CONTENT ADAPTIVE COMPLEXITY REDUCTION SCHEME FOR QUALITY/FIDELITY SCALABLE HEVC

Hamid Reza Tohidypour, Mahsa T. Pourazad[1,2], and Panos Nasiopoulos[1]
[1]Department of Electrical & Computer Engineering, University of British Columbia, Canada
[2]TELUS Communications Inc., Canada

ABSTRACT

There has been significant interest in developing a scalable version of the High Efficiency Video Coding (HEVC) standard. As expected, the HEVC scalable video version increases the complexity of the codec compared to the non-scalable counterpart. In this paper, we propose an adaptive early-termination interlayer motion prediction mode search that significantly reduces HEVC/SVC's coding complexity by up to 85.77%, while maintaining the overall bitrate.

Index Terms— Scalable HEVC, video compression, low complexity compression

1. INTRODUCTION

Video applications have become part of our everyday lives, using a variety of devices with different screen resolutions, processing capabilities and network bandwidth requirements. This means each video stream needs to be encoded in a way that is compatible with a specific viewing device and all of these coded streams must be transmitted separately (simulcast coding). This approach is computationally expensive and requires large amounts of bandwidth. One solution is to use Scalable Video Coding (SVC) which enables multicast service and video transmission to heterogeneous clients with different capabilities [1][2]. An SVC stream consists of a base layer and one or more enhancement layers. On the decoder side, based on the type of the application and supported complexity level, the appropriate part of an SVC bit stream will be decoded. While SVC enables the video coding system to deliver different versions of the same video content within the same bit stream, it significantly increases coding complexity.

The latest scalable video coding standard (H.264/SVC) is an extension of H.264/AVC standard [1]. Recently the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) has introduced a new compression standard, known as the High Efficiency Video Coding (HEVC), which has substantially higher compression capabilities (up to 45.54% in terms of bit rate) than H.264/AVC standard, but with higher computational complexity [3]. Due to HEVC's compression performance, there has been significant interest in developing a scalable version of this standard. One obvious challenge in this case is the complexity of the scalable HEVC (HEVC/SVC) implementation. To the best of our knowledge, there is no existing work on the reduction of the HEVC/SVC complexity. To this end, in our study we propose a scheme to effectively reduce the complexity of the SNR/Quality/Fidelity scalable HEVC. Our scheme utilizes the correlation between the enhancement layers and the base layer to reduce the inter and intra prediction computational complexity. This is done by utilizing specific coding information from the base layer such as motion homogeneity, prediction modes, and the rate distortion cost in the coding process of the enhancement layer.

The rest of this paper is organized as follows: Section 2 includes a short overview on the HEVC standard. Section 3 elaborates on our proposed method. Performance evaluation of our method is presented in Section 4 and the conclusion is drawn in Section 5.

2. HEVC OVERVIEW

HEVC utilizes a quad-tree based coding structure with support for coding units of more diverse sizes than that of macro-blocks in H.264/AVC. The basic block in HEVC is known as the Largest Coding Unit (LCU), which is 64x64 and can be re-cursively split into smaller Coding Units (CU), which in turn can be split into small Prediction Units (PU) and Transform Units (TU). HEVC employs more complicated intra prediction modes and more flexible motion compensation than H.264/AVC to reduce the spatial and temporal redundancies [3]. For intra prediction, HEVC uses 35 luma intra prediction modes compared to 9 used in H.264/AVC. Furthermore, intra prediction can be done at different block sizes, ranging from 4x4 to 64x64.

In the case of inter-prediction, HEVC introduces a technique called "motion merge". For every inter-coded PU, the encoder can choose between 1) the motion merge mode, 2) the SKIP mode, or 3) explicit encoding of motion parameters. The motion merge mode involves creating a list of previously coded neighboring (spatially or temporally) PUs (called candidates) for the PU being encoded. The motion information for the current PU is copied from one selected candidate, avoiding the need to encode a motion vector for the PU; instead only the index of a candidate in the motion merge list is encoded as well as the residual data.

In the SKIP mode, the encoder signals the index of a motion merge candidate and the motion parameters for the current PU are copied from the selected candidate, without sending any residual data.

In explicit coding, inter-coded CUs can use Symmetric and Asymmetric Motion Partitions (AMP). AMPs allow for asymmetrical splitting of a CU into smaller PUs. AMP can be used on CUs of size 64x64 down to 16x16, improving coding efficiency since it allows PUs to more accurately conform to the shape of objects, without requiring further splitting [3]. Each PU, coded using inter-prediction, has a set of motion parameters, which consists of a motion vector, a reference picture index and a reference list flag.

