Upper Coding Unit Depth Rate Distortion Cost Comparison Based Fast Intra and Interlayer Inter Prediction Mode Selection for Enhancement Layer in All Intra Spatial SHVC

Pei Liu, Guojing Zhu, Gaoxing Chen and Takeshi Ikenaga
Graduate School of Information, Production and Systems, Waseda University
Fukuoka 808-0135, Japan
E-mail: liupei@asagi.waseda.jp

Abstract: SHVC is the scalability extension of the video compression standard high efficiency video coding (HEVC). It encodes the video in different layers, called base layer (BL) and enhancement layer (EL). The additional interlayer inter prediction (ILIP) mode employed in EL increases the complexity. This paper proposes a fast mode selection method for EL in all intra spatial SHVC. The rate distortion cost (RD-cost) difference between intra prediction and ILIP in upper coding unit (CU) depth is calculated as a prejudgment condition to select the prediction mode in lower CU depth. The proposed method can achieve 52.65% and 58.00% encoding time saving for 1.5x and 2x sequences respectively with negligible BD-rate increase compared with SHM-8.0. Compared with related work, this paper can get over 30% time saving increase on average.

Keywords-- **SHVC, Mode Selection, Enhancement Layer, Coding Unit Depth, Rate Distortion Cost**

1. Introduction

Nowadays, with the development of network and information technology, the demand for video streaming to mobile devices such as smart phones, tablet computers or laptops is rising. These devices have broad variety of screen sizes, computing capabilities and networks [1], which are not supported by just utilizing high efficiency video coding (HEVC). As a result, the scalability extension of HEVC, which is known as SHVC, becomes more and more important.

One of the scalabilities specified in SHVC is all intra spatial scalability. It has the advantage of no error propagation. Also, it is not affected by sharp motion and is able to support devices with different resolution needs. SHVC encodes the video in different layers. In this scalability, one base layer (BL) and one enhancement layer (EL) with higher resolution are utilized [2]. The prediction process of BL only includes intra prediction part, while EL conducts both intra prediction and an additional interlayer inter prediction (ILIP), which utilizes the up-sampled interlayer reference pictures from reference layer.

Frames in SHVC are divided into multiple coding tree units (CTUs). As shown in Figure 1, the prediction process starts from a largest coding unit (LCU) which consists of 64x64 pixels, the block in this size means coding unit (CU) depth 0. Then, a LCU is split into CU depth 1, 2 and 3 using a quadtree structure. The further split of a LCU means the lower CU depth in this paper. Both intra prediction and ILIP should be conducted from top to bottom of all the CU depths.

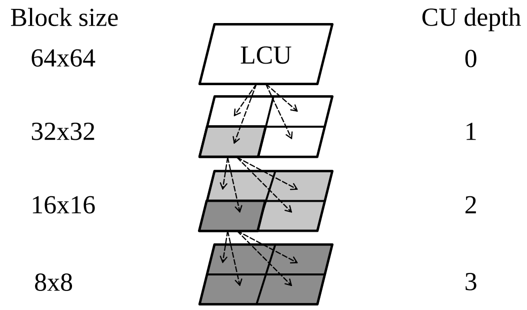

Figure 1. CU Split and CU Depth in SHVC

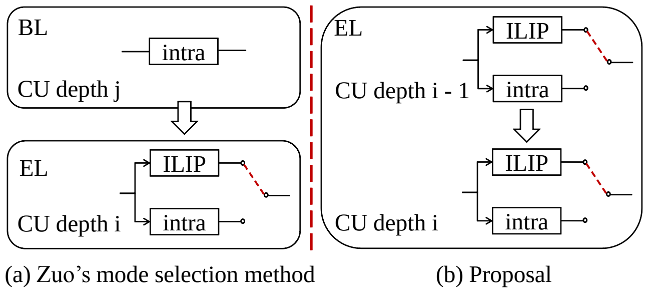

Figure 2. Concept Comparison between Zuo's Mode Selection Method and Our Proposal

Rate distortion cost (RD-cost) of all the mode candidates should be calculated to derive the best mode and CU depth with the smallest RD-cost. This recursive prediction process is time consuming, and the additional ILIP will definitely cause the coding time of EL to increase as well. Thus, one main purpose of works on SHVC is to reduce the coding time of EL while maintain the quality as much as possible.

