IMPROVED INTRA ANGULAR PREDICTION BY DCT-BASED INTERPOLATION FILTER

Shohei Matsuo, Seishi Takamura, and Hirohisa Jozawa

NTT Cyber Space Laboratories, NTT Corporation

ABSTRACT

This paper proposes a new intra angular prediction method exploiting interpolation technique whose tap length is greater than two. Intra prediction direction of High efficiency video coding (HEVC) standard is finer than that of H.264/AVC. A method to make reference samples for the intra angular prediction plays an important role in the coding performance. In this proposal, the reference samples are generated by the conventional 2-tap filter or a 4-tap DCT-based interpolation filter. The proposal improves the intra prediction performance especially for small prediction units such as 4×4 and 8×8. The average coding gains against the anchor of HEVC test model (HM5.0) were about 0.3%, 0.5%, and 0.5% for Y, Cb, and Cr. The maximum coding gains were about 1.8%, 2.5%, and 3.4% for each component. The average run-times of encoding and decoding were about 99.87% and 99.49%, respectively.

Index Terms— HEVC, video coding, intra angular prediction, DCT-based interpolation filter

1. INTRODUCTION

In the emerging video coding standard, High Efficiency Video Coding (HEVC), well-designed block structures and many new coding tools are introduced. Coding units (CU), prediction units (PU), and transform units (TU) are independently defined and each coding function is performed for the corresponding unit sizes. Each unit size can be greater than 16×16: the size of the macro block of H.264/AVC. Therefore, HEVC can effectively compress the 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 50% bit rate reduction compared to H.264/AVC based on subjective evaluation [4]. However, the improvement of intra prediction is smaller than that of inter prediction [5]. New coding tools that improve the intra prediction are necessary.

H.264/AVC employs nine kinds of intra prediction modes: eight directional prediction modes and one DC mode. The directional prediction modes are effective for texture where the direction is orthogonal to a local gradient. On the other hand, the DC mode is effective for a smooth area. 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" [6]. 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.

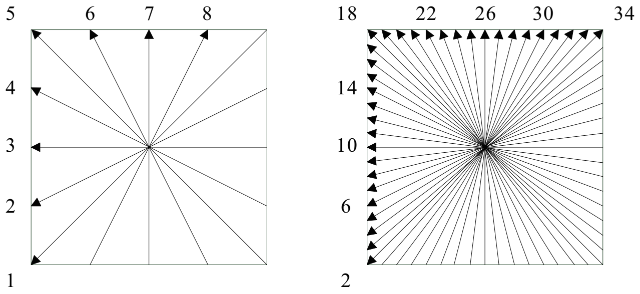

Fig. 1. Intra prediction directions of H.264/AVC (left) and HEVC (right).

The intra angular prediction of HEVC can generate more flexible predicted samples than the directional intra prediction of H.264/AVC, as shown in Fig. 1, and effectively reduce the prediction error energy. When the angular prediction mode is selected, the reference samples are generated by linear interpolation. Specifically, the following equation defined in the HEVC working draft [6] is used for generating the predicted samples. Note that the equation is tentative because HEVC is not finalized yet.

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

where $p[x,y]$ 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 interpolated by a 2-tap linear interpolation filter. In this paper, for intra angular prediction, an interpolation filter whose tap length is greater than 2 is applied to the generation of reference samples at the fractional positions.

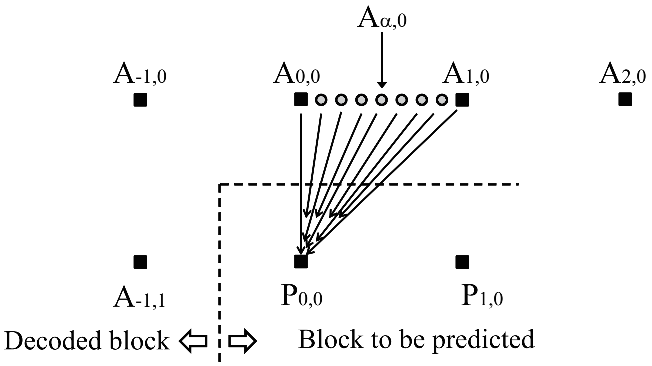

Fig. 2. Intra angular prediction of HEVC.

2. IMPROVED INTRA ANGULAR PREDICTION

2.1. Basic idea

This section explains the motivation of this proposal. Assume that $A_{i,j}$ and $P_{i,j}$ are the locally-decoded reference samples and the predicted samples, respectively, as shown in Fig. 2. 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 $P_{0,1}$, $P_{0,2}$, $\cdots$, $P_{0,b}$. The parameter b denotes the PU size.

