Cross-Plane Chroma Enhancement for SHVC Inter-Layer Prediction

Jie Dong, Yan Ye, Yuwen He
InterDigital Communications, Inc.
San Diego, USA
{jie.dong, yan.ye, yuwen.he}@interdigital.com

***Abstract*— This paper proposes a cross-plane chroma enhancement (CPCE) scheme to enhance the chroma planes of the inter layer reference (ILR) pictures for the Scalable extensions of HEVC (SHVC), the on-going scalable video coding project in JCT-VC. The CPCE scheme restores the blurred edges and textures in the chroma planes using the corresponding information from the luma plane. Experimental results under the SHVC common test conditions show that the average BD-rate reductions for the Cb and Cr chroma planes are as much as -7.5% and -8.5%, respectively, when compared with SHM-1.0.**

I. Introduction

The Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T/SG16/Q.6/VCEG and ISO/IEC/MPEG has been working on developing the new generation video coding standard and its extensions since April 2010. In January 2013, the first version of the standard [1] was finalized and submitted for approval, and was later approved by ITU-T as H.265 and by ISO/IEC as MPEG-H Part-2 High Efficiency Video Coding (HEVC). The first version of H.265/HEVC, used for single-layer video coding, employs the traditional hybrid coding (i.e., predictive coding plus transform coding) framework. H.265/HEVC doubles the coding efficiency of its predecessor H.264/AVC, due to a number of new technical designs, including more flexible coding unit size and transform unit partitions, more advanced in-loop filters, more accurate texture and motion prediction, and so on [2].

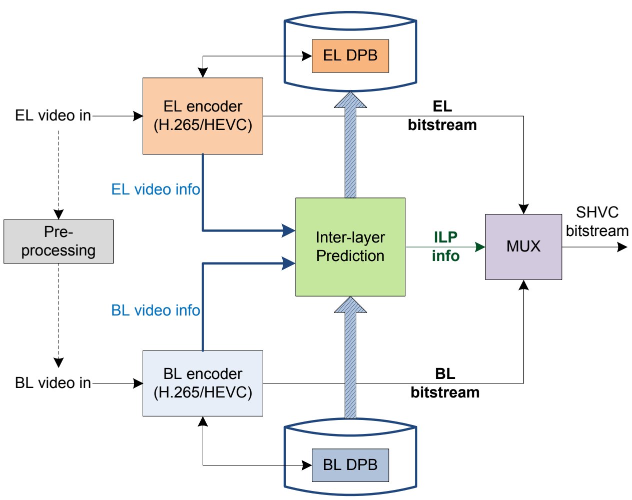

Fig. 1 Block diagram of two-layer SHVC encoder

The scalable extension of H.265/HEVC, named Scalable High-efficiency Video Coding (SHVC), is one of the on-going extensions projects under JCT-VC. SHVC will be a new scalable video coding standard built on the top of H.265/HEVC to provide the solution to improve the quality of experience for video applications running on devices with different capabilities over heterogeneous networks. An SHVC bitstream is coded at the highest representation of the source content, including quality and spatio-temporal resolution, and can be partitioned into layers, with each layer corresponding to a different representation of the source content. Among the layers, the base layer (BL) provides the lowest representation of the source signal, whereas higher representation of the video signal can be obtained by decoding each extra enhancement layer (EL). An SHVC BL bitstream can be coded using either single layer H.265/HEVC or single layer H.264/AVC codec; this is called standards scalability or hybrid codec scalability.

At the time of writing, SHVC [3] adopts multi-loop prediction structure, which means that an EL encoder needs the reconstruction of the inter prediction blocks in the reference/lower layers for prediction. Fig. 1 shows how the reconstructed BL picture is used for prediction in EL coding in an exemplary two-layer SHVC encoder. The BL input video is coded by H.265/HEVC and the reconstructed BL picture is processed by the Inter-layer Prediction (ILP) module before being inserted into the EL decoded picture buffer (DPB). The EL coding reuses H.265/HEVC encoder with the processed BL reconstructed picture as an additional reference picture, called the inter layer reference (ILR) picture. The ILP module performs at least the inverse processing of the pre-processing applied to the EL input video that generates the BL input video. For example, for spatial scalability, the pre-processing is downsampling, and the ILP module performs at least upsampling. Since the quality of the ILR pictures has a significant impact on the EL coding efficiency, additional enhancements can be performed on the ILR picture by the ILP module. The parameters used for such enhancements, shown as ILP info in Fig. 1, are multiplexed together with BL and EL bitstreams to form an SHVC bitstream.

