Modification of Intra Angular Prediction in HEVC

Shohei Matsuo, Seishi Takamura and Atsushi Shimizu†
†NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka, Japan
E-mail: matsuo.shohei@lab.ntt.co.jp Tel: +81-46-859-2070

Abstract— **Intra prediction of the emerging High Efficiency Video Coding (HEVC) standard has new features that the existing video coding standard H.264/AVC does not have. One example is that new angular prediction modes are added. Finer prediction directions enable to reduce the prediction error energy by making the predicted signals more flexibly. A method to make reference samples for the intra angular prediction plays an important role in terms of the coding efficiency. In the angular prediction of HEVC, a simple 2-tap linear filter is used to make reference samples. In this paper, the reference samples are generated by the conventional 2-tap linear filter or a DCT-based interpolation filter. The proposal improves the intra prediction performance especially for small prediction units such as 4x4 and 8x8. The average coding gains against the anchor of HEVC test model (HM6.0) were about 0.34% and 0.31%, when the tap length of DCT-IF is set to four and six, respectively. The maximum coding gains were about 2.2%, 3.3%, and 3.9% for each component (Y, Cb, and, Cr). In the case of four tap interpolation, the average run-times of encoding and decoding were about 102.84% and 100.96%, respectively.**

I. INTRODUCTION

Flexible block structures and many new coding tools are introduced into the emerging video coding standard: High Efficiency Video Coding (HEVC). An input image is divided into blocks called coding units (CU) based on quad-tree segmentation. Each CU can be independently separated into processing units for prediction and transformation. They are called prediction units (PU) and transform units (TU), respectively. Each coding function is performed for the corresponding unit sizes. Each unit size can be greater than 16x16: the size of the macro block of H.264/AVC. Prediction and transformation of larger block sizes are effective for compression of high resolution video.

New coding tools are, for example, sample adaptive offset (SAO) [1], DCT-based interpolation filter (DCT-IF) [2], internal bit depth increase (IBDI) [3] and so on. By using those new functionalities, it was reported that HEVC achieves about half bitrate of H.264/AVC based on subjective evaluation [4]. However, the improvement of intra prediction is smaller than that of inter prediction [5]. Therefore, it is necessary to invent new coding tools that improve the intra prediction. Some approaches for the improvement of intra prediction in HEVC have been proposed so far. The examples are short distance intra prediction (SDIP) [6] and bidirectional intra prediction (BIP) [7]. SDIP is one of line/pixel based intra prediction methods. By dividing the PU into lines or non-square blocks, SDIP can reduce the energy of prediction residuals by reducing the distance between the predicted pixel and its reference pixel. On the other hand, BIP calculates a weighted sum of two kinds of unidirectional intra-predicted signals according to the distance between the predicted pixel and the reference pixel. These two approaches provided good coding gains. However, there is a problem the conventional methods increase the computational complexity, especially for at the decoder side.

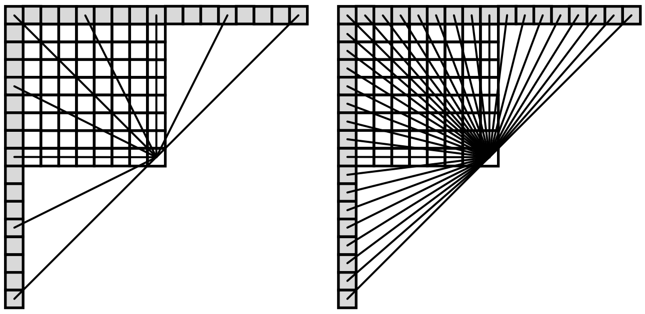

Fig. 1 Intra prediction directions of H.264/AVC (left) and HEVC (right).

H.264/AVC employs nine kinds of intra prediction modes: one DC mode and eight directional prediction modes. The DC mode calculates an average value of the available reference pixels and sets the average value to all pixels included in the block as the predicted value. This mode is effective for a smoothed area. The directional prediction modes are effective for complicated texture where the direction is orthogonal to a local gradient. When the prediction direction increases, it is possible to predict the texture more flexibly. In HEVC, 35 kinds of intra prediction modes are defined as follows: a planar mode, a DC mode, and 33 directional modes called "intra angular prediction" [8]. In the case of chrominance prediction, one additional mode where the predicted samples are made by luminance samples is prepared. It is important to improve the performance of the intra angular prediction because it accounts for a large part of the intra prediction modes.