To achieve this purpose, there are two main approaches, one is to reduce the complexity of original SHVC modules, and the other is to do fast mode selection before original mode decision process. For each approach, there are some papers concerned. For example, one of the proposals in [3] is using the mode information of co-located BL CUs to restrict the intra prediction modes and the mode number in EL. However, the time reduction of this kind of approach is not much in general. For fast mode decision method, one of the proposals in Zuo's paper [4] is to conditionally skip CU depth in EL based on co-located BL CU depth when the possibility to choose it is relatively low, as shown in the left part of Figure 2. For example, when the co-located BL CU depth is 0, the possibility of selecting CU depth 3 with intra prdiction is 4.7%, Zuo proposes to skip intra prediction in CU depth 3 for EL under this condition.

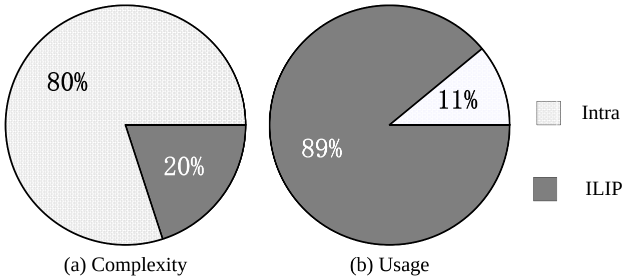

Figure 3. Complexity and Usage Percent of Intra and ILIP

We choose the second approach and our target is exactly the same as Zuo's mode selection method. Zuo's main idea is based on the correlation between BL and EL. However, intra prediction is used to obtain edge information in one frame, on the contrast, ILIP is an inter prediction mode between frames, which uses zero motion vector. It means these two kinds of prediction modes focus on different features of a sequence, thus CU split of ILIP can't be derived from the CU depth of intra mode in BL directly. Additionally, Zuo sets CU depth skip conditions based on BL and EL CU depth correlation of two sequences for 2 times resolution ratio. Nevertheless, different sequences have different features, and in some other sequences, the CU depth correlation between BL and EL is not as obvious as the sample sequences in Zuo's paper, so this kind of method is not suitable for other sequences and other spatial ratios.

In order to solve the remaining problems and achieve large time saving, this paper chooses to just focus on EL, and proposes a fast mode selection method between intra prediction and ILIP using upper CU depth RD-cost comparison, as shown in the right part of Figure 2.

The remaining parts of the paper are arranged as the following. In Section 2, the motivation for choosing this kind of proposal is explained. In Section 3, the proposal is interpreted in detail. Section 4 describes the experiment conditions and the experiment results. The final section draws the conclusion of this paper.

2. Motivation

The motivation for the proposed upper CU depth RD-cost comparison based fast intra and ILIP mode selection method includes three parts.

First, intra prediction and ILIP process take much complexity during the whole coding process. This is because SHVC inherits the 35 intra prediction modes presented in HEVC, and ILIP is an additional prediction method put forward in SHVC, which will increase the coding time. Also, the recursive prediction process using rate distortion optimization (RDO), which means trying all and selecting the best, is very time consuming.

Second, Figure 3 shows that in all intra spatial SHVC, intra prediction occupies about 80% complexity with only 11% usage, while ILIP takes 89% usage with just 20% complexity in EL. It represents that the usage is not in direct proportion to complexity. Thus, mode selection is necessary.

Third, both intra prediction and ILIP are conducted for all the four kinds of CU depths in EL. For that different

Table 1. Percentage of Selecting Intra in CU Depth 0 for EL

Sequence	Percentage of selecting intra in CU depth 0 for EL (%)
Traffic	0.001
PeopleOnStreet	0.049
Kimono	0.161
ParkScene	0.049
Cactus	0.008
BasketballDrive	0.424
BQTerrace	0.659

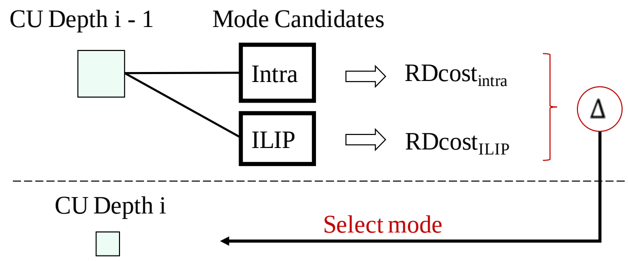

Figure 4. Proposed Upper CU Depth RD-cost Comparison based Mode Selection Method

sizes of CUs for one LCU represent different parts of the same picture, there is correlation between adjacent CU depths, which means the information of upper CU depth can be utilized when the prediction process of lower CU depth is conducted.

