IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 October 2023, accepted 16 November 2023, date of publication 20 November 2023,
date of current version 28 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335043

RESEARCH ARTICLE

Adaptive Interpolation Filter Using Correlation for Inter Prediction

MIN-KYEONG CHOI AND YUNG-LYUL LEE, (Senior Member, IEEE)
Department of Computer Engineering, Sejong University, Seoul 05006, South Korea

Corresponding author: Yung-Lyul Lee (yllee@sejong.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government through MSIT under Grant RS-2023-00219051.

ABSTRACT In video coding, interpolation filters play an important role in motion-compensated prediction. In Versatile Video Coding (VVC), adaptive motion vector resolution (AMVR) that makes use of fractional samples by 8-tap Discrete Cosine Transform-based interpolation filter (DCT-IF), 7-tap DCT-IF, or 6-tap Smoothing interpolation filter (SIF) was adopted. This paper proposes an adaptive interpolation filter to improve motion-compensated prediction. A 12-tap DCT-IF is designed to improve the filter response in the high frequency range and applied by a proposed filter selection method using correlation that considers the frequency characteristics of reference block found from motion estimation. Experimental results show that the proposed method achieved the overall Bjøntegaard Delta (BD)-rate gains of -0.27%, -0.12%, and -0.01% for Y, Cb, and Cr components, respectively, under the Random Access (RA) configuration compared with VVC.

INDEX TERMS Video coding, versatile video coding (VVC), motion-compensated prediction, interpolation filter, inter prediction, adaptive motion vector resolution (AMVR).

I. INTRODUCTION

In order to jointly develop the next generation video coding standard, the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) organized the Joint Video Exploration Team (JVET) in October 2015, and Versatile Video Coding (VVC/H.266) standardization [1] was completed in July 2020. VVC achieved a bit rate reduction of about 40% [2] compared to High Efficiency Video Coding (HEVC/H.265) and VVC inherits the framework of block-based hybrid coding like Advanced Video Coding (AVC/H.264) [3], [4] and HEVC [5], [6].

Since sequential pictures in a video signal have high temporal redundancy, inter prediction, which aims to reduce temporal redundancy, significantly contributes to the capability of video compression and is essential to the hybrid video coding scheme.

In VVC [1], [7], a lot of dominant coding tools have been adopted to improve inter prediction. The coding tools are divided into three categories based on the inter prediction scheme: advanced motion vector prediction (AMVP), motion vector merge, and others. In normal AMVP mode, when constructing the candidate list with two candidates, spatial motion vector predictor (MVP) from spatial neighbour Coding Units (CUs) and temporal MVP from temporal neighbor CUs are considered as candidates first like HEVC. If the number of candidates does not construct two candidates, history-based motion vector prediction (HMVP) that the motion information of a previously coded block is stored in a table and used as MVP for the current CU [8], [9] is used in VVC. If the number of candidates does not still construct two candidates, zero motion vectors are used as MVP.

In addition, symmetric motion vector difference (SMVD) [10], [11] that the motion information is derived using the assumption of linear motion in bi-prediction mode was adopted in AMVP. In contrast to AVC and HEVC, which employ a fixed motion vector resolution, adaptive motion vector resolution (AMVR) [12] according to motion vector difference at quarter luma sample, half luma sample, integer luma sample, or four luma sample unit was adopted. When generating fractional-samples, AMVR uses 8-tap Discrete Cosine Transform-based interpolation filter (DCT-IF), 7-tap

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

DCT-IF, or 6-tap smoothing interpolation filter (SIF) [13], [14].

In motion-compensated prediction, accurate fractional samples can better capture continuous motion. Therefore, interpolation filters have changed with each new standard development. In AVC [4], when the motion vector points to a fractional sample position for luma samples, the prediction signal consisting of corresponding samples is obtained using interpolation to generate samples of fractional positions. The prediction values at half-sample positions are obtained by applying a one-dimensional 6-tap FIR filter. Prediction values at quarter-sample positions are generated by averaging samples at integer- and half-sample positions. The fractional sample interpolation for luma samples in HEVC [6] uses a DCT-based 8-tap filter for the half sample positions and a DCT-based 7-tap filter for the quarter sample positions.