In this paper, we propose a cross-plane chroma enhancement (CPCE) algorithm, as an additional ILP enhancement, to improve an ILR picture's chroma planes and thus the EL chroma coding efficiency. A color video signal is usually decomposed into Y, Cb, and Cr planes. Compared with the Y plane, the Cb and Cr planes usually have lower energy, meaning the pixel intensities in Cb and Cr planes have smaller variance. During compression and quantization, the edges and tex-

978-1-4799-0294-1/13/$31.00 ©2013 IEEE

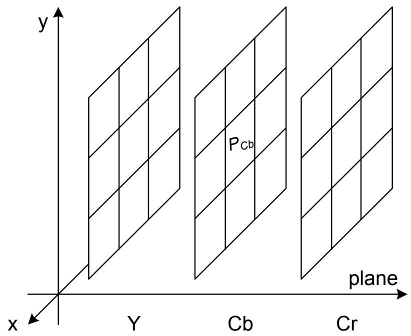

Fig. 2 3-D support region to filter a Cb pixel

tures in the chroma signals can be too delicate to be preserved. It is observed that with medium to high QPs, the chroma planes of the ILR pictures are severely blurred, whereas luma plane still has fair to good quality. We propose to restore the blurred edges and textures in the chroma planes using the corresponding information from the luma plane, because the three planes have high correlation in the edge structure. Specifically, for each chroma pixel in Cb and Cr planes, an appropriate offset calculated based on the values of corresponding 3×4 luma neighboring pixels is added. Compared with SHM-1.0 [4], the proposed CPCE algorithm achieves average {Y, Cb, Cr} BD-rate reduction of {-0.8%, -7.5%, -9.1%}, {-0.3%, -9.2%, -9.7%}, {-0.2%, -6.4%, -7.2%}, and {-0.2%, -6.9%, -7.9%} for coding configurations of all intra (AI), random access (RA), low delay P (LDP), and low delay B (LDB), respectively.

II. PROPOSED CROSS-PLANE CHROMA ENHANCEMENT

A. Framework

Without loss of generality, we discuss the CPCE scheme only for the Cb plane. The same scheme is applied for the Cr plane. To reduce the distortion between a degraded image and the corresponding original image, a well-known method is to use the two images as inputs to the least minimum mean square error (LMMSE) estimator, derive an FIR filter, and apply the filter on the degraded image. This method was widely used in adaptive interpolation filter (AIF) [5] and adaptive loop filter (ALF) [6] designs, and was found to improve the efficiency especially for luma coding. However, directly applying this method cannot efficiently reduce the distortion between the blurred Cb plane in an ILR picture and the original Cb plane in the corresponding EL picture, because the FIR filter derived by the LMMSE estimator has low-pass characteristics and cannot effectively restore the severely blurred edges in the Cb plane of an ILR picture.

In the proposed CPCE framework, to improve a Cb pixel, denoted as P_{Cb}, the commonly used 2-D non-separable filter is extended to 3-D, including 2-D in the space domain and 1-D in the plane domain, as shown in Fig. 2. By doing this, the edge information in the other two planes can be used as the guidance to improve the quality of the Cb plane. To derive the optimal 3-D filter, we also use the LMMSE estimator, which needs to take four images as the inputs: the Y, Cb, and Cr planes in an ILR picture and the original Cb plane in the corresponding EL picture to be coded, denoted as Y, Cb, Cr, and S_{Cb}, respectively.