3. Proposed Method

As the prediction process is done from top to bottom, which is mentioned in Section 1, the information of upper CU depth can be utilized in lower CU depth. This paper proposes an upper CU depth RD-cost comparison based fast intra and ILIP mode selection method for EL in all intra spatial SHVC. It means recording the RD-cost information of intra prediciton and ILIP seperately at first, then the difference between intra RD-cost and ILIP RD-cost is utilized as a prejudgment condition in one CU depth to decide which mode to be selected in lower CU depth.

For that CU depth 0 is the topside CU depth in the prediction process, no upper depth RD-cost information can be used. After testing all the sequences, statistic data show some trends in the percentage of modes and CU depths. As shown in Table 1, intra prediction in CU depth 0 is selected as the final prediction mode for less than 1% during EL prediction process when quantization parameter (QP) equals 22. It means that intra prediction is rarely selected in CU depth 0 when doing EL prediction. Thus, we propose to select ILIP and skip intra prediction in CU depth 0 of EL for time reduction. Even if we skip this mode, the coding efficiency will almost not be affected.

For that only ILIP is conducted in CU depth 0, no RD-cost compasison can be got between intra prediction and ILIP. As a result, both intra prediction and ILIP are conducted in CU depth 1 as it is in original SHVC.

Table 2. Experiment Results Comparison between Zuo's Method and Our Proposed Method in SHM-8.0

Sequences	Zuo's method (1.5x)		Proposed method (1.5x)		Zuo's method (2x)		Proposed method (2x)	
	Y BD-rate (%)	TS (%)	Y BD-rate (%)	TS (%)	Y BD-rate (%)	TS (%)	Y BD-rate (%)	TS (%)
Traffic					0.1	16.63	0.3	58.50
PeopleOnStreet					0.1	15.63	0.5	58.25
Kimono	0.5	30.75	0.0	52.63	0.6	31.88	0.1	59.13
ParkScene	0.5	18.63	0.0	52.38	0.4	19.88	0.1	58.50
Cactus	0.4	18.63	0.1	52.63	0.7	20.00	1.6	57.75
BasketballDrive	0.6	24.38	0.4	52.63	1.2	26.63	2.5	56.38
BQTerrace	0.3	18.13	0.3	53.00	0.5	19.13	2.0	57.50
Average	0.5	22.10	0.2	52.65	0.5	21.40	1.0	58.00

The concept of upper CU depth rate distortion cost (RD-cost) comparison based fast intra and ILIP mode selection method is shown in Figure 3. Here, i means the depth of current CU. As RD-cost comparison based mode selection method is applied in CU depth 2 and 3, i equals 2 or 3 in our proposal. CU depth i - 1 means the upper CU depth, which has already been coded. Δ is defined as

$$\Delta = \mathrm{RDcost}_{\mathrm{intra}} - \mathrm{RDcost}_{\mathrm{ILIP}} \quad (1)$$

$\mathrm{RDcost}_{\mathrm{intra}}$ means the RD-cost of the best mode derived by intra prediction, and $\mathrm{RDcost}_{\mathrm{ILIP}}$ means the RD-cost of the best mode selected by ILIP in CU depth i. Here, RD-cost means the cost it takes for coding if choosing this mode, so the smaller RD-cost means the better mode. First, we record the RD-cost of intra prediction and ILIP in CU depth $\mathrm{i}-1$, and then judge the value of Δ to select the mode we need to use in CU depth i. When the value of Δ differs, it represents that the blocks have different kinds of features. Thus we set the mode selection conditions in CU depth i accoording to the value of Δ as the following

I. Δ > Threshold 1, select ILIP.
II. Δ < Threshold 2, select intra prediction.
III. Other conditions, do both intra prediction and ILIP.