Several studies have also been done to enhance performance of interpolation filters for motion-compensated prediction. In [15], Lv et al. proposed a resolution-adaptive interpolation filter in which 4-tap filter was used high-definition video, and 6-tap and 10-tap filter were used to relatively small resolution videos. The resolution-adaptive filter achieved bit-saving compared to HEVC, but it is not common in block-based video coding to use different interpolation filters according to video resolutions. Kim and Lee [16] proposed 11-tap Discrete Sine Transform-based interpolation filter (DST-IF) and 12-tap DST-IF instead of 7-tap DCT-IF and 8-tap DCT-IF to use filters that highlight high frequency components in HEVC. The DST-IFs improve coding performance in low resolution video, but in high-resolution video, the coding performance was decreased. Deep learning-based methods have also been studied to improve interpolation filter performance. In [17] and [18], Convolutional Neural Network (CNN)-based interpolation filters were used by Zhang et al. and Muran et al. in HEVC and VVC respectively. The limitations of DCT-based interpolation filters, which might not be able to adapt to various video contents, were made up for by the CNN-based interpolation filters. However, in [17], the complexity of this method was high due to the very deep network structure. Despite the fact that [18] is proposed for low-complexity inter prediction, its encoding and decoding times were increased by 7.6 and 2.6 times, respectively, compared with VVC. To improve coding performance for chroma components as well as luma components, long tap DCT-IF was proposed by Xie et al. [19], [20] to replace existing 4-tap DCT-IF. The 6-tap DCT-IF improved the filtering especially in the high frequency range and the method achieved the overall BD-rate gains of −0.05%, −1.23%, and -1.25% for Y, Cb, and Cr components, respectively, under the Random Access (RA) configuration compared with Enhanced Compression Model (ECM) [21] beyond VVC. The method was adopted in ECM due to chroma coding gain.

In this paper, an adaptive interpolation filter according to the characteristics of block are presented. First, to generate more accurate fractional sample for motion estimated-reference blocks with high frequency characteristics, a 12-tap DCT-IF with better high frequency characteristics than the VVC existing filter is designed. Second, an interpolation filter selection method is proposed in the consideration of frequency characteristics on each reference block by calculating normalized correlation. The experimental results verify that the 12-tap DCT-IF and the filter selection method achieve an improvement of coding performance compared with VVC.

The rest of this paper is organized as follows. In Section II, the proposed method based on 12-tap DCT-IF and interpolation filter selection is described. Next, experimental results are presented in Section III, and Section IV concludes the paper.

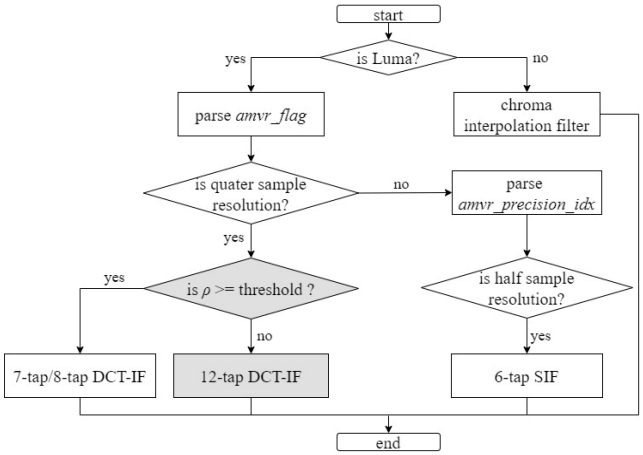

FIGURE 1. Flow chart of the proposed method added in the VVC method. The proposed gray parts are added to VVC and ρ is correlation.

II. PROPOSED METHOD

A 12-tap DCT-based interpolation filter for quarter sample positions and an interpolation filter selection method based on correlation are proposed. Fig. 1 shows a flow chart of the proposed method, where integer luma sample and four luma sample units in AMVR are not considered because interpolation is not necessary in motion-compensated prediction. The gray parts in Fig. 1 represent the proposed method added to the VVC method and details are followed in Section A, B, C and D.

A. INTERPOLATION FILTER IN VVC

If the two gray parts are removed from Fig.1, the flow chart becomes the VVC method in AMVR.

In VVC, quarter samples (1/4, 2/4, 3/4) are generated by 7-tap/8-tap DCT-IF in case of quarter sample resolution and half samples are generated 6-tap SIF [13], [14] when *amvr_precision_idx* indicates half sample resolution.

The DCT-based interpolation filter coefficients are derived from DCT-II in (1) and IDCT-II (Inverse DCT-II) in (2), where $X(k)$ is the DCT-II coefficients and $x(n)$ is the IDCT-II

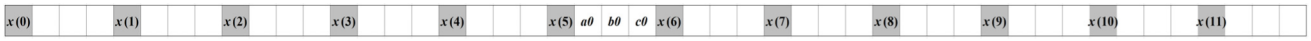

FIGURE 2. An example of the integer pixel positions and fractional pixel positions in horizontal direction. Capital letters represent integer sample positions, and small letters represent fractional sample positions.

TABLE 1. Interpolation filter coefficients in VVC.

