) bitstream should have minimal changes compared to BL H.265/HEVC bitstream and needs to be parsable with an H.265/HEVC decoder with minor modifications. The most straight forward scalability tool in SHVC is the inter-layer prediction (ILP) which makes the (upsampled) base layer picture, known as inter-layer reference (ILR) picture, available for EL sample prediction. As the BL and EL are using different color gamut, color gamut mapping is required to transform the data of the BL with a color gamut (e.g. ITU-R BT.709 8-bit) to the data of the EL with another color gamut (e.g. ITU-R BT.2020 10-bit).

There are several approaches proposed for color gamut mapping such as simple bit shifts of BL data, shift-offset [4] and gain-offset model [5]. The improved method, further explained in subsection 2.2, are piecewise linear prediction model [6], region-based gain-offset predictor [7], and 3D Look-Up Table [8] and [9]. In all the existing methods, the mapping process is applied on ILR picture, either before or after upsampling. It should be noted that most of these approaches require changes in the lower level processing of SHVC standard. Furthermore, some changes in bitstream syntax and parsing are needed as some parameters for color gamut mapping need to be separately signaled.

In this paper, we propose to use the weighted prediction tool of the SHVC standard for color gamut mapping. In this method, the weighed prediction motion compensation is enabled for the ILR picture inserted in the EL reference picture list. The weight and offset values are calculated in a frame basis and are signaled in the slice header using the syntax already available in SHVC standard. As a result, no change is applied to the standard, while still achieving most of the coding efficiency which is about 6.8% and 3.6% on average, for the All Intra (AI) and Random Access (RA) configurations, respectively.

The rest of the paper is organized as below. Color gamut concept and the existing mapping methods are reviewed in section 2. The proposed method based on the weighted prediction tool of SHVC standard is introduced in section 3. Simulation results are provided in section 4, followed by conclusions in section 5.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

2 BACKGROUND AND EXISTING METHODS

2.1 Color Gamut of BT.709 and BT.2020

The difference of two color spaces of BT.709 and BT.2020 can graphically be presented with the chromaticity diagram in Figure 1, in which the color gamut spaces of BT.709 and BT2020 are shown inside the space of surface color. As shown in Figure 1, they have the same white point definition in the middle, the blue color has almost the same range but red and green corners are expanded to include a larger range of colors.

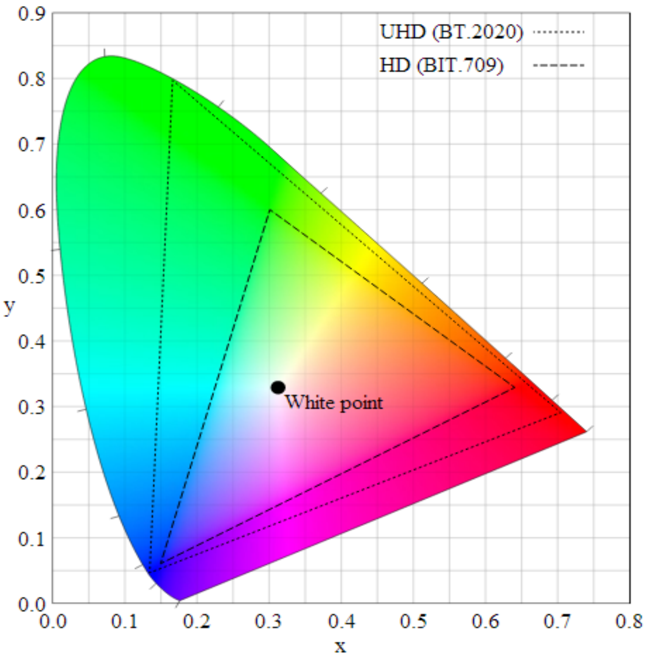

Figure 1. The Chromaticity diagram showing the difference between HD (BT.709) and UHD (BT.2020)

The relation between two color spaces can be extracted by analyzing the relation of each color space to the common intermediate, known as XYZ. The output of a digital camera is a video with wide color range which is represented with XYZ color space with 16-bit raw data. The picture data is then converted from XYZ color space to BT.709 and BT.2020 picture format using the same steps as below:

1- Converting XYZ to linear RGB using the definitions in each color space
2- Converting linear RGB to non-linear RGB (R'G'B') using the non-linear scalar function of (1) which individually operates on each color component and is the same for both color space

$$f(x) = \begin{cases} 4.5x & 0 \le x < \beta \\ \alpha x^{0.45} - (1-\alpha) & \beta \le x \le 1 \end{cases} \quad (1)$$
$$\alpha = 1.099 \qquad \beta = 0.018$$

3- Converting R'G'B' to YCrCb using the transform matrix defined in each color space
4- Chroma downsampling, for example, to 4:2:0
5- Spatial resolution downsampling (only for BT.709)

2.2 Existing Methods for Color Gamut Mapping

There are several approaches proposed for color gamut mapping to be applied either before or after upsampling. The simplest prediction model is BitShift which shifts BL data to the left by appropriate number of bit to compensate only bit-depth difference of the BL and EL. The improved model is shift-offset where an offset is added to data after shifting them. In the other method, a linear gain-offset model is used to individually operate on each color component [4]. The improved version of this method is a piecewise linear prediction model [6] which uses two different sets of weight and offset for two different ranges of each color component of the BL. The regions are defined by comparing the pixel values to a breakpoint value as a threshold. In the other similar method, each picture is divided to several rectangular regions (e.g. 16 regions), each has its own gain-offset predictor [7].

The more sophisticated approach is based on a 3D Look-Up Table [8] and [9] where YCrCb domain of the based layer is divided to NxNxN grid, and YCrCb values of the EL are defined for all (N+1)x(N+1)x(N+1) node. Then, for each YCrCb of the BL, the appropriate region is extracted and the YCrCb of the BL is calculated by the YCrCb of the EL of the neighboring nodes using a tri-linear interpolation. In [8], BL data has been upsampled and bit-depth converted to 10-bit, and then the color gamut conversion is applied. But in [9], the color gamut mapping with bit-depth conversion is applied first on BL data, and then, the resulting picture is upsampled to the EL resolution. The results in [9] show that combining the bit-depth conversion with color gamut mapping results in better quality performance than combining the bit-depth conversion with upsampling process.

It should be noted that most of the above methods require a change in process and bitstream syntax of SHVC standard. For example, required parameters such as weights, offsets and thresholds, are signaled in the EL bit-stream.

3 THE PROPOSED METHOD

The data of a pixel from BT.709 color space can be converted to BT.2020 color space by reversely converting the data to XYZ color space following the BT.709 specification described in subsection 2.1, and then converting the XYZ data to BT.2020 following the corresponding specifications. It is, however, computationally expensive as requires several matrix-based transforms and non-linear mapping.

It has been shown in the literature that the linear weight-offset prediction is a computationally friendly method that can capture the relation between two different color spaces [10]. In the existing methods, the weight-offset mapping has been implemented as the inter-layer processing this. This approach not only requires a change in the processes of the standard, but also imposes changes in SHVC bit-stream.

In this paper, we have proposed to use the weighted prediction tool, already available in SHVC standard, to realize the weight-offset mapping of ILR's data. It should be reminded that in the current draft of SHVC [11], the upsampled BL (i.e. ILR) picture is added the reference list of the EL and is available for EL sample prediction. Hence, the weighted prediction tool can be enabled for ILR picture to compensate the color gamut deference using the weight-offset prediction model. In this approach, as we are using the standard tool, the BL picture is first upsampled, and then the color gamut mapping is applied by weighted prediction.

3.1 Bit-depth compensation

As described in the introduction, the bit-depth of the BL (BT.709) can be less or equal to that of the EL (BT.2020). In the case that they are using the same bit precision, no bit-depth conversion is needed. In this case, no extra process is required to be applied between the processes of upsampling the BL picture and applying the weighted prediction.

On the other hand, bit-depth conversion is required when two layers are using different bit-precisions. In fact, the bit-depth difference of two layers is not directly related to color gamut different. In other words, the BL and EL with the same color gamut may also have different bit-depth, or vice versa. Hence, the bit-depth conversion is applied first to compensate the difference in bit precision of two layers.