In condition I, "Δ > Threshold 1" means RD-cost of ILIP is much smaller than intra prediction, so that ILIP is much better than intra prediction in CU depth $\mathrm{i}-1$, thus we still select ILIP in CU depth i and skip intra prediction to reduce coding time. In condition II, "Δ < Threshold 2" means intra prediction is much better than ILIP, thus we still select intra in CU depth i and skip ILIP. For other cases, there is no obvious difference between intra prediction and ILIP in CU depth $\mathrm{i}-1$, we can't judge whether intra prediction or ILIP is better in CU depth i before the original prediction process, so both intra prediction and ILIP are conducted in CU depth i. When the situation meets condition 1 or 2 and i equals 2, the selected mode is inherited in CU depth $\mathrm{i}+1$. The RD-cost comparison of intra prediction and ILIP is utilized for the parts where both intra prediction and ILIP are conducted. That's to say, when we choose to select intra in CU depth 2, then we still select intra in CU depth 3, and it is the same with ILIP. For the parts that both intra prediction and ILIP are conducted in CU depth 2, three mode selection conditions are checked to decide the mode to be selected in CU depth 3.

The reason for choosing RD-cost comparison is that the original SHVC algorithm uses the rate distortion optimization (RDO) process to derive the best mode with the smallest RD-cost. So that, RD-cost is the only criterion used to decide the final mode, and no additional calculation is needed in our proposal.

After testing all the sequences, RD-cost data show that, when the difference of RD-cost between the best modes of intra prediction and ILIP is big enough, which means the predictions using these two kinds of modes have obvious different efficiencies, there's a high possibility that even the CU is split into lower depth, the prediction mode selection between intra and ILIP doesn't change. That's to say, if we choose intra prediction in CU depth 1, and the RD-cost difference between intra prediction and ILIP is big, then intra prediction will still be selected in CU depth 2.

When the difference between Threshold 1 and Threshold 2 is small, modes will be skipped for more CUs, and time saving will increase with more quality loss. On the contrast, when the difference between two thresholds is large, less time saving and quality loss will be achieved. The threshold can be changed to get the result we want.

4. Experimental Result

4.1 Experiment conditions

The proposed fast mode selection method is implemented in SHVC test model 8.0 (SHM-8.0). For evaluation, two classes of sequences recommended in JCT-VC [5] are tested. Class A includes two sequences, and is only tested for 2x spatial ratio, which means the resolution of EL is twice of BL both in the width and height, while class B includes five sequences, and is tested for both 1.5x and 2x spatial ratio in all intra configuration. One BL and one EL are tested in the experiment. The quantization parameters (QP) of BL and EL are set as $\mathrm{QP}_{\mathrm{BL}} \in \{22, 26, 30, 34\}$ and $\mathrm{QP}_{\mathrm{EL}} = \mathrm{QP}_{\mathrm{BL}} + \Delta\mathrm{QP}$ respectively, $\Delta\mathrm{QP} \in \{0, 2\}$.

There are two kinds of evaluation criterions, BD-rate and time saving (TS). BD-rate is the combination of bitrate and peak signal to noise ratio (PSNR), BD-rate increase means compression efficiency loss and is calculated as [6]

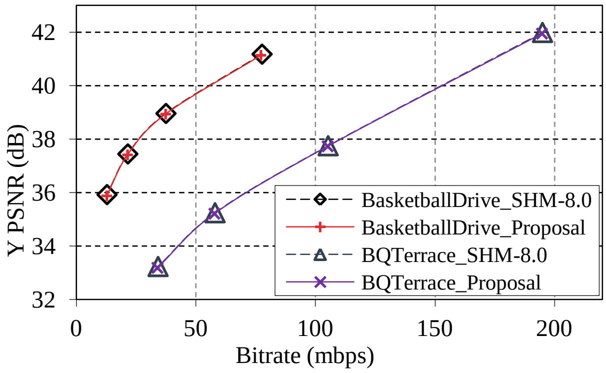

Figure 5. RD-curve Comparison for 1.5x Sequences

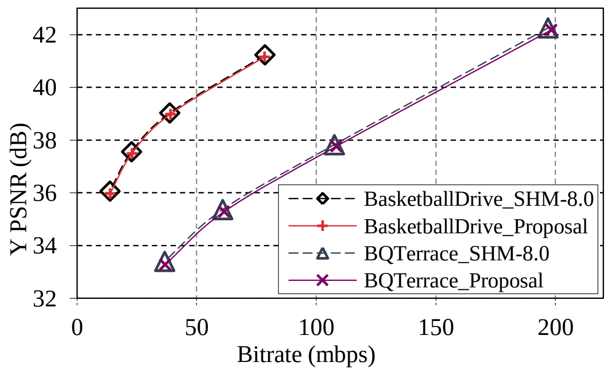

Figure 6. RD-curve Comparison for 2x Sequences

introduces. TS increase means coding complexity becomes lower. Here, TS is defined as

$$TS = \frac{Time_{SHM-8.0} - Time_{proposal}}{Time_{SHM-8.0}} \times 100\% \quad (2)$$