Index i	0	1	2	3	4	5	6	7
1/4-pixel filter coefficients[i] at i=3.25	-1	4	-10	58	17	-5	1	0
	7-tap DCT-IF							
1/2-pixel filter coefficients[i] at i=3.5	-1	4	-11	40	40	-11	4	-1
	8-tap DCT-IF							
1/2-pixel filter coefficients[i] at i=3.5	0	3	9	20	20	9	3	0
	6-tap SIF							

coefficients.

$$X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} c_k x(n) \cos\frac{(n+1/2)\pi k}{N} \quad (1)$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} c_k X(k) \cos\frac{(n+1/2)\pi k}{N} \quad (2)$$

where

$$c_k = \begin{cases} \frac{1}{\sqrt{2}}, k = 0 \\ 1, k = 1, 2, \ldots, N-1 \end{cases}$$

Equation (1) is substituted for (2) to produce the following DCT-IF equation:

$$x(n) = \frac{2}{N} \sum_{m=0}^{N-1} x(m) \sum_{k=0}^{N-1} c_k^2 \cos\frac{(m+1/2)\pi k}{N} \cos\frac{(n+1/2)\pi k}{N} \quad (3)$$

For example, the half pixel position, where $n = 3 + 1/2$, in the 8-tap DCT-IF is derived in (3) as a linear combination of the cosine coefficients and $x(m)$, $m = 0, 1, \ldots, 7$.

The 6-tap SIF is derived from Gaussian window as follows:

$$w[n] = exp\left(-\frac{1}{2}\left(\frac{n - \frac{N-1}{2}}{\frac{\sigma(N-1)}{2}}\right)^2\right) \quad (4)$$

where

$$\sigma = 0.3$$

where N represents the width, in samples, of a symmetrical window function $w[n]$, $n = 0, 1, \ldots, 5$ [14].

Table 1 shows the interpolation filter coefficients at 1/4 and 1/2 positions in VVC, where the Index i represents the integer sample positions. The 3/4-pixel filter coefficient order is the reverse of the 1/4-pixel filter coefficient order.

B. 12-TAP DISCRETE COSINE TRANSFORM-BASED INTERPOLATION FILTER

In the similar way to 8-tap DCT-IF and 7-tap DCT-IF, the 12-tap DCT-IF coefficients are derived from DCT-IF equation in (3) of Section A, where N is equal to 12. For example, the half pixel position, where $n = 5 + 1/2$, in the 12-tap DCT-IF is derived in (3) as a linear combination of the cosine coefficients and $x(m)$, $m = 0, 1, \ldots, 11$ as follows:

$$x(5.5) = \frac{1}{6}(x(0) \cdot \frac{1}{2} + x(1) \cdot \cos\frac{1.5}{12}\pi \cdot \cos\frac{6}{12}\pi + \ldots + x(11) \cdot \cos\frac{126.5}{12}\pi \cdot \cos\frac{66}{12}\pi) \quad (5)$$

Similarly, the quarter pixel position, where $n = 5 + 1/4$, in the 12-tap DCT-IF is derived as a linear combination of cosine coefficients and $x(m)$, $m = 0, 1, \ldots, 11$ as follows

$$x(5.25) = \frac{1}{6}(x(0) \cdot \frac{1}{2} + x(1) \cdot \cos\frac{1.5}{12}\pi \cdot \cos\frac{5.75}{12}\pi + \ldots + x(11) \cdot \cos\frac{126.5}{12}\pi \cdot \cos\frac{63.25}{12}\pi) \quad (6)$$

The Table 2 shows the proposed 12-tap DCT-IF coefficients derived by scaling with 128. Index i in Table 2 represents the integer sample positions where filtering is applied to generate fractional samples. The 3/4-pixel filter coefficient order is the reverse of the 1/4-pixel filter coefficient order.

Fig. 2 shows an example of the integer pixel positions and fractional pixel positions in horizontal direction. $x(0)$ to $x(11)$ mean integer sample positions corresponding to $x(m)$ in (5) or (6), and small letters ($a0$, $b0$, and $c0$) mean fractional samples to interpolate. Equations (7) and (8) are the examples of generating the quarter sample and half sample by using the coefficients in Table 2.

$$a0 = (-1 \cdot x(0) + 3 \cdot x(1) - 6 \cdot x(2) + 11 \cdot x(3) - 22 \cdot x(4) + 115 \cdot x(5) + 38 \cdot x(6) - 16 \cdot x(7) + 9 \cdot x(8) - 5 \cdot x(9) + 3 \cdot x(10) - 1 \cdot x(11)) \gg 7 \quad (7)$$

$$b0 = (-1 \cdot x(0) + 4 \cdot x(1) - 8 \cdot x(2) + 14 \cdot x(3) - 26 \cdot x(4) + 81 \cdot x(5) + 81 \cdot x(6) - 26 \cdot x(7) + 14 \cdot x(8) - 8 \cdot x(9) + 4 \cdot x(10) - 1 \cdot x(11)) \gg 7 \quad (8)$$

Horizontal/vertical padding on the reference picture boundary is applied as in the VVC test Model VTM-16.0 if filter length is extended outside the reference picture boundary.