The BL picture is upsampled to generate ILR picture with a higher spatial resolution (e.g. 2x) and the same bit-depth of the BL (BD_{BL}) data. It is noted that the intermediate samples after upsampling operation are stored in 16-bits. In the straightforward implementation, shown in Figure 2-(a), these intermediate samples are first rounded to the BL bit-depth and then shifted back to higher bit-depth of EL (BD_{EL}). This cascaded operation reduces the coding efficiency as the precision of intermediate samples are lost. Furthermore, additional bit-shift operations are required which increase the complexity.

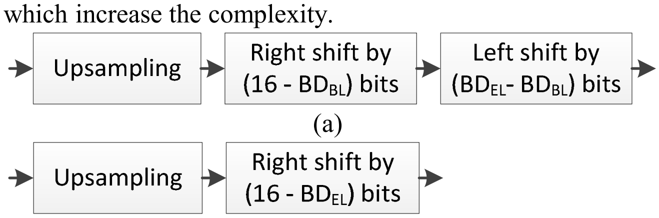

Figure 2. Bit-depth compensation (a) The straightforward implementation (b) The proposed joint upsampling and bit shift

Instead of this cascaded operation of upsampling and bit-depth increase, we instead propose to perform upsampling and bit-depth increase jointly (JUsSh). This is illustrated in Figure 2-(b) and requires a simple change to the upsampling process. As seen in Figure 2-(b), the final stage of upsampling is adjusted so that the intermediate samples are shifted not to BL bit-depth but to EL bit-depth. This allows keeping the intermediate precision as high as possible and eliminates a redundant bit-depth shift operation. Because of these reasons, this method was accepted as a part of inter-layer processing tool in SHVC standard.

3.2 Weight-offset parameter calculation

Weight and offset parameters of the weighted prediction tool should be optimally selected for ILR picture to capture the color gamut different by least square error (LSE) method which minimizes the distortion between the original EL picture and the mapped BL picture as presented in (2),

$$\{w_C, o_C\}=\text{argmin}\left\{\left(C_{EL}-(w_C.C_{BL}+o_C)\right)^2\right\} \quad (2)$$

where C_{EL} and C_{BL} are different color components (Y, Cr, or Cb) of the EL and BL respectively, and w_C and o_C are the weight and offset parameters the color component of C. The calculated weight and offset parameters are appropriately converted to fix precision factors and signaled in the slice header for each frame.

4 SIMULATION RESULTS

The R-D performance results of the proposed method have been reported in this section and have also been compared with other existing methods. Our anchor software is the SHM3.0.1 provided by JCT-VC for the development of color gamut support for scalable extension of H.265/HEVC.

In our experiments, we have used the JCT-VC common test conditions for SHVC [13]. The test sequences for color gamut scalability includes five test sequences as presented in Table 1, each has the following picture formats:

- UHD (3840x2160, BT.2020, 10-bit): used as the EL input
- HD (1920x1080, BT.709, 10-bit): used as the BL input
- HD (1920x1080, BT.709, 8-bit): used as the BL input

Table 1: Test sequences for color gamut scalability

Seq. ID	Sequence Name	Frames per second	Number of frames
BD	Birthday	60	300
BF1	BirthdayFlashPart1	60	300
BF2	BirthdayFlashPart2	60	300
PK	Parakeets	50	250
TC	TableCar	60	300

As seen above, the input to the EL is always UHD picture format with 10-bit data, and the input to the BL is HD picture format with either 10-bit or 8-bit data

Four different quantization parameters (QP) of 22, 26, 30, and 34 are used for BL encoding. And two different tests of DeltaQP0 and DeltaQO2 are used for EL encoding in which the difference between QPs of EL and BL is 0 and 2, respectively. BL sequences are obtained by downsampling the full resolution sequence by factor of two (2x) using the default downsampler defined in H.265/HEVC scalable extension call for proposals [14].

R-D performance is separately calculated using BD-rate [15] for luma and chroma components, where negative number indicates coding performance gain. BD-rate_{YUV} is the weighted average of the BD-rate values of three components, and is calculated using (3).

$$\text{BD-rate}_{YUV} = \frac{(8\times\text{BD-rate}_{Y} + \text{BD-rate}_{U} + \text{BD-rate}_{V})}{10} \quad (3)$$

4.1 Coding efficiency performance

The details of the R-D performance (BD-rate) of the weighted prediction (WP) approach compared to the basic scheme of BitShift have been reported in Table 2 for all test sequences and two configurations of All Intra (AI) and Random Access (RA). In this case, the BL input is HD picture format with 8-bit data, and weight-offset parameters are adaptively (Adp) calculated for each frame using the LSE method described in subsection 3.2. The average values of different categories have also been reported in this table. The average BD-rate for DeltaQP2 is higher than that of DeltaQP0, because the quality of temporal reference frames in DeltaQP2 case are lower than that in DeltaQP0, hence ILR picture is referred more. For a similar reason, the proposed method shows better R-D performance with AI than RA as there is no temporal reference frame.