As shown in Fig. 1, compared to the directional intra prediction of H.264/AVC, the intra angular prediction of HEVC can generate more flexible predicted samples and effectively reduce the prediction error energy. When the angular prediction mode is selected, the reference samples are generated by the following equation defined in the HEVC working draft [8] to make the predicted samples.

$$p[x,y]=((32-i)*r[x+j+1]+i*r[x+j+2]+16)>>5 \qquad (1)$$

In Eq. (1), $p[x, y]$ and $r[x]$ denote the predicted samples and the reference samples at the integer positions, respectively. The parameter i and j denote the multiplication factor and the index determined by the intra prediction mode and the position y. From Eq. (1), the predicted samples are generated by adding the weighted values of neighboring reference samples. In other words, the reference samples are made by a 2-tap linear interpolation filter. In this paper, for intra angular prediction in HEVC, an interpolation filter whose tap length is equal to or greater than 2 is employed for the generation of reference samples at the fractional positions.

II. Modification of Intra Angular Prediction

A. Basic Idea

This section describes the motivation and the benefit of this proposal. As shown in Fig. 2, assume that $A_{i,j}$ and $P_{i,j}$ are the locally-decoded reference samples and the predicted samples, respectively. The index i and j denote spatial coordinates for horizontal and vertical position, respectively. When the vertical mode is selected, for example, the value of $A_{0,0}$ is copied to $P_{0,1}$, $P_{0,2}$, …, $P_{0,b}$. In the same manner, the value of $A_{1,0}$ is copied to $P_{1,1}$, $P_{1,2}$, …, $P_{1,b}$.The parameter b denotes the selected PU size. When the down-left prediction mode is selected, the value of $A_{1,0}$ is copied to $P_{0,0}$ and that of $A_{2,0}$ is copied to $P_{1,0}$ and $P_{0,1}$, and so on. There is no difference between the eight prediction modes of H.264/AVC and HEVC as shown in Fig. 1.

When the angular prediction modes which H.264/AVC does not have are chosen, the reference samples $A_{\alpha,0}$ ($0<\alpha<1$), which are the gray circles in Fig. 2, are generated to make the predicted samples. In the implementation of the reference software called HEVC test model (HM) [9], the reference samples are generated by two samples such as $A_{0,0}$ and $A_{1,0}$ with the 2-tap linear interpolation filter of Eq.(1). The filter tap length is fixed regardless of the characteristics of the input image and coding information. When the PU size is larger, the texture tends to be flat and simple. Each reference sample may have the similar characteristics. Therefore, the 2-tap conventional interpolation would be sufficient. In addition, in the large PU size, the pixels at the bottom-right corner in the PU are far from the reference samples. There is a possibility that the effect of modification of interpolation filter is small.

However, when the PU size is smaller, the texture would be complicated. In addition, the pixels at the bottom-right corner in the PU are close to the reference samples. Therefore, the longer-tap interpolation may be able to generate the predicted samples that reduce the prediction error energy. DCT-IF [2] was proposed and it is currently employed in the HM software because of its effectiveness and unified design. In this paper, when the intra angular prediction is selected, $n(\geq 2)$-tap DCT-IF is applied to smaller PU sizes. In the example of Fig. 2, four samples such as $A_{-1,0}$, $A_{0,0}$, $A_{1,0}$, and $A_{2,0}$ are used to interpolate the reference samples at the fractional positions $A_{\alpha,0}$. DCT-IF is currently applied to the interpolation of the proposal. Of course, other interpolation approaches can be also easily applied to the proposal.

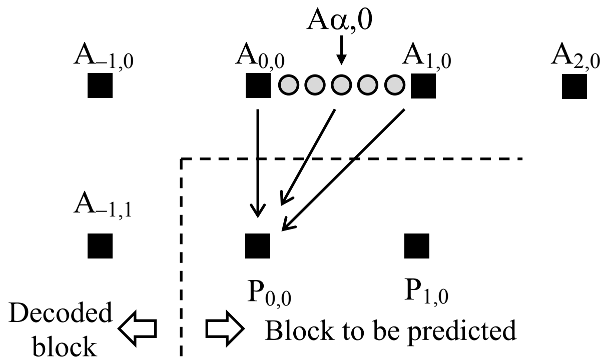

Fig. 2 Intra angular prediction in HEVC.