C. COMPARISON OF INTERPOLATION FILTERS

The magnitude responses of the half pixel position of interpolation filters are shown by the graphs in Fig. 3, where the x-axis represents the normalized frequency in 0 to 1,

TABLE 2. The 12-TAP DCT-based interpolation filter coefficients.

Index i	0	1	2	3	4	5	6	7	8	9	10	11
1/4-pixel filter[i] at i = 5.25	-1	3	-6	11	-22	115	38	-16	9	-5	3	-1
1/2-pixel filter[i] at i = 5.5	-1	4	-8	14	-26	81	81	-26	14	-8	4	-1

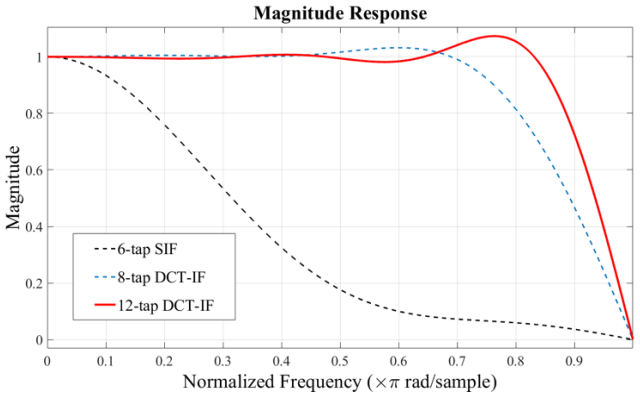

FIGURE 3. Magnitude responses at half pixel position of interpolation filters. The dotted lines represent magnitude responses of the existing interpolation filters in VVC, and the solid red line represents magnitude response of the proposed 12-tap DCT-based interpolation filter.

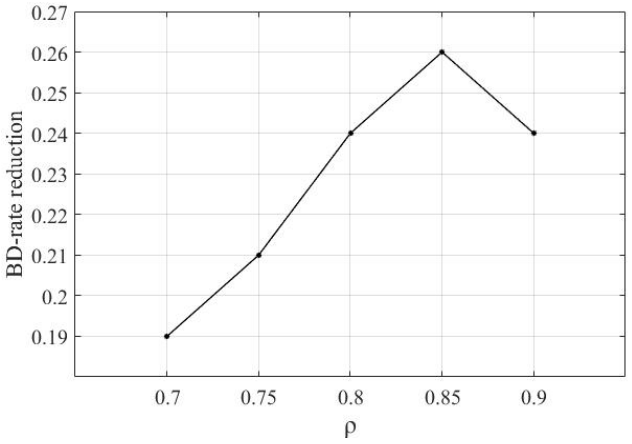

FIGURE 4. Performance of the proposed method with different ρ values.

where 1 corresponds to π radian and the y-axis represents the magnitude response. The 6-tap SIF can be seen to have a strong low-pass characteristics, which is useful for reducing high frequency components. On the other hand, the 8-tap DCT-IF and the 12-tap DCT-IF have a wide passband that passes not only low frequency components but also high frequency components. Especially, the wider passband of the proposed 12-tap DCT-IF represented in solid red line allows for the preservation of the input signal's strong high frequency information.

D. CORRELATION-BASED INTERPOLATION FILTER SELECTION METHOD

For more accurate fractional sample generation, an interpolation filter selection method was developed considering the frequency characteristics of the reference block using correlation. The correlation in (9) is computed from the motion-estimated reference block, where N is the height of the reference block, M is the width of the reference block, $x_(i,j)$ is the sample at (i, j) position of the reference block, and $\bar{x}$ is the mean value of the reference block.

$$\rho = \frac{\sum_{i=0}^{N-2}\sum_{j=0}^{M-2}(x_{(i,j)}-\bar{x})(x_{(i+1,j+1)}-\bar{x})}{\sqrt{\sum_{i=0}^{N-2}\sum_{j=0}^{M-2}(x_{(i,j)}-\bar{x})^2}\sqrt{\sum_{i=0}^{N-2}\sum_{j=0}^{M-2}(x_{(i+1,j+1)}-\bar{x})^2}} \quad (9)$$

High ρ value means low frequency characteristics in the reference block, while low ρ value means high frequency characteristics. Therefore, the proposed 12-tap DCT-IF is selected to interpolate high frequency components in fractional motion compensation. If the correlation value of the reference block is less than the predefined threshold, then the 12-tap DCT-IF is used, otherwise the existing VVC interpolation filters are used as shown in Fig. 1.