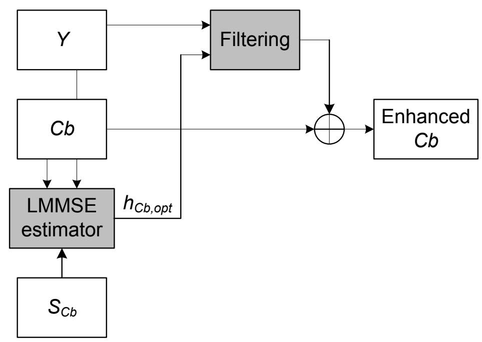

Fig. 3 Block diagram of the proposed CPCE scheme for Cb plane

The 3-D filter for the Cb plane is denoted as h_{Cb}, which consists of three 2-D filters applied on three planes, denoted as $h_{Cb}(Y)$, $h_{Cb}(Cb)$, $h_{Cb}(Cr)$, respectively. After applying h_{Cb} to the three planes of an ILR picture, the differential plane between the enhanced and the original Cb planes is calculated as in (1),

$$Diff|_{h_{Cb}} = Y \otimes h_{Cb}(Y) + Cb \otimes h_{Cb}(Cb) + Cr \otimes h_{Cb}(Cr) - S_{Cb} \quad (1)$$

where the operator ⊗ means 2-D convolution. The optimal h_{Cb} minimizes the energy of $Diff|_{h_{Cb}}$, as shown in (2),

$$h_{Cb,opt} = \arg\min_{h_{Cb}} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \left(Diff|_{h_{Cb}}(x,y)\right)^2 \quad (2)$$

where W and H represent the width and height of $Diff|_{h_{Cb}}$.

Transmitting two real-valued 3-D filters for Cb and Cr planes with sufficient precision on a picture basis is usually unaffordable. Furthermore, to enhance a chroma pixel, (3×M×N-1) additions and 3×M×N multiplications are required, of which the complexity is too high. Therefore, the ideal CPCE framework needs to be simplified.

We studied the optimal 3-D filter $h_{Cb,opt}$, trained by (2), and had the following observations. First, $h_{Cb}(Cr)$ has little contribution to enhance a Cb pixel, because Cb and Cr planes are differential signals for blue and red, respectively, and have little correlation. Second, when $h_{Cb}(Cb)$, which has low-pass characteristics, is replaced by an identity filter (i.e., a 1-tap filter with the coefficient equal to 1.0), the performance degradation is negligible. Third, CPCE's performance cannot be further improved, if $h_{Cb}(Y)$ is larger than a sufficient size. Base on these observations, we made three simplifications: removing $h_{Cb}(Cr)$ from h_{Cb}, reducing $h_{Cb}(Cb)$ to an identity filter, and finding a sufficient size for $h_{Cb}(Y)$. As a result, $h_{Cb,opt}$ is reduced to a 2-D filter applied only on the Y plane of the ILR picture, and (1) is re-written as in (3).

$$Diff|_{h_{Cb}} = Y \otimes h_{Cb}(Y) + Cb - S_{Cb} \quad (3)$$

Fig. 3 shows the block diagram of the proposed CPCE scheme for the Cb plane. The LMMSE estimator takes the Y and Cb planes of the ILR picture and the original Cb plane as inputs, and derives the optimal filter $h_{Cb,opt}(Y)$ based on (2) and (3). Since $h_{Cb,opt}(Y)$ is a high-pass filter, the output of the filtering process contains Y plane's edge and texture information, which is added to the blurred Cb plane for edge and texture enhancement and restoration. The following subsections introduce the implementation details of the proposed CPCE scheme in SHVC.

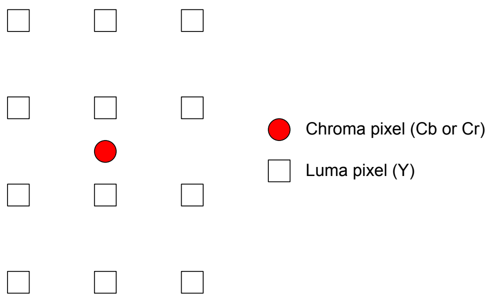

Fig. 4 Relative Luma and Chroma pixel positions with 4:2:0 color subsampling format.