When the angular prediction mode is selected, the reference samples such as $A_{\alpha,0}$ ($0< \alpha <1$), which are represented as gray circles in Fig. 2, are generated to make the predicted samples. In the current implementation of the reference software called HEVC test model (HM) [7], the reference samples are generated by using two samples such as $A_{0,0}$ and $A_{1,0}$ with the 2-tap linear interpolation filter. 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, when the PU size is larger, 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, $n(>2)$-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.

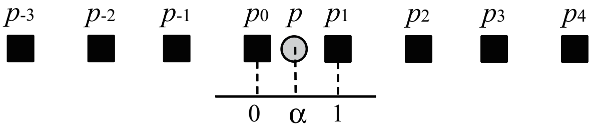

Fig. 3. Integer pixels (black squares) and an interpolated pixel (a shadowed circle).

2.2. DCT-based interpolation filter

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

$$C_k = \frac{1}{M} \sum_{l=-M+1}^{M} p(l) \cos\Big(\frac{(2l-1+2M)k\pi}{4M}\Big) \quad (2)$$

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

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

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

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

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.

2.3. Details of encoding and decoding algorithm

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 $n(>2)$-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 $m \times m$, the tap length of the interpolation filter is set to 2. Otherwise, it is set to n.
Step 3: Input the value of the intra prediction mode. If the input value is equal to the angular prediction, go to **Step 4**. Otherwise, the corresponding conventional intra prediction process is performed in the same way as the HM software. After the process is finished, go to **Step 5**.
Step 4: Intra angular prediction is performed by using the

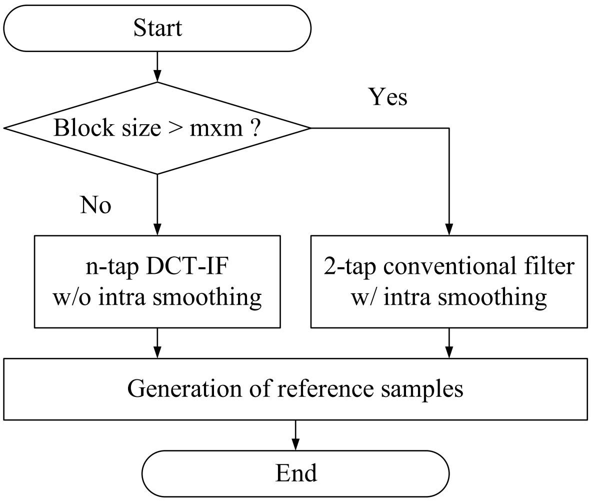

Fig. 4. Flow chart of proposed intra angular prediction.

interpolation filter defined in **Step 2**. Specifically, if the tap length is equal to 2, the conventional angular prediction is performed. In this case, the intra smoothing filter whose coefficients are [1/4, 1/2, 1/4] is performed to the reference samples at the integer position. Otherwise, n-tap DCT-IF is performed and the intra smoothing filter is not performed. The reason for skipping the intra smoothing filter is to reduce the computational complexity. 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 encoding and decoding complexity, when the tap length is equal to four, the intra smoothing is turned off.

3. EXPERIMENTS

The proposal was implemented in HM 5.0. It was applied to chrominance components as well as luminance component. The configurations of the proposal are listed in Table 1. In this experiment, the tap length n of DCT-IF is set to 4. The threshold value of the PU size m is set to 8. The phase parameter σ [2] is set to 0.07 for luminance and 0 for chrominance. The precise filter coefficients are described in Tables 2 and 3. The coefficients are derived based on Eq. (4). Note that the filter coefficient values shown in Tables 2 and 3 are normalized. The values of m and σ are decided by preliminary experimental results. The other configurations are subject to the common test condition [8] used in the standardization activity for HEVC. Three clusters are defined: High Efficiency (HE), Low Complexity (LC), and HE with IBDI (HE10). Each cluster has four GOP structures: All Intra (AI), Random Access (RA), Low Delay with B pictures (LB), and Low Delay with P pictures (LP). The resolution of the test sequences ranges 416×240 to 2560×1600. All frames of the sequences were encoded with four quantization parameters: QP(I)=$\{22, 27, 32, 37\}$.

Table 1. Configurations of the proposal

Prediction unit size	Tap length	Intra smoothing
4×4	4 (DCT-IF)	Off
8×8		
16×16	2 (same as HM)	On
32×32		
64×64		

The simulation results of the AI cases are shown in Table 4. The gains in BD-rate [9] against the anchor of HM 5.0 were calculated. The overall average gains were about 0.3%, 0.8%, 0.7% (Y, Cb, Cr in HE and HE10), 0.3%, 0.2%, and 0.2% (Y, Cb, Cr in LC). The maximum coding gains were about 1.8%, 2.5% and 3.4% (Y, Cb, Cr) for the sequence "BasketballDrill" in LC case. 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. The average run-times of encoding and decoding were about 99.87% and 99.49%, respectively. This complexity reduction comes from skipping of the intra smoothing process.