The ρ threshold was determined through experiments. To select the optimal threshold, the JVET common test condition (CTC) [22] with RA configuration, where 32 frames in classes A1 and A2 sequences and 64 frames in classes B, C, D sequences were used. The threshold ρ was set to 0.85, because the Bjøntegaard Delta (BD)-rate [23], [24] is most effectively reduced when the threshold is 0.85 in Fig. 4. From Figure 4, the fixed $\rho = 0.85$ was set for all test sequences. That is, if $\rho \leq 0.85$, then the 12-tap DCT-IF is used, otherwise, the existing VVC interpolation filters are used.

III. EXPERIMANTAL RESULTS

The proposed method was implemented on the VVC test model VTM-16.0 [25] under the JVET CTC [22] with Random Access (RA) configurations. The Single Instruction Multiple Data (SIMD) operation was used to apply the proposed interpolation filter for fractional pixel generation. The sequences of classes A1, A2, B, C, and D were tested with Quantization Parameter (QP) values of 22, 27, 32,

TABLE 3. Information on test video sequences for each class.

Class	Sequence name	Picture size	Picture count	Picture rate	Bit depth
A1	Tango2	3840×2160	294	60	10
	FoodMarket4	3840×2160	300	60	10
	Campfire	3840×2160	300	30	10
A2	CatRobot1	3840×2160	300	60	10
	DaylightRoad2	3840×2160	300	60	10
	ParkRunning3	3840×2160	300	50	10
B	MarketPlace	1920×1080	600	60	10
	RitualDance	1920×1080	600	60	10
	Cactus	1920×1080	500	50	8
	BasketballDrive	1920×1080	500	50	8
	BQTerrace	1920×1080	600	60	8
C	RaceHorses	832×480	300	30	8
	BQMall	832×480	600	60	8
	PartyScene	832×480	500	50	8
	BasketballDrill	832×480	500	50	8
D	RaceHorses	416×240	300	30	8
	BQSquare	416×240	600	60	8
	BlowingBubbles	416×240	500	50	8
	BasketballPass	416×240	500	50	8

and 37, respectively. Table 3 shows the sequence name, picture size, picture count, picture rate, and bit depth for the JVET CTC video sequences, and Table 4 shows the experimental results of the proposed method compared with the VTM-16.0 anchor.

The bit rate reduction ratio over the anchor in Table 4 is shown by the BD-rate [23], [24], in which the negative sign of BD-rate means the bit-saving of the proposed method compared with the VVC anchor in the same PSNR reference. The computational complexity of the proposed method is calculated as the relative computational complexity ratio of the proposed method to the VVC anchor, as shown in (10). All tests were performed on homogeneous computers of Intel Xeon E5-1620 v2 (3.7GHz) with 64 cores and 32GB RAM. In table 4, the separate "EncT" and "DecT" that mean the computational complexity increases compared with VTM-16.0 in the encoder and decoder for all test sequences, respectively, are computed by (10).

$$Time\_Inc = \frac{Time_{proposed\ method}}{Time_{anchor}} \times 100(\%) \tag{10}$$

As shown in Table 4, the proposed method achieves the overall BD-rate gains of -0.27%, −0.12% and −0.01% for Y, Cb, and Cr components, respectively, with the average encoding time of 120% and the average decoding time of 130%, respectively, compared with the VVC anchor. −0.27% luma (Y) component coding gain based on correlation calculation in VTM-16.0 is meaningful. As a reference, the percentage of CU blocks selected in the proposed method are shown in Table 5, where W represents the width of CU, and H represents the height of CU.

−1.67% and −1.13% BD-gains in the BQSquare and Blowing Bubbles sequences of class D are obtained, and −1.13% BD-gain in the PartyScene sequence of class C is obtained. But there are almost no BD-gains in the classes A1, A2, and B. Since high-resolution images have high correlation between pixels, the application of 12-tap DCT-IF with high-frequency characteristics does not improve performance. To improve the performance of A1 and A2 class sequences, the ρ value was changed independently only for A1 and A2 sequences, but there was no performance improvement due to the high correlation between pixels.