B. High-Pass Filter Training

At the time of writing, SHVC considers scalable coding of YUV4:2:0 input signals. For YUV 4:2:0 input signals, we propose the size of $h_{Cb}(Y)$ to be 3×4, where one chroma sample corresponds to a neighborhood of 3×4 luma samples, as shown in Fig. 4.

At the encoder side, to code each EL picture, two optimal high-pass filters, $h_{Cb,opt}$ and $h_{Cr,opt}$, are analytically derived to enhance the Cb and Cr planes of the ILR picture, respectively, using the LMMSE estimator. The objective functions, formulated by (4) and (5), find $h_{Cb,opt}$ and $h_{Cr,opt}$, such that MSE between the original EL chroma planes and the enhanced chroma planes in the ILR picture is minimized.

$$h_{Cb,opt} = \arg\min_{h_{Cb}} E\left[\left(\sum_{j=-2}^{1}\sum_{i=-1}^{1} h_{Cb}(i,j)Y(2x-i,2y-j) + Cb(x,y) - S_{Cb}(x,y)\right)^2\right] \quad (4)$$

$$h_{Cr,opt} = \arg\min_{h_{Cr}} E\left[\left(\sum_{j=-2}^{1}\sum_{i=-1}^{1} h_{Cr}(i,j)Y(2x-i,2y-j) + Cr(x,y) - S_{Cr}(x,y)\right)^2\right] \quad (5)$$

C. Quantization and Signaling of the High-Pass Filters

The optimal high-pass filters $h_{Cb,opt}$ and $h_{Cr,opt}$, obtained in Section II.B, have real-valued coefficients, and therefore need to be quantized before transmission.

Without loss of generality, here we only explain how to quantize and signal $h_{Cb,opt}$. Though the summation of all the coefficients in $h_{Cb,opt}$ is always zero, the magnitudes of individual coefficients vary significantly with the content of the picture. $h_{Cb,opt}$ for pictures full of sharp edges has larger magnitude coefficients, and vice versa. Therefore, to quantize a coefficient $h_{Cb,opt}(i,j)$ (i = -1, 0, 1, j = -2, -1, 0, 1), we use a 16-level uniform quantizer, of which the quantization step size, denoted as Q_{step}, is adapted at the picture-level, as shown in (6),

$$h_{Cb,opt}(i,j)/Q_{step} = f_{Cb}(i,j) \quad (6)$$

where $f_{Cb}(i,j)$ denotes the quantized coefficient and is an integer with the dynamic range from -8 to 7 (4-bit representation). In practice, the division in (6) is approximated by multiplication and right shifting, and (6) is re-written as in (7),

$$h_{Cb,opt}(i,j) = f_{Cb}(i,j) \times \frac{Q_{Cb}}{2^{N_{Cb}}} = (f_{Cb}(i,j) \times Q_{Cb}) \gg N_{Cb} \quad (7)$$

where Q_{Cb} and N_{Cb} are the two quantization parameters transmitted for each picture.

As an approximation to the optimal high-pass filter, f_{Cb} also has the zero-summation constraint. Therefore, only 11 filter coefficients are explicitly coded, and the remaining one can be implicitly derived.

Altogether, for each of the two chroma planes, 61 bits are needed to signal the side information, when CPCE is enabled, including:

1. 1-bit flag, indicating whether CPCE is enabled for the current chroma plane,
2. 11 filter coefficients, each of which with 4-bit precision,
3. Q_{Cb} represented by 11 bits (10 bits for magnitude and 1 bit for sign),
4. N_{Cb} represented by 5 bits.

D. Process of Cross-Plane Chroma Enhancement

As in Section II.C, only the enhancement for the Cb plane of the ILR picture is discussed as an example. To enhance a Cb pixel located at position (x,y) (see the red circle in Fig. 4 as an example), high-pass filter f_{Cb} is first applied to the corresponding 3×4 luma neighboring pixels to generate the intermediate result $z(x,y)$, as shown in (8).

$$z(x,y) = \sum_{j=-2}^{1}\sum_{i=-1}^{1} f_{Cb}(i,j)Y(2x-i,2y-j) \quad (8)$$

Then, $z(x,y)$ is scaled by $Q_{Cb}/2^{N_{Cb}}$, as in (7), to the normal range, denoted as $o(x,y)$. The integer realization of this scaling step is shown below,

$$o(x,y) = \mathrm{Sign}(Q_{Cb}z(x,y))\left((\mathrm{Abs}(Q_{Cb}z(x,y)) + 2^{N_{Cb}-1}) \gg N_{Cb}\right) \quad (9)$$

where function Abs(x) returns the absolution value of x, and Sign(x) returns 1, 0, and -1, when x is positive, equal to 0, and negative, respectively.