3. OUR PROPOSED SCHEME

The focus of our study is to reduce the complexity of SNR/Quality scalable HEVC by minimizing the redundant computations involved in intra and inter prediction process while encoding the enhancement layer. The following subsections elaborate on our proposed scheme.

3.1. Adaptive search range

In the inter prediction process, selecting a large search range leads to high computational costs, while selecting a small search range produces poor matching results. The optimal motion search range would result in reduced complexity without hampering the compression performance.

In the case of scalable video coding, this may be possible since there is a correlation between the base layer and the enhancement layer, the MVs of the base layer and those of the enhancement layer are also correlated. In our study we utilize this correlation to select the proper motion search range for the enhancement layer based on the motion information of the base layer. Our approach is inspired by the scheme proposed for the existing H.264/SVC standard in [4]. We classify the LCUs within each frame in the base layer to three different groups: 1) with homogeneous motion, 2) with moderate motion, and 3) with complex motion. This is achieved by defining the motion vector deviation (so called MV homogeneity) of each LCU as follows:

$$MVH_{m,n} = MVHx_{m,n} + MVHy_{m,n} \quad (1)$$

$$MVHx_{m,n} = \frac{1}{T} \sum_{(i,j)\in LCU_{m,n}} \left| mvx_{i,j} - \frac{1}{T} \sum_{(i,j)\in CU_{m,n}} mvx_{i,j} \right|$$

$$MVHy_{m,n} = \frac{1}{T} \sum_{(i,j)\in LCU_{m,n}} \left| mvy_{i,j} - \frac{1}{T} \sum_{(i,j)\in CU_{m,n}} mvy_{i,j} \right|$$

where T is the total number of MVs assigned to all CUs within the LCU, m and n are the coordinates of the LCU, $mvx_{i,j}$ and $mvy_{i,j}$ are the horizontal and vertical components of motion vector of the CU with the coordinates of (i, j), respectively. Once the motion vector deviation of each LCU is available, we can classify the LCUs as follows:

$$\begin{cases} MVH_{m,n} < T_1 & : \text{LCU} \in \text{region with homogeneous motion} \\ T_1 \le MVH_{m,n} < T_2 & : \text{LCU} \in \text{region with moderate motion} \\ MVH_{m,n} \ge T_2 & : \text{LCU} \in \text{region with complex motion} \end{cases} \quad (2)$$

where T_1 and T_2 are the threshold values defined based on the average motion vector homogeneity (MVH_{ave}) of the whole frame as follows:

$$T_1 = MVH_{ave} \quad (3)$$

$$T_2 = 0.5 \times MVH_{ave}^2$$

$$MVH_{ave} = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (MVHx_{m,n} + MVHy_{m,n})$$

where M and N are total number of LCU rows and columns respectively in each frame. Once the LCUs in the base layer are classified, the motion search range for the co-located LCU in the enhancement layer is adaptively adjusted as follows:

$$SR' = \begin{cases} round(SR/16) & \text{LCU} \in \text{region with homogenious motion} \\ round(SR/4) & \text{LCU} \in \text{region with normal motion} \\ SR & \text{LCU} \in \text{region with complex motion} \end{cases} \quad (4)$$

where SR is the defined motion search range for the base layer and SR' is the adjusted search range of the LCU in the enhancement layer. Note that depending on the class of the LCU in the base layer, the search range of the co-located LCU in the enhancement layer is adjusted, and all the CUs within that LCU will have the same adjusted motion search range setting. As it can be observed from (4), the search range can become quite small, depending on the type of the LCU. Taking into account that there might be several CUs (up to 64) within a LCU, this scheme will significantly reduce the computational cost.

3.2. Early termination mode search

Note that during inter prediction in HEVC, the encoder goes through all three inter prediction modes, first checking for the skip and merge modes, which are computationally less expensive compared to the explicit motion vector encoding process. Our objective here is to implement an early termination (ET) mode-search, so that the encoder does not need to go through all the modes, thus significantly reducing the computational complexity.