Table 2: R-D performance (BD-rate%) results for weighted prediction with adaptive weight selection (BL: 8-bit, Anchor: BitShift)

Delta QP	Test Seq.	AI			RA		
		Y	U	V	Y	U	V
0	BD	-5.4	-3.3	-5.0	-3.0	-1.2	-1.6
	BDF1	-6.3	-6.3	-9.1	-2.8	-1.8	-4.2
	BDF2	-5.9	-6.0	-7.7	-2.0	-1.0	-3.0
	PK	-10.7	-7.4	-12.1	-4.7	1.7	-3.0
	TC	-2.4	-1.8	-2.8	-1.7	0.5	-0.7
	Avr.	-6.1	-4.9	-7.3	-2.8	-0.4	-2.5
2	BD	-6.0	-2.8	-8.4	-4.2	-1.0	-3.9
	BDF1	-7.8	-7.6	-14.8	-4.2	-2.5	-8.5
	BDF2	-7.3	-7.3	-13.2	-3.0	-1.0	-6.1
	PK	-11.1	-10.6	-16.8	-6.4	-3.3	-11.0
	TC	-3.1	-2.0	-4.7	-2.4	0.1	-2.1
	Avr.	-7.1	-6.1	-11.6	-4.1	-1.5	-6.3
Average		-6.6	-5.5	-9.5	-3.4	-1.0	-4.4

The average BD-rate performance (BD-rate$_{YUV}$) of different combinations of the proposed method and other existing gain-offset methods has been presented in Table 3 for different encoding configurations. Results in Table 3 indicate that the proposed method achieves a very similar coding efficiency to the other methods using gain-offset predictor. It should be noted that the proposed method can be realized with coding tools in SHVC standard, but the other method required changes in the processes and bit-stream syntax of SHVC standard.

According to Table 3, when the BL has 8-bit data, the proposed joint upsampling and bit shifting (JUsSh) achieves 0.2% and 0.8% coding efficiency on average for AI and RA configurations, respectively. But the JUsSh method does not result in BD-rate gain when BL is 10-bit, because in this case both the BL and EL have 10-bit data and bit shifting is not required at all.

For the EL pictures, coded as B-frame, two ILR pictures are available in the reference list. In one test case, referred to as WP(Adp), the calculated weight-offset parameter by LSE method is used for both ILR pictures. In the other test case, referred to as WP(Adp+Fix), the ILR picture in reference list0 uses the calculated weight-offset parameter, but the ILR picture in reference list1 uses fixed weight and offset factors of 1.0 and 0.0, respectively (i.e. weighted prediction is not applied for the inter-layer reference frame in list1). The benefit of this configuration is that for a specific block the encoder can decide which weight-offset parameter is more suitable in R-D sense. Based on the results in Table 3, WP(Adp+Fix) results in about 0.3% better coding efficiency than WP(Adp). But both the methods of WP(Adp) and WP(Adp+Fix) shows the same gain for AI configuration, because in this case, all the EL pictures have only one ILR picture as they are coded as P-frame.

In the case that the BL is 8-bit, the proposed WP method when used with JUsSh (i.e "JUsSh+WP(Adp) − JUsSh" in Table 3) shows a higher gain than when WP method is used with BitShift (i.e. "BitShift+WP(Adp)" in Table 3). This means that JUsSh help the WP brings a higher gain.