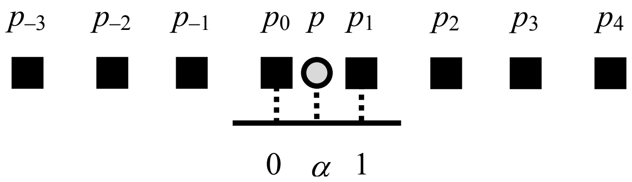

Fig. 3 Integer pixel positions (black squares) and an interpolated pixel (a shadowed circle) in DCT-IF.

B. DCT-based Interpolation Filter

Details of DCT-IF are explained in this section. Assume that the integer pixels are $\{p_i\}$ ($i = -(M-1), \ldots, M$) and that the fractional pixel p is in position α ($0<\alpha<1$) as shown in Fig. 3. The forward DCT yields the transformed coefficients set in the following equation.

$$c_k = \frac{1}{M}\sum_{l=-M+1}^{M} p(l)\cos\left(\frac{(2l-1+2M)k\pi}{4M}\right) \tag{2}$$

The inverse DCT returns exactly $p(x)$ for integer pixels $x = \{-(M-1), \ldots, M\}$.

$$p(x) = \frac{C_0}{2} + \sum_{k=1}^{2M-1} C_k \cos\left(\frac{\pi(2x-1+2M)k}{4M}\right) \tag{3}$$

The pixel value at a fractional position can be derived by using the corresponding shift (α) as the basis function argument.

$$p(\alpha) = \frac{C_0}{2} + \sum_{k=1}^{2M-1} C_k \cos\left(\frac{\pi(2\alpha-1+2M)k}{4M}\right) \tag{4}$$

From Eq. (4), filter coefficients for any fractional position can be calculated when the position α of the interpolated pixel and the number M of the integer pixels are given. In the HM software, quarter-pixel accuracy interpolation is employed. Modified 8-tap and 4-tap DCT-IF are used for the motion compensation of luminance and chrominance, respectively.

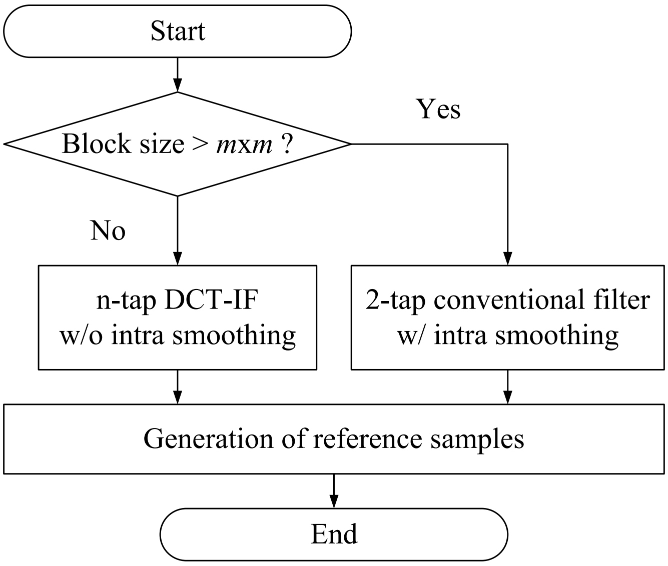

Fig. 4 Detailed flow chart of the proposal.

C. *Detailed Algorithm of the proposal*

The detailed encoding and decoding algorithm of the proposal is summarized as shown in Fig. 4. In the following explanation, the conventional 2-tap or n-tap DCT-IF are used to interpolate the reference samples for the intra angular prediction. The interpolation filter to be used is determined by the PU size.

Step 1: Set m that indicates a threshold value of the PU size.

Step 2: Input the PU size of intra prediction. If the PU size is larger than mxm, the conventional 2-tap interpolation filter is selected. Otherwise, n-tap DCT-IF is selected.

Step 3: Input the value of the intra prediction mode. If the input value is equal to the finer angular prediction, go to **Step 4**. Otherwise, the corresponding conventional intra prediction process is performed in the same way as the HM software does. After the process is finished, go to **Step 5**.

Step 4: Intra angular prediction is performed by using the interpolation filter defined in **Step 2**. When the conventional angular prediction is performed, in the same manner as the HM software, the intra smoothing filter whose coefficients are [1/4, 1/2, 1/4] is performed to the reference samples at the integer position. When the n-tap DCT-IF is performed, the intra smoothing filter is not performed. The reason for skipping the intra smoothing filter is to reduce the complexity of encoding and decoding. After the angular prediction is performed, go to **Step 5**.

Step 5: Move on to the next prediction block. Repeat the loop (**Step 1** to **5**) until all the intra prediction process is completed.