The HEVC encoder, in the inter/intra prediction mode selection process, calculates the Rate Distortion (RD) cost for each mode and the one with minimum RD cost is selected. In mode search, if the RD cost of the current to-be-coded CU in the enhancement layer is predicted from the already coded CUs in the base layer and enhancement layer, once the RD cost of a mode is close or equal to the predicted RD cost, the mode search can be terminated. This will

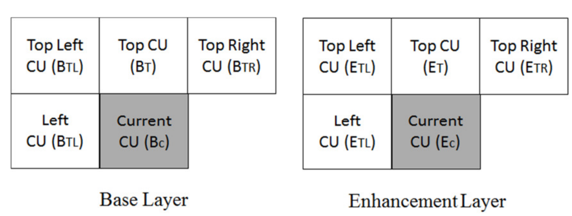

Figure 1 Current CU and its four spatial neighbors of base layer and Enhancement layer

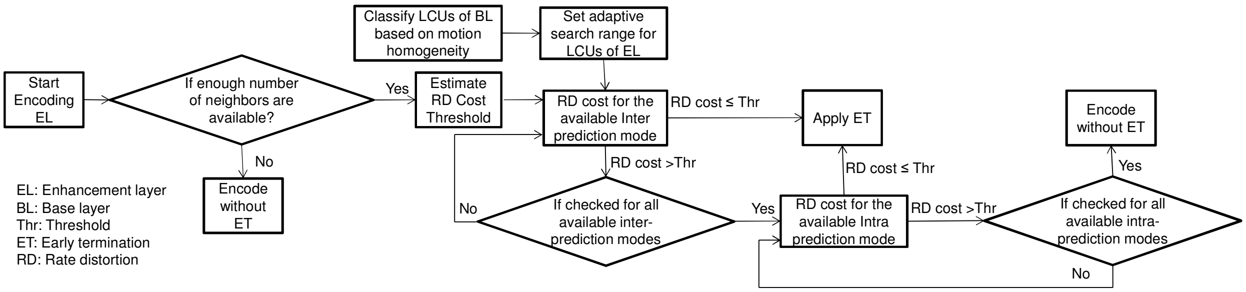

Figure 2 Block diagram of our HEVC-based complexity reduction scheme

significantly reduce the computational complexity. In order to find a prediction for the RD cost of the current CU in the enhancement layer, the RD cost of the already coded CUs in the enhancement layer and that of their co-located CUs in the base layer is utilized. Figure 1 shows an example of the arrangement of the CUs whose information is utilized to predict the RD cost of the to-be-coded CU in the enhancement layer. The neighboring CUs in the enhancement layer are similar to the candidates that HEVC chooses for the merge mode motion search. Inspired by [4], we assume that there is an additive model between the RD cost of the CUs in the enhancement layer and their co-located CUs in the base layer as follows:

$$\mathrm{RDcostE}_{C_{predict}} = (\alpha_0 \frac{\mathrm{RDcostE_T}}{\mathrm{RDcostB_T}} + \alpha_1 \frac{\mathrm{RDcostE_L}}{\mathrm{RDcostB_L}} + \alpha_2 \frac{\mathrm{RDcostE_{TL}}}{\mathrm{RDcostB_{TL}}} + \alpha_3 \frac{\mathrm{RDcostE_{TR}}}{\mathrm{RDcostB_{TR}}})\ * \mathrm{RDcostB_c} \quad (5)$$

where $\mathrm{RDcostE}_{Cpredict}$ is the predicted RD cost of current CU in the enhancement layer, $RDcostB_c$ is RD cost of the co-located CU in the base layer, $RDcostE_T$, $RDcostE_L$, $RDcostE_{TL}$ and $RDcostE_{TR}$ denote the RD cost of the four spatial neighbors of the current CU (see Figure 1 for the arrangement of CUs), $RDcostB_T$, $RDcostB_L$,$RDcostB_{TL}$ *and* $RDcostB_{TR}$ are the RD cost values of the corresponding CUs in base layer, and α_0, α_1, α_2 and α_3 are weighting constants. We compute these weighting constants in the following sub-section.

Once the predicted RD cost for the current CU is available, we define a threshold for early termination of mode search in the enhancement layer as follows:

$$Thr = min(\mathrm{RDcostE_T}, \mathrm{RDcostE_L}, \mathrm{RDcostE_{TL}}, \mathrm{RDcostE_{TR}}, \mathrm{RDcostE_{Cpredict}}) \quad (6)$$

Basically, using this threshold the encoder instead of testing all the modes, it terminates the mode search if the RD cost of a mode is less than the threshold, and selects that mode as the best one. Otherwise, it continues testing other modes till this criterion is met. Note that this scheme is applied to the CUs with at least two already-coded neighboring CUs. In case that the size of co-located CUs in the base layer is larger than the one in the enhancement layer, the RD cost of the co-located CU is normalized to its size and the RD cost used in our calculation is equal to the size of the CU in the enhancement layer times the normalized RD cost. On the other hand, if the co-located area in the base layer is portioned to more than one CU, the sum of the RD costs of these CUs is used.