The running-times in the encoder and decoder are increased, because the 12-tap DCT-IF is applied if 2-D correlation ρ calculated for each reference block is less than 0.85. Therefore, we carried out additional experiments to reduce the complexity of ρ computation. Equation (11) is computed from the reference block as well, where $\rho_{1st\ row}$ represents the correlation value on the first row of the reference block, and $\rho_{1st\ col}$ represents the correlation value on the first column of the reference block.

$$\rho' = (\rho_{1st\ row} + \rho_{1st\ col}) \gg 1 \tag{11}$$

where

$$\rho_{1strow} = \frac{\sum_{i=0}^{M-2}(x_{(1,i)} - \overline{x_{1strow}})(x_{(1,i+1)} - \overline{x_{1strow}})}{\sqrt{\sum_{i=0}^{M-2}(x_{(1,i)} - \overline{x_{1strow}})^2}\sqrt{\sum_{i=0}^{M-2}(x_{1,i+1)} - \overline{x_{1strow}})^2}}$$

$$\rho_{1stcol} = \frac{\sum_{i=0}^{N-2}(x_{(i,1)} - \overline{x_{1stcol}})(x_{(i+1,1)} - \overline{x_{1stcol}})}{\sqrt{\sum_{i=0}^{N-2}(x_{(i,1)} - \overline{x_{1stcol}})^2}\sqrt{\sum_{i=0}^{N-2}(x_{(i+1,1)} - \overline{x_{1stcol}})^2}}$$

where N is the height of the reference block, M is the width of the reference block, $x_{(1,i)}$ is the sample at first row of the reference block, $x_{(i,1)}$ is the sample at first column of the reference block, $\overline{x_{1st\ row}}$ is the mean value at first row of the reference block and $\overline{x_{1st\ col}}$ is the mean value at first column of the reference block.

In the simple version of the proposed method, ρ' in (11) instead of ρ in (9) is used to apply the 12-tap DCT-IF with the same threshold as the proposed method. Table 4 also shows the experimental results of the simple version of the proposed method.

As shown in the right part of Table 4, the simple version achieved the overall BD-rate gains of −0.27%, −0.06% and 0.01% for Y, Cb, and Cr components, respectively. The BD-rate gains in Cb and Cr components in the simple version are slightly poorer than those in the proposed version with the reduction of computation complexity in the encoder and decoder, compared with the proposed method. When compared to the VVC anchor, the encoding and decoding times of 120% and 130%, respectively, in the proposed are reduced to those of 105% and 113%, respectively, in the simple version.

The simplified version that calculates 1-D correlations on the first row and first column reduces the computational

TABLE 4. Experimental results of the proposed methods.

Random Access Main 10										
Method		Proposed Method					Simple Version of the Proposed Method				
Class	Sequence name	Y	Cb	Cr	EncT	DecT	Y	Cb	Cr	EncT	DecT
A1	Tango2	-0.01%	-0.16%	0.27%	121%	113%	0.00%	0.54%	0.17%	106%	103%
	FoodMarket4	0.11%	-0.22%	-0.19%	123%	136%	0.11%	-0.66%	-0.18%	105%	116%
	Campfire	0.00%	-0.05%	0.18%	109%	121%	0.03%	0.06%	0.10%	102%	117%
A2	CatRobot1	-0.09%	-0.11%	0.32%	119%	139%	0.18%	-0.27%	0.11%	105%	118%
	DaylightRoad2	-0.09%	-0.17%	-0.03%	118%	141%	-0.08%	-0.13%	0.11%	105%	119%
	ParkRunning3	0.03%	0.01%	-0.08%	113%	121%	0.00%	-0.09%	0.08%	105%	102%
B	MarketPlace	0.00%	0.02%	-0.14%	121%	129%	-0.03%	0.04%	-0.10%	103%	111%
	RitualDance	0.01%	-0.12%	0.05%	119%	126%	0.02%	-0.05%	-0.07%	103%	111%
	Cactus	-0.03%	-0.17%	0.11%	114%	131%	-0.03%	-0.08%	0.11%	98%	112%
	BasketballDrive	-0.04%	0.19%	0.00%	118%	125%	-0.13%	0.15%	0.01%	101%	109%
	BQTerrace	0.07%	0.02%	-0.19%	122%	130%	0.08%	-0.01%	-0.01%	101%	109%
C	BasketballDrill	-0.20%	-0.07%	-0.06%	129%	132%	-0.21%	0.19%	-0.04%	111%	121%
	BQMall	-0.46%	-0.13%	-0.26%	130%	139%	-0.47%	-0.03%	0.08%	111%	120%
	PartyScene	-1.12%	-0.30%	-0.37%	119%	120%	-1.11%	-0.49%	-0.37%	110%	119%
	RaceHorses	-0.05%	0.03%	0.19%	119%	120%	-0.06%	-0.02%	0.19%	114%	117%
D	BassketballPass	-0.19%	-0.19%	0.14%	122%	134%	-0.29%	-0.40%	0.29%	105%	113%
	BQSquare	-1.67%	-0.76%	-0.76%	119%	137%	-1.66%	-0.78%	-0.83%	100%	119%
	BlowingBubbles	-1.13%	-0.14%	0.21%	121%	139%	-1.15%	-0.23%	0.20%	103%	116%
	RaceHorses	-0.28%	-0.03%	0.34%	120%	126%	-0.26%	-0.21%	0.05%	106%	112%
Class A1		0.03%	-0.14%	0.09%	118%	129%	0.05%	-0.02%	0.03%	104%	110%
Class A2		-0.05%	-0.09%	0.07%	117%	136%	0.03%	-0.17%	0.10%	105%	113%
Class B		0.00%	-0.01%	-0.03%	119%	128%	-0.02%	0.01%	-0.01%	101%	110%
Class C		-0.46%	-0.12%	-0.13%	124%	129%	-0.46%	-0.09%	-0.04%	112%	119%
Class D		-0.82%	-0.28%	-0.02%	121%	134%	-0.84%	-0.40%	-0.07%	104%	115%
Overall		**-0.27%**	**-0.12%**	**-0.01%**	**120%**	**130%**	**-0.27%**	**-0.06%**	**0.01%**	**105%**	**113%**