In this paper, the proposed encoder and decoder share the tap length n and the threshold value of the block size m. These two values are not encoded nor sent to the decoder. Therefore, there is no additional syntax related to the proposed intra prediction. To reduce the number of the operation for the filtering of intra prediction, when n-tap DCT-IF is employed, the intra smoothing is turned off.

Table 1 Configurations of the proposal (m=8, fixed).

PU size	Tap length	Intra smoothing
4x4	n (DCT-IF)	Off
8x8		
16x16	2 (same as HM)	On
32x32		
64x64		

III. EXPERIMENTS

A. *Experimental Conditions*

The proposed method was implemented in HM 6.0. It was applied to chrominance components as well as luminance component. Table 1 shows the configurations of the proposal. In the experiments, the threshold value of the PU size m was set to 8, which was fixed, and three kinds of tap length n in DCT-IF were tested: n={2, 4, 6}. The filter coefficients are derived based on Eq. (4) and they are modified with the phase parameter σ [2]. The modification is represented by the following equation.

$$c_i = \begin{cases} (1-\sigma)c_i + \sigma c_{i+1} & (i = -M+1) \\ \sigma c_{i-1} + (1-2\sigma)c_i + \sigma c_{i+1} & (-M+1 < i < M) \\ (1-\sigma)c_i + \sigma c_{i-1} & (i = M) \end{cases} \quad (5)$$

where c_i denotes the filter coefficients and the parameter i means the position of the filter. M is equal to half of the tap length n. For each tap length, ten kinds of σ are used: σ={0, 0.01, …, 0.09} for luminance. As for the chrominance, σ is set to 0 for all cases, which means the filter coefficients are not changed from the values derived by Eq. (4). In this paper, exploiting symmetry property, the number of the required filter coefficients is halved. The proposal requires not 31 but 16 sets of filters. The other configurations are the same as the common test condition [10] used in the standardization activity for HEVC. Two test sets are defined: Main profile and High Efficiency with IBDI (HE10). The resolution of the test sequences ranges 416x240 to 2560x1600. All frames of the input sequences were encoded with all-intra coding and four quantization parameters: QP(I)={22, 27, 32, 37}.

B. *Results and Analysis*

The experimental results are shown in Tables 2 and 3. The gains in BD-rate [11] against the HM6.0 anchor were calculated. The run-times of encoding and decoding are the average values of ten kinds of σ. The overall average gains of Main profile were about 0.34% and 0.31% for n=4 and 6, respectively. In the case of n=2, coding loss of 0.02% was observed. The most effective value of σ seemed to be 0.05. In the case of n=4 and σ=0.05, the maximum coding gains were about 2.2%, 3.3% and 3.9% (Y, Cb, Cr) for the sequence "BasketballDrill" in Main profile. The reason the sequence offered the high coding gain is that it has some slanted textures. Therefore, the selected ratio of the intra angular prediction is high, which results in the high coding efficiency.

Table 2 BD-rates and run-times (Main profile, Luminance).

σ	n=2	n=4	n=6
0	0.11%	–0.21%	–0.09%
0.01	0.09%	–0.33%	–0.21%
0.02	0.04%	–0.36%	–0.29%
0.03	0.02%	–0.39%	–0.32%
0.04	–0.01%	–0.40%	–0.37%
0.05	–0.02%	–0.41%	–0.40%
0.06	–0.02%	–0.39%	–0.38%
0.07	–0.03%	–0.35%	–0.37%
0.08	–0.02%	–0.32%	–0.33%
0.09	0.01%	–0.26%	–0.30%
Average	0.02%	–0.34%	–0.31%
Enc-time	101.89%	102.84%	106.28%
Dec-time	100.26%	100.96%	101.48%

Table 3 BD-rates and run-times (HE10, Luminance).

σ	n=2	n=4	n=6
0	0.10%	–0.09%	–0.04%
0.01	0.09%	–0.20%	–0.15%
0.02	0.05%	–0.23%	–0.23%
0.03	0.03%	–0.27%	–0.26%
0.04	0.01%	–0.29%	–0.31%
0.05	–0.01%	–0.30%	–0.33%
0.06	–0.01%	–0.28%	–0.33%
0.07	–0.01%	–0.26%	–0.31%
0.08	0.00%	–0.23%	–0.29%
0.09	0.02%	–0.18%	–0.25%
Average	0.03%	–0.23%	–0.25%
Enc-time	101.63%	102.51%	105.62%
Dec-time	100.13%	100.83%	101.25%

The average run-times of encoding and decoding in Main profile were about 101.89% and 100.26% (n=2), 102.84% and 100.96% (n=4), 106.28% and 101.48% (n=6). It was confirmed that the run-times of encoding and decoding increase in proportion to the tap length. When the tap length increases, the required operation such as addition and multiplication also increases. In terms of the balance of coding performance and computational complexity, the best tap length seems to be 4. When the intra smoothing is turned on, the overall average coding gain was 0.38% and increased by 0.04% compared to 0.34% of n=4.