The threshold defined in (6) is also used in intra prediction of the enhancement layer to further reduce the complexity of encoder. Figure 2 provides a block diagram of our proposed complexity reduction scheme.

3.3. Determining weighting constants

In order to find the proper weighting constants in equation (5), the Linear Least Square method is used. Our objective is to minimize the difference between the predicted RD cost and the real RD cost of the best mode (without using ET) for the current to-be-coded CU in the enhancement layer. Our objective is formulated as follows:

$$argmin_{\alpha_i} |(S - S')^2| \quad (7)$$

where S is a matrix that contains the real RD cost values of the best modes selected by HEVC for the current CU in the enhancement layer ($\mathrm{RDcostE_C}$) divided by $\mathrm{RDcostB_C}$, S' denotes a matrix which contains the predicted RD cost of the current CU ($\mathrm{RDcostE_{Cpredict}}$) divided by $\mathrm{RDcostB_C}$. We can re-write S' as follows:

$$S' = QM = \begin{bmatrix} q_{11}, q_{12}, q_{13}, q_{14} \\ q_{21}, q_{22}, q_{23}, q_{24} \\ . \quad . \quad . \quad . \\ . \quad . \quad . \quad . \\ q_{n1}, q_{n2}, q_{n3}, q_{n4} \end{bmatrix} [\alpha_0, \alpha_1, \alpha_2, \alpha_3]^T \quad (8)$$

where

$$q_{i1} = \frac{\mathrm{RDcosE_{Ti}}}{\mathrm{RDcostB_{Ti}}},\ q_{i2} = \frac{\mathrm{RDcosE_{Li}}}{\mathrm{RDcostB_{Li}}},\ q_{i3} = \frac{\mathrm{RDcosE_{TLi}}}{\mathrm{RDcostB_{TLi}}},\ q_{i4} = \frac{\mathrm{RDcosE_{TRi}}}{\mathrm{RDcostB_{TRi}}} \quad i=1,2,3,\ldots,n$$

Thus, the weighting constants are calculated as follows:

$$M = (Q^T Q)^{-1} Q^T S \quad (9)$$

We use a train dataset (five representative video sequences) to calculate the weighting constants. We code the video streams, record the real RD cost values, calculate the predicted RD cost based on equation (5), and find the weighting constants based on equation (9). In the case that all four spatial neighbors (T, L, TL and TR) are available (see Figure 1), the estimated weighting constants are as follows: $[\alpha_0, \alpha_1, \alpha_2, \alpha_3] = [0.35, 0.32, 0.16, 0.17]$. When