TABLE 5. The percentage of CU blocks that selected in the proposed method.

W \ H	4	8	16	32	64
4	-	24.8%	26.4%	24.8%	22.9%
8	24.5%	30.1%	24.9%	21.0%	18.8%
16	24.8%	24.5%	18.2%	18.3%	16.6%
32	25.0%	21.4%	19.0%	18.5%	17.6%
64	24.7%	19.0%	19.2%	18.6%	18.3%

complexity. However, the BD-rates are slightly decreased in Cb and Cr components because the correlation value is slightly less accurate than the proposed method.

Therefore, the simplified version can be used for low complexity application.

12-tap DCT-IF was also proposed in [16], but the sum of interpolation filter coefficient weights of [16] is 64, and that of the proposed interpolation filter in this paper is 128, so the curves of the two 12-tap DCT-IFs are slightly different. When the 12-tap DST-IF in [16] is applied instead of the 12-tap DCT-IF in the proposed method, the 12-tap DST-IF resulted in decrease of the overall BD-rates of 0.42%, 0.36% and 0.55% for Y, Cb, Cr components, respectively, compared with the VVC anchor. Many efficient intra and inter coding tools adopted in VVC reduced the performance of the DST-IF that showed good performance in HEVC.

IV. CONCLUSION

In this paper, an adaptive interpolation method is proposed with a 12-tap DCT-IF and correlation-based filter selection.

We designed the 12-tap DCT-IF that preserves the strong high frequency components of the input signal and applied the interpolation filter selection method based on the frequency characteristics of the reference block using correlation. The proposed method demonstrated coding performance with overall BD-rate gains of -0.27%, -0.12% and -0.01% for Y, Cb, and Cr components, respectively, compared with VVC anchor. The simplified version of the proposed method is applicable to block-based next-generation video coding standards because it shows BD-rate gain of -0.27% with a 105% encoding time increase and a 113% decoding time increase. In particular, the proposed method based on correlation shows better BD-rate improvements in low resolution sequence rather than high resolution. In order to improve the coding efficiency more, the interpolation study in high resolution video is required for the future topic.

REFERENCES

[1] B. Bross, J. Chen, S. Liu, and Y.-K. Wang, *Versatile Video Coding (Draft 10)*, document JVET-T2001, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, Oct. 2020.

[2] F. Bossen, X. Li, and K. Suehring, *AHG Report: Test Model Software Development (AGH3)*, document JVET-S0003, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Jul. 2020.

[3] *Advanced Video Coding (AVC)*, Standard ISO/IEC14496-10, May 2003.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003, doi: 10.1109/TCSVT.2003.815165.

[5] *High Efficient Video Coding (HEVC)*, Standard ISO/IEC 23008-2, Apr. 2013.

[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and 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, doi: 10.1109/TCSVT.2012.2221191.

[7] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021, doi: 10.1109/TCSVT.2021.3101953.

[8] L. Zhang, K. Zhang, H. Liu, Y. Wang, P. Zhao, and D. Hong, *CE4: History-Based Motion Vector Prediction (Test 4.4.7)*, document JVET-L0266, ITU-T/ISO/IEC, Joint Video Experts Team, Macao, China, Oct. 2018.

[9] L. Zhang, K. Zhang, H. Liu, H. C. Chuang, Y. Wang, J. Xu, P. Zhao, and D. Hong, "History-based motion vector prediction in versatile video coding," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2019, pp. 43–52.

[10] H. Chen, H. Yang, and J. Chen, *Symmetrical Mode for Bi-Prediction*, document JVET-J0063 ITU-T/ISO/IEC, Joint Video Experts Team, San Diego, CA, USA, Apr. 2018.

[11] J. Luo and Y. He, *CE4-Related: Simplified Symmetric MVD Based on CE4.4.3*, document JVET-M0444 ITU-T/ISO/IEC, Joint Video Experts Team, Marrakesh, Morocco, Jan. 2019.