The offset value $o(x,y)$ is then added to $Cb(x,y)$ to obtained the enhanced Cb pixel, denoted as $Cb_{enh}(x,y)$,

$$Cb_{enh}(x,y) = Cb(x,y) + o(x,y). \quad (10)$$

III. EXPERIMENTAL RESULTS

The proposed CPCE scheme was integrated into SHVC's reference software SHM-1.0 [4] and tested under the common test conditions [7]. Seven video sequences: two for Class A 2560×1600 and five for Class B 1080*p*, are coded with spatial scalability (2x and 1.5x) and SNR scalability. Each sequence is coded by four types of coding structures: all intra (AI), random access (RA), low delay P (LDP), and low delay B (LDB). For spatial scalability, the BL sequences are coded at four *QP*s (22, 26, 30, and 34) and the EL *QP*s are set to BL *QP* + Δ*QP*, where Δ*QP* is equal to 0 or 2. For SNR scalability, the BL *QP*s are (26, 30, 34, and 38) and the Δ*QP* for EL is -6 and -4. With a set of four *QP*s, four data points, each consisting of the PSNR of the coded EL sequence and the total bits of the BL and EL bitstreams, form an operational R-D curve over a wide

TABLE I BD-rate reduction compared with SHM-1.0 (%)

	AI 2x			AI 1.5x		
	Y	Cb	Cr	Y	Cb	Cr
Class A	-0.9	-7.8	-6.2			
Class B	-0.8	-6.4	-8.5	-0.7	-8.2	-10.4
Average	-0.8	-6.8	-7.8	-0.7	-8.2	-10.4

	RA 2x			RA 1.5x			RA SNR		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A	-0.5	-11.2	-7.7				-0.4	-9.0	-5.4
Class B	-0.3	-8.0	-9.7	-0.3	-10.5	-12.1	-0.3	-7.7	-8.6
Average	-0.4	-9.0	-9.1	-0.3	-10.5	-12.1	-0.3	-8.0	-7.7

	LDP 2x			LDP 1.5x			LDP SNR		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A	-0.2	-9.3	-5.2				-0.3	-7.3	-4.4
Class B	-0.2	-4.6	-5.8	-0.2	-7.6	-10.1	-0.2	-5.2	-6.1
Average	-0.2	-5.9	-5.6	-0.2	-7.6	-10.1	-0.2	-5.8	-5.6

	LDB 2x			LDB 1.5x			LDB SNR		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A	-0.2	-9.7	-5.8				-0.3	-8.0	-5.0
Class B	-0.2	-5.1	-6.7	-0.2	-8.2	-10.8	-0.2	-5.6	-6.9
Average	-0.2	-6.4	-6.4	-0.2	-8.2	-10.8	-0.2	-6.3	-6.3

range of bit-rates, and the average bit-rate reduction compared with SHM-1.0, known as BD-rate [8], is calculated and used to measure the R-D performance of a proposed coding tool.

TABLE I shows the percentage of average BD-rate reduction provided by the CPCE scheme, compared with SHM-1.0. As can be seen, significant bit-rate reductions for Cb and Cr, ranging from 5.6% to 12.1%, are achieved across all the test conditions. 1.5x spatial scalability achieves higher gains than 2x and SNR scalability, because inter layer prediction is more frequently used in 1.5x spatial scalability and thus the quality of ILR picture has a more significant impact on the scalable coding efficiency. Since the proposed scheme reduces the overall bit-rate, BD-rate reductions for the Y plane, ranging from 0.2% to 0.8%, are also observed. Fig. 5 and Fig. 6 show two exemplary operational R-D curves with 0.5 dB or higher improvement by the CPCE scheme. As shown, the efficiency of EL chroma coding is significantly improved by the CPCE scheme and even outperforms the single-layer HEVC chroma coding (see the green lines).