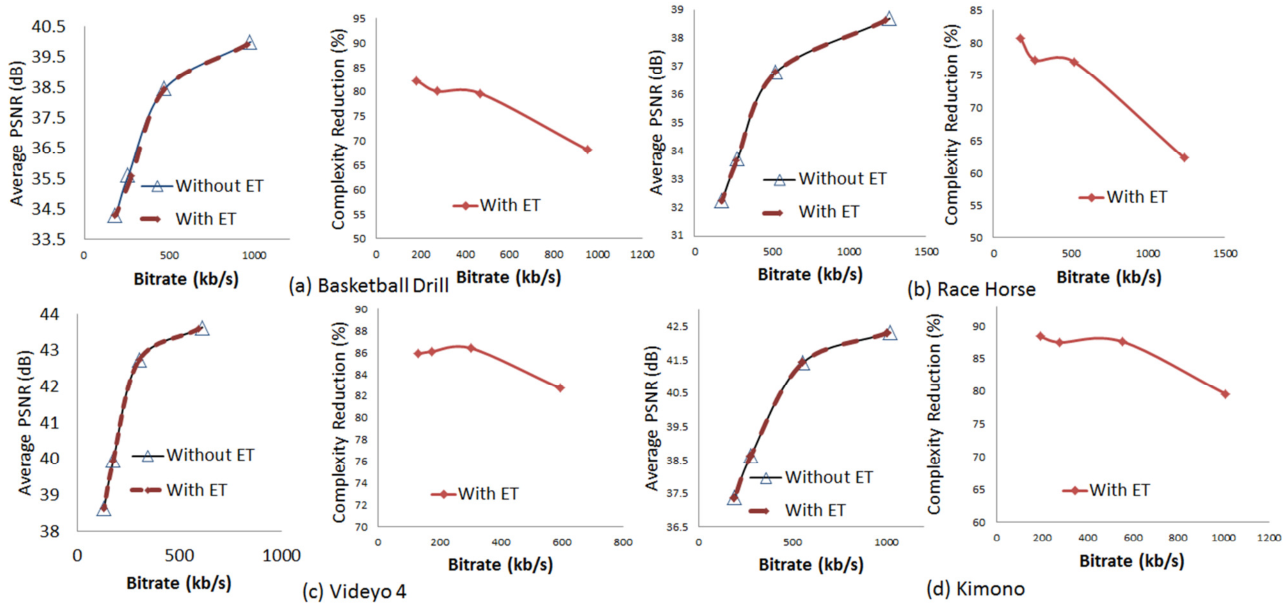

Figure 3 Rate distortion and the complexity reduction comparison

$RDcostE_{TR}$ is not available, $[\alpha_0, \alpha_1, \alpha_2, \alpha_3] = [0.4505, 0.4055, 0.1404, 0]$. If $RDcostE_{TL}$ is not available – which means that the $RDcostE_L$ is not available either - we use two upper neighbors to predict the RD cost, and the weighting constants are $[\alpha_0, \alpha_1, \alpha_2, \alpha_3] = [0.5194, 0, 0, 0.4806]$. The weighting constants of the top and left neighboring CUs when available are larger than the others, denoting that they are more correlated with the current CU.

4. RESULTS AND DISCUSSION

In our experiment, four test videos from the data set provided by MPEG for HEVC Call for proposals [5] were used (see Table I). Note that the train data used for finding the weighting constants was not included in our test videos. Our method was implemented on the HM 5.2 software version of HEVC, with Random Access High Efficiency (RA-HE) configuration (hierarchal B pictures, GOP length 8, ALF, SAO and RDOQ were enabled) [6]. The QPs used for the base layer and enhancement layer (QpB , QpE) are as follows: (32, 23), (32, 28), (38, 32) and (40, 36).

Figure 3 shows the RD curves of the test video sequences, where bitrate is the average total bitrate of the scalable video stream (base layer+ enhancement layer). As can be observed, our scheme minimally hampers the bitrate.

Figure 3 also illustrates the percentage of mode-search complexity reduction for each stream. In our study the complexity is computed based on the number of times the encoder searches for the best mode. For example, for inter prediction, for every search point the complexity measure is equal to 1. For the skip mode, the complexity measure is equal to 1. For the Merge mode the complexity is up to 5, depending on the available candidates. By adding up these complexity values when coding the enhancement layer, we find the total complexity measure. For intra modes, we also compute the number of candidates which the encoder checks to encode a CU. As can be seen from the complexity curves in Figure 3, our proposed scheme substantially reduces the computational cost without hampering the total bitrate.

Table 1 summarizes the effect of our scheme in terms of bitrate, PSNR and complexity for each stream. As can be observed, our scheme reduces the complexity up to 85.77% at a maximum cost of 3.51% bitrate increase.

Table I Impact of the proposed scheme on Bitrate, PSNR and Complexity

Name	Resolution, Frame Rate (fps)	Average PSNR Degrade	Average Bitrate Increase	Average Complexity Reduction
BasketballDrill	832x480, 50	0.082 dB	3.51%	77.57%
Kimono	1920x1080, 24	0.005 dB	0.036%	85.77%
Race Horse	832x480, 30	0.012 dB	0.02%	74.35%
Videyo 4	1280x720, 60	0.016 dB	0.26 %	85.27%

5. CONCLUSIONS

In this paper, we proposed a content adaptive complexity reduction scheme for SNR/Quality scalable HEVC. In our scheme the information of the base layer is utilized to facilitate the inter prediction and intra prediction mode selection process in the enhancement layers by avoiding redundant computations.

Performance evaluations show that our approach results in significant HEVC/SVC coding complexity reduction (up to 85.77%) while minimally hampering the overall bitrate.

5. REFERENCES

[1] A. H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17 (9), pp. 1103-1120, Sep. 2007.

[2] J.R. Ohm, "Scalable Video Coding" (Presentation Slides), Shenzhen workshop, Oct. 2007.

[3] M. T. Pourazad, C. Doutre, M. Azimi, and P. Nasiopoulos, "HEVC: The New Gold Standard for Video Compression," IEEE Consumer Electronic Magazine, vol.1 , issue 3, pp. 36-46, July 2012.

[4] L. Shen, Z. Zhang, "Content adaptive motion estimation algorithm for coarse-grain scalability", IEEE Trans. on Image Processing, pp. 2582-2591, Vol. 21, No.2, March 2012.

[5] ISO/IEC JTC1/SC29/WG11,"Joint Call for Proposals on Video Com-pression Technology", N11113, January 2010.

[6] F. Bossen, "Common test conditions and software reference configura-tions", JCTVC-F900, July 2011.