The simulation results of the inter-frame coding are shown in Table 5. Except for the Cr case of LP HE10, overall average gains were improved. The gains of chrominance were higher than that of luminance. It was confirmed that the proposal improved the coding efficiency with almost the same computational complexity as HM.

4. CONCLUSION

In this paper, an improved method of intra angular prediction in HEVC was proposed. The proposal employs a 4-tap DCT-IF to interpolate the reference samples at the fractional positions, especially for the small PU sizes such as 4×4 and 8×8. The average coding gains against the HM software were 0.3%, 0.8%, 0.7% (Y, Cb, Cr in HE and HE10), 0.3%, 0.2% and 0.2% (Y, Cb, Cr in LC). The maximum coding gains were about 1.8%, 2.5% and 3.4% (Y, Cb, Cr) for the sequence "BasketballDrill" in LC case. The average run-times of encoding and decoding were 99.87% and 99.49%, respectively. The future work will include investigating the performance of different tap length of DCT-IF and applying different kinds of interpolation filters other than DCT-IF.

Table 2. Filter coefficient values of 4-tap DCT-IF in case of σ=0.07 for luminance.

Position	Filter coefficients	Position	Filter coefficients	Position	Filter coefficients
0/32	N/A	11/32	{ −3, 45, 25, −3 }	22/32	{ −3, 23, 46, −2 }
1/32	{ 3, 55, 6, 0 }	12/32	{ −3, 43, 27, −3 }	23/32	{ −3, 21, 48, −2 }
2/32	{ 3, 54, 8, −1 }	13/32	{ −3, 41, 29, −3 }	24/32	{ −2, 19, 49, −2 }
3/32	{ 2, 53, 10, −1 }	14/32	{ −3, 39, 31, −3 }	25/32	{ −2, 17, 50, −1 }
4/32	{ 1, 53, 11, −1 }	15/32	{ −3, 37, 33, −3 }	26/32	{ −2, 15, 52, −1 }
5/32	{ 0, 53, 13, −2 }	16/32	{ −3, 35, 35, −3 }	27/32	{ −2, 13, 53, 0 }
6/32	{ −1, 52, 15, −2 }	17/32	{ −3, 33, 37, −3 }	28/32	{ −1, 11, 53, 1 }
7/32	{ −1, 50, 17, −2 }	18/32	{ −3, 31, 39, −3 }	29/32	{ −1, 10, 53, 2 }
8/32	{ −2, 49, 19, −2 }	19/32	{ −3, 29, 41, −3 }	30/32	{ −1, 8, 54, 3 }
9/32	{ −2, 48, 21, −3 }	20/32	{ −3, 27, 43, −3 }	31/32	{ 0 , 6, 55, 3 }
10/16	{ −2, 46, 23, −3 }	21/32	{ −3, 25, 45, −3 }	32/32	N/A

Table 3. Filter coefficient values of 4-tap DCT-IF in case of σ=0 for chrominance.

Position	Filter coefficients	Position	Filter coefficients	Position	Filter coefficients
0/32	N/A	11/32	{ −7, 50, 26, −5 }	22/32	{ −5, 23, 53, −7 }
1/32	{ −1, 63, 2, 0 }	12/32	{ −7, 48, 29, −6 }	23/32	{ −4, 20, 54, −6 }
2/32	{ −2, 63, 4, −1 }	13/32	{ −7, 46, 31, −6 }	24/32	{ −4, 18, 56, −6 }
3/32	{ −3, 62, 6, −1 }	14/32	{ −7, 43, 34, −6 }	25/32	{ −3, 16, 57, −6 }
4/32	{ −4, 62, 8, −2 }	15/32	{ −7, 41, 36, −6 }	26/32	{ −3, 13, 59, −5 }
5/32	{ −4, 59, 11, −2 }	16/32	{ −7, 39, 39, −7 }	27/32	{ −2, 11, 59, −4 }
6/32	{ −5, 59, 13, −3 }	17/32	{ −6, 36, 41, −7 }	28/32	{ −2, 8, 62, −4 }
7/32	{ −6, 57, 16, −3 }	18/32	{ −6, 34, 43, −7 }	29/32	{ −1, 6, 62, −3 }
8/32	{ −6, 56, 18, −4 }	19/32	{ −6, 31, 46, −7 }	30/32	{ −1, 4, 63, −2 }
9/32	{ −6, 54, 20, −4 }	20/32	{ −6, 29, 48, −7 }	31/32	{ 0 , 2, 63, −1 }
10/16	{ −7, 53, 23, −5 }	21/32	{ −5, 26, 50, −7 }	32/32	N/A

5. 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, Daegu, KR, 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, Daegu, KR, 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, San Jose, USA, 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, SanJose, CA, USA, 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, Geneva, CH, November 2011.