[12] J. Chen, W.-J. Chien, N. Hu, V. Seregin, M. Karczewicz, and X. Li, *Enhanced Motion Vector Difference Coding*, document JVET-D0123 ITU-T/ISO/IEC, Joint Video Exploration Team, Chengdu, China, Oct. 2016.

[13] A. Henkel, B. Bross, M. Winken, P. Keydel, H. Schwarz, D. Marpe, and T. Wiegand, *Non-CE4: Switched Half-Pel Interpolation Filter*, document JVET-N0309 ITU-T/ISO/IEC, Joint Video Experts Team, Geneva, Switzerland, Mar. 2019.

[14] A. Henkel, I. Zupancic, B. Bross, M. Winken, H. Schwarz, D. Marpe, and T. Wiegand, "Alternative half-sample interpolation filters for versatile video coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 2053–2057.

[15] H. Lv, R. Wang, Y. Li, C. Zhu, H. Jia, X. Xie, and W. Gao, "A resolution-adaptive interpolation filter for video codec," in *Proc. ISCAS*, Melbourne, VIC, Australia, 2014, pp. 542–545.

[16] M. Kim and Y.-L. Lee, "Discrete sine transform-based interpolation filter for video compression," *Symmetry*, vol. 9, no. 11, p. 257, Nov. 2017, doi: 10.3390/sym9110257.

[17] H. Zhang, L. Song, Z. Luo, and X. Yang, "Learning a convolutional neural network for fractional interpolation in HEVC inter coding," in *Proc. VCIP*, St. Petersburg, FL, USA, Dec. 2017, pp. 1–4.

[18] L. Murn, S. Blasi, A. F. Smeaton, and M. Mrak, "Improved CNN-based learning of interpolation filters for low-complexity inter prediction in video coding," *IEEE Open J. Signal Process.*, vol. 2, pp. 453–465, 2021, doi: 10.1109/OJSP.2021.3089439.

[19] X. Xie, K. Zhang, L. Zhang, M. Wang, J. Li, and S. Wang, "Efficient interpolation filters for chroma motion compensation in video coding," in *Proc. VCIP*, Suzhou, China, Dec. 2022, pp. 1–5.

[20] X. Xie, K. Zhang, L. Zhang, J. Ki, M. Wang, and S. Wang, *Non-EE2: Long Tap Interpolation Filtering on Chroma Components*, document JVET-Y0172 ITU-T/ISO/IEC, Joint Video Experts Team, Jan. 2022.

[21] M. Coban, F. L. Léannec, K. Naser, J. Ström, and L. Zhang, *Algorithm Description of Enhanced Compression Model 5 (ECM 5)*, document JVET-Z2025, Apr. 2022.

[22] F. Bossen, J. Boyce, K. Suehring, X. Li, and V. Seregin, *VTM Common Test Conditions and Software Reference Configurations for SDR Video*, document JVET-T2010, Oct. 2020.

[23] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD Curves*, document VCEG-M33, ITU-T SG16 Q 6 Video Coding Experts Group (VCEG), Apr. 2001.

[24] G. Bjøntegaard, *Improvements of the BD-PSNR Model*, document VCEG AI11, ITU-T SG16 Q 6 Video Coding Experts Group (VCEG), Jul. 2008.

[25] *Versatile Video Coding (VVC) Test Model (VTM-16.0) Reference Software*. Accessed: Mar. 2022. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/tags/VTM-16.0

MIN-KYEONG CHOI received the B.S. degree in computer engineering from Sejong University, Seoul, South Korea, in 2022, where she is currently pursuing the M.S. degree in computer engineering. Her current research interests include image and video compression, image processing, and future video coding technologies.

YUNG-LYUL LEE (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, South Korea, in 1985 and 1987, respectively, and the Ph.D. degree in electrical and electronic engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1999. He was a Principal Researcher with the Samsung Electronics Research and Development Center, from 1987 to 2001. He has been a Professor with the Department of Computer Engineering, Sejong University, Seoul, since 2001. He was a Visiting Scholar with The University of Texas at Arlington, USA, from September 2006 to August 2007. He had contribution documents to AVC/H.264, high-efficiency video coding (HEVC), and versatile video coding (VVC) standards and had patents in AVC/H.264, HEVC, and VVC standards. His current research interests include video compression, image processing, 3-D video coding, deep learning, and CNN-based video coding. He received the Minister Prize from the Ministry of Commerce, Industry and Energy, South Korea, in November 2006; and the Korea Science Technology Superiority Paper Prize, in October 2006. He was the Chairperson of the Korea Institute of Broadcasting and Media Engineering, in 2020.