In the experiment, we set Threshold 1 as 100 and Threshold 2 as 0 in CU depth 1. In CU depth 2, we set Threshold 1 and Threshold 2 as 0.

4.2 Experiment results

Zuo's fast mode decision method is based on SHM-5.0, as it is the different version, we implement it in SHM-8.0. Table 2 shows the experiment results comparison between Zuo's method and our proposal in aspects of BD-rate of Y component and TS. The proposed algorithm in this paper can achieve 52.65% and 58.00% time saving with 0.2% and 1.0% BD-rate increase for 1.5x and 2x sequences respectively on average compared with SHM-8.0. The time reduction is stable for all the sequences. Compared with Zuo's work, our proposal can obtain -0.3% and 0.5% BD-rate increase with 30.55% and 36.60% TS increase for 1.5x and 2x sequences separately. It means our proposal is better in both TS and coding efficiency for 1.5x sequences and can obtain much more time saving with a little BD-rate increase for 2x sequences. This is mainly because the mode skip conditions in Zuo's paper are only set for 2x spatial ratio sequences. Even compared with the results shown in Zuo's paper, our proposal can get about 7.64% and 16.15% more time saving with 0.125% and 0.88% BD-rate increase, for 1.5x and 2x sequences respectively, when testing the same sequences.

In the experiment, the worst two cases of our proposal are BasketballDrive and BQTerrace for 2x sequences, which increase BD-rate more. This is because there are more detail parts which need deeper CU split to decide the final mode and depth in these sequences. As shown in Figure 5 and Figure 6, we draw the RD-curves, which represent BD-rate, of these two sequences for both 1.5x and 2x spatial ratio, when $\Delta QP = 0$. The curves in lines represent the results of our proposed method, and the curves in dashed lines represent the result of original SHM-8.0. The curves for each sequence are close to each other, therefore, our proposal can achieve the time saving and maintain the coding efficiency as well.

5. Conclusion

This paper proposes a fast enhancement layer mode selection algorithm for all intra spatial scalability in SHVC. The RD-cost difference of intra prediction and ILIP in upper CU depth is considered as a prejudgment condition to select possible prediction mode before the original mode decision process in lower CU depth, the needless mode is skipped for complexity reduction. As is shown in the experiment results, 52.65% and 58.00% time saving can be achieved for 1.5x and 2x spatial ratio sequences respectively compared with SHM-8.0. Compared with Zuo's work, 30.55% and 36.60% more time saving can be obtained. BD-rate increase is negligible compared with time saving increase.

Acknowledgment

This work was supported by KAKENHI (26280016).

References

[1] P. Helle, H.Lakshman, M.Siekmann, J. Stegemann, T. Hinz, H. Schwarz, D.Marpe, T. Wiegand, " A Scalable Video Coding Extension of HEVC," *Data Compression Conference (DCC)*, 2013.

[2] J. Chen, J. Boyce, Y. Ye and M. M. Hannuksela, "Scalable HEVC (SHVC) Test Model 8(SHM 8)," *JCT-VC S1007*, Oct. 2014.

[3] G. Zhu, G. Chen, T. Ikenaga, "Fast Enhancement Layer Intra Coding Based on Inter-channel Correlations and TU depth Correlation in SHVC", *CSPA 2016*, Mar. 2016.

[4] X. Zuo and L. Yu, "Fast mode decision method for all intra spatial scalability in SHVC," *IEEE Visual Communications and Image Processing Conference*, pp. 394-397, Dec. 2014.

[5] V. Seregin and Y. He, "Common SHM test conditions and software reference configurations," *JCT-VC Q1009*, Apr. 2014.

[6] G. Bjontegaard, "Calculation of average PSNR differences between RD curves", *Video Coding Experts Group (VCEG)*, VCEG-M33, Austin, Texas, USA, April. 2001.