[6] 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, SanJose, CA, USA, February 2012.

[7] "HM reference software," December 2011, `http://hevc.kw.bbc.co.uk/trac/browser/jctvc-hm/tags/HM-5.0`.

[8] F. Bossen, "Common test conditions and software reference configurations," JCTVC-F900, 6th JCT-VC Meeting, Torino, IT, July 2011.

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

Table 4. Simulation results for all-intra coding.

Class and resolution of the test sequences	AI HE			AI LC			AI HE10		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A (2560×1600)	0.03	−0.48	−0.28	0.05	0.05	0.05	0.04	−0.55	−0.23
Class B (1920×1080)	−0.23	−0.84	−0.89	−0.22	−0.24	−0.30	−0.23	−0.84	−0.86
Class C (832× 480)	−0.52	−1.10	−1.07	−0.58	−0.75	−0.94	−0.51	−1.08	−1.07
Class D (416× 240)	−0.33	−0.75	−0.78	−0.36	−0.36	−0.40	−0.32	−0.76	−0.78
Class E (1280× 720)	−0.39	−0.69	−0.69	−0.52	0.68	0.62	−0.39	−0.74	−0.65
Class F (832×480 to 1280×720)	−0.41	−0.58	−0.60	−0.39	−0.40	−0.72	−0.38	−0.54	−0.57
Overall average of AI	−0.30	−0.75	−0.73	−0.31	−0.17	−0.24	−0.29	−0.75	−0.70
Enc-time [%]	99.96			99.63			100.01		
Dec-time [%]	99.49			99.64			99.34		

Table 5. Simulation results for inter-frame coding.

Class and resolution of the test sequences	RA HE			RA LC			RA HE10		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A (2560×1600)	0.00	−0.73	−0.69	0.02	−0.71	−0.19	0.08	−0.72	0.16
Class B (1920×1080)	−0.13	−0.94	−0.87	−0.14	−0.92	−1.06	−0.07	−1.04	−1.03
Class C (832× 480)	−0.25	−0.83	−0.72	−0.28	−1.02	−1.29	−0.20	−0.78	−0.79
Class D (416× 240)	−0.16	−0.72	−0.47	−0.23	−0.32	−0.81	−0.16	−0.85	−0.21
Class E (1280× 720)	N/A								
Class F (832×480 to 1280×720)	−0.23	−0.59	−0.48	−0.17	−0.86	−0.86	−0.09	−0.43	−0.29
Overall average of RA	−0.14	−0.81	−0.70	−0.15	−0.75	−0.85	−0.09	−0.86	−0.50
Enc-time [%]	100.01			99.96			100.00		
Dec-time [%]	99.67			99.69			99.86		

Class and resolution of the test sequences	LB HE			LB LC			LB HE10		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A (2560×1600)	N/A								
Class B (1920×1080)	−0.02	−0.43	−0.14	−0.08	−0.88	−0.63	−0.02	−0.17	−0.25
Class C (832× 480)	−0.12	−0.48	−0.23	−0.13	−0.28	−0.89	−0.11	−0.07	−0.23
Class D (416× 240)	−0.02	−0.80	−0.48	−0.17	0.25	−0.68	−0.07	−0.57	0.23
Class E (1280× 720)	0.02	0.22	−1.21	−0.25	−0.51	−1.43	−0.07	−0.13	−0.63
Class F (832×480 to 1280×720)	−0.27	−0.24	−0.70	0.14	−1.15	−0.64	−0.12	−0.13	−0.26
Overall average of LB	−0.08	−0.38	−0.50	−0.09	−0.53	−0.82	−0.07	−0.21	−0.21
Enc-time [%]	100.04			99.96			100.08		
Dec-time [%]	100.34			99.82			99.85		

Class and resolution of the test sequences	LP HE			LP LC			LP HE10		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A (2560×1600)	N/A								
Class B (1920×1080)	−0.04	−0.34	−0.55	−0.07	−0.47	−1.09	−0.02	−0.25	−0.03
Class C (832× 480)	−0.05	−0.30	−0.45	−0.15	−0.32	−0.60	−0.14	−0.11	0.04
Class D (416× 240)	−0.04	−0.05	−0.06	−0.19	−0.34	−0.39	−0.02	−0.62	0.69
Class E (1280× 720)	−0.04	−0.65	−0.59	−0.15	−0.52	−2.04	−0.17	0.05	−0.41
Class F (832×480 to 1280×720)	−0.13	−0.28	−0.46	−0.01	−0.65	−0.08	−0.36	−0.23	−0.28
Overall average of LP	−0.06	−0.31	−0.42	−0.11	−0.46	−0.79	−0.13	−0.25	0.02
Enc-time [%]	100.10			100.04			100.07		
Dec-time [%]	99.74			100.36			99.83		