Table 3: Average R-D performance (BD-rate$_{YUV}$%) results for different methods and configurations (Anchor: BitShift)

Method	BL: 8-bit		BL: 10-bit	
	AI	RA	AI	RA
JUsSh	-0.8	-0.2	0.0	0.0
JUsSh+WP(Adp)	-6.8	-3.3	-6.5	-3.5
JUsSh+WP(Adp+Fix)	-6.8	-3.6	-6.5	-3.8
BitShift+WP(Adp)	-5.5	-2.9	-6.5	-3.5
BitShift+WP(Adp+Fix)	-5.5	-3.2	-6.5	-3.8
Weighted Prediction [5]	-6.2	-3.4	-6.6	-3.9
Piecewise gain-offset [6]	-4.6	-3.1	-4.3	-3.0
Region-based gain-offset [7]	-7.1	-2.8	-7.8	-3.2
3D LUT [9]	-16.1	-10.5	-16.1	-11.2

5 CONCLUSION

In this paper we proposed using the weighted prediction tool, already available in the SHVC standard, to map the ILR picture from the base layer color gamut to the enhancement layer color gamut. In addition, we also proposed a high-precision bit-depth mapping of the base layer to the enhancement layer. These methods enable coding of ultra-high definition content in BT.2020 color gamut in a scalable manner with high coding efficiency, where the base layer is of high definition resolution with BT.709 color gamut. The proposed methods utilize the tools already present in the SHVC standard; hence they will be useful in practical use-cases.

REFERENCES

[1] ITU-R Recommendation BT.2020 "Parameter values for UHDTV systems for production and international programme exchange" April 2012.

[2] ITU-R Recommendation BT.709 "Parameter values for the HDTV standards for production and international programme exchange" Dec. 2010.

[3] M. M. Hannuksela, K. Ugur, J. Lainema, D. Rusanovskyy, J. Chen, V. Seregin, Y. Wang, Y. Chen, L. Guo, M. Karczewicz, Y. Ye, J. Boyce, "Test Model for Scalable Extensions of High Efficiency Video Coding (HEVC)", Technical Report JCTVC-L0453, January 2013.

[4] L. Kerofsky, A. Segall, S.-H. Kim, and K. Misra, "Color Gamut Scalable Video Coding: New Results", Technical Report JCTVC-L0334, January 2013.

[5] X. Li, V. Seregin, J. Chen, K. Rapaka, Y. Chen, M. Karczewicz, "Non-SCE4: Weighted Prediction Based Color Gamut Scalability", Technical Report JCTVC-O0180, November 2013.

[6] C. Auyeung, "SCE4: Results of test 5.4-model1 on piecewise linear color space predictor", Technical Report JCTVC-O0196, November 2013.

[7] C. Auyeung, "Non-SCE4: Picture and region adaptive gain-offset prediction for color space scalability", Technical Report JCTVC-O0195, November 2013.

[8] P. Bordes, P. Andrivon, F. Hiron, P. Salmon, R. Boitard, "SCE4: Results on 5.3-test1 and 5.3-test2", Technical Report JCTVC-O0159, November 2013.

[9] Y. He, Y. Ye, J. Dong, "Non-SCE4/AHG14: Combined bit-depth and color gamut conversion with 3D LUT for SHVC color gamut scalability", Technical Report JCTVC-O0161, November 2013.

[10] L. Kerofsky, A. Segall, S.-H. Kim, "Color Gamut Scalable Video Coding", Proceedings of Data Compression Conference (DCC), p.p. 211-220, March 2013.

[11] M. M. Hannuksela, K. Ugur, J. Lainema, D. Rusanovskyy, J. Chen, V. Seregin, Y.-K. Wang, Y. Chen, L. Guo, M. Karczewicz, Y. Ye, J. Boyce, "Common Specification Text for Scalable and Multi-view Extensions", Technical Report JCTVC-L0452, January 2013.

[12] A. Aminlou, K. Ugur, M. M. Hannuksela, E. Alshina, A. Alshin, "SCE4: Test 5.1 results on bit-depth and color-gamut scalability", Technical Report JCTVC-O0161, November 2013.

[13] X. Li, J. Boyce, P. Onno and Y. Ye, "Common test condition and software reference configurations for the scalable test model," ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Document JCTVC-L1009, January 2013.

[14] G. Sullivan and J.R. Ohm, "Joint Call for Proposals on Scalable Video Coding Extensions of High Efficiency Video Coding (HEVC)," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 2012.

[15] G. Bjontegaard, "Improvements of the BD-PSNR model," ITU-T SG16 Q, vol. 6, pp. 35, 2008.