The scalable coding performance without and with the proposed CPCE are compared with simulcast, respectively, in Table II. Without CPCE (i.e., using the original SHM-1.0), the chroma scalable gain significantly lags the luma scalable gain in the RA, LDP, and LDB cases. By enabling CPCE for SHM-1.0, the performance gap between luma and chroma is eliminated, proving that the ILR pictures with CPCE can provide comparable benefits for luma and chroma coding in the EL.

IV. Conclusion

This paper proposes the CPCE scheme to enhance the chroma planes of the ILR pictures for SHVC. The CPCE scheme restores the blurred edges and textures in the chroma planes using the corresponding edge information from the luma plane. Experimental results show that, compared with SHM-1.0, the proposed scheme achieves average BD-rate reductions of -7.5% and -8.5% for the Cb and Cr planes, respectively.

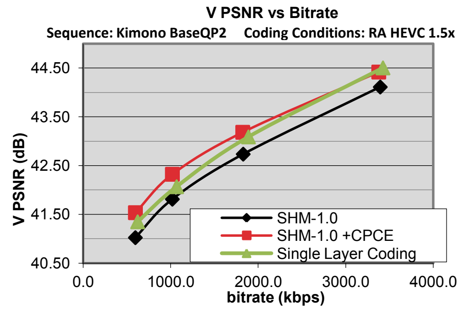

Fig. 5 Cr plane of *Kimono* (1.5x spatial scalability, RA, Δ*QP*=2)

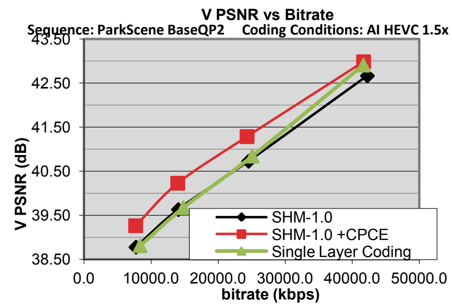

Fig. 6 Cr plane of *ParkScene* (1.5x spatial scalability, AI, Δ*QP*=2)

TABLE II Scalable coding w/o and w/ CPCE vs. simulcast

	SHM-1.0			SHM-1.0 + CPCE		
	Y	Cb	Cr	Y	Cb	Cr
AI	-26.8%	-26.4%	-26.7%	-27.4%	-31.6%	-32.7%
RA	-20.6%	-12.2%	-10.4%	-20.9%	-20.1%	-18.5%
LDP	-15.4%	-8.1%	-6.2%	-15.6%	-13.9%	-12.5%
LDB	-14.7%	-8.2%	-6.3%	-14.9%	-14.4%	-13.2%

References

[1] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, T. Wiegand, "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Consent)," JCT-VC document, JCTVC-L1003, Geneva, Switzerland, Jan. 2013.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans.Circuits Syst. Video Technol.*, Vol.22, No.12, pp.1649-1668, Dec. 2012.

[3] J. Chen, J. Boyce, Y. Ye, M. M. Hannuksela, "SHVC Working Draft 2," JCT-VC document, JCTVC-M1008, Incheon, South Korea, Apri. 2013.

[4] https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/tags/SHM-1.0/.

[5] Y. Vatis, J. Ostermann, "Adaptive interpolation filter for H.264/AVC," *IEEE Trans.Circuits Syst. Video Technol.*, Vol.19, No. 2, pp.179–192, Feb. 2009.

[6] C.-Y. Chen, *et al.*, "The adaptive loop filter in the HEVC standard," 2012 Applications of Digital Image Processing XXXV, San Diego, USA, Aug. 12-16, 2012.

[7] X. Li, J. Boyce, P. Onno, Y. Ye, "Common SHM test conditions and software reference configurations," JCT-VC document, JCTVC-L1009, Geneva, Switzerland, Jan. 2013.

[8] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," document VCEG-M33, ITU-T SG16/Q6, Apr. 2001.