As for the coding performance and run-times of HE10, the same tendency as Main profile was observed. However, the average coding gains of HE10 were less than those of Main profile. In HE10, some coding tools such as adaptive loop filter (ALF) [12], non-square quadtree transforms (NSQT) [13], and linear model for chrominance intra prediction (LMchroma) [14] are additionally turned on. These coding tools effectively reduce the prediction error energy. Thus, the coding gains of the proposal relatively attenuate. Moreover, adding some coding tools results in the increment of the run-times of HE10 compared to those of Main profile. Consequently, the increased complexity of the proposal seems to be small in the case of HE10.

IV. CONCLUSIONS

In this paper, an improved method of intra angular prediction in HEVC was proposed. The proposal employs n-tap DCT-IF to interpolate the reference samples at the fractional positions, especially for the small PU sizes such as 4x4 and 8x8. The average coding gains against the HM 6.0 were about 0.34% and 0.31% for n=4 and 6, respectively. The maximum coding gains were about 2.2%, 3.3% and 3.9% (Y, Cb, Cr) for the sequence "BasketballDrill" in Main profile in the case of n=4 and σ=0.05. The average run-times of encoding and decoding were 101.89% and 100.26% (n=2), 102.84% and 100.96% (n=4), 106.28% and 101.48% (n=6). The future work will include investigating the performance of different threshold m of PU size and applying different kinds of interpolation filters other than DCT-IF.

REFERENCES

[1] C.-M. Fu, C.-Y. Chen, Y.-W. Huang and S. Lei, "CE8 subset3: picture quadtree adaptive offset," JCTVC-D122, *4th JCT-VC Meeting*, January 2011.

[2] E. Alshina, J. Chen, A. Alshin, N. Shlyakhov and W.-J. Han, "CE3: Experimental results of DCTIF by Samsung," JCTVC-D344, *4th JCT-VC Meeting*, January 2011.

[3] T. Chujoh, and R. Noda, "Internal bit depth increase except frame memory," ITU-T SG16 Q.6 Document, VCEG-AF07, *32nd VCEG Meeting*, April 2007.

[4] TK. Tan, A. Fujibayashi, Y. Suzuki, and J. Takiue, "[AHG 8] Objective and subjective evaluation of HM5.0," JCTVC-H0116, *8th JCT-VC Meeting*, February 2012.

[5] B. Li, G. J. Sullivan, and J. Xu, "Comparison of compression performance of HEVC working draft 4 with AVC high profile," JCTVC-G399, *7th JCT-VC Meeting*, November 2011.

[6] X.Cao, et al. "CE6.b Report on Short Distance Intra Prediction Method," JCTVC-D299, *4th JCT-VC Meeting*, January 2011.

[7] T. Shiodera, A. Tanizawa, T. Chujoh, and T. Yamakage, "CE6 Subset A: Bidirectional intra prediction," JCTVC-D108, *4th JCT-VC Meeting*, January 2011.

[8] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 6," JCTVC-H1003, *8th JCT-VC Meeting*, February 2012.

[9] "HM reference software version 6.0," February 2012, http://hevc.kw.bbc.co.uk/trac/browser/jctvc-hm/tags/HM-6.0.

[10] F. Bossen, "Common test conditions and software reference configurations," JCTVC-H1100, *8th JCT-VC Meeting*, February 2012.

[11] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," VCEG-M33, ITU-T SG16 Q.6 Document, *13th VCEG Meeting*, April 2001.

[12] T. Chujoh, T. Yamakage, and G. Yasuda, "Improvement of quadtree-based adaptive loop filter," ITU-T SG16 Q.6 Document, C324, October 2009.

[13] Y. Yuan et al., "Asymmetric motion partition with OBMC and non-square TU," JCTVC-E376, *5th JCT-VC Meeting*, March 2011.

[14] J. Chen and J. Kim, "CE6.a.4: Chroma intra prediction by reconstructed luma samples," JCTVC-E266, *5th JCT-VC Meeting*, March 2011.