Adaptive Mode Selection for Low Complexity Enhancement Layer Encoding of SHVC

Kazuki Kuroda[1], Takafumi Katayama[1], Tian Song[2], and Takashi Shimamoto[2]
[1]Electrical and Electronic Engineering, Graduate School of Engineering,
Tokushima University, Minami-Jyosanjima 2-1, Tokushima City, 770-8506, Japan
Email: {k.kuroda,katayama}@ee.tokushima-u.ac.jp
[2] Computer Systems Engineering, Institute of Technology and Science, Graduate School of Engineering,
The University of Tokushima, Tokushima, 770-0814, Japan
E-mail : {tiansong,simamoto}@ee.tokushima-u.ac.jp

Abstract: In this work, we proposed a low complexity algorithm using the adaptive mode selection based on the sum of squared errors of the correlated pixels and motion vector. To achieve low computation complexity, the temporal correlation and the spatial correlation is used for the mode decision. The proposed algorithm is evaluated by the reference software of SHVC. The simulation results show that the proposed algorithm can achieve over 20% computation complexity reduction comparing to the original SHVC algorithm.

Keywords—**SHVC, Merge mode, Motion estimation, Complexity reduction, Coding efficiency**

1. Introduction

With the recent growth of internet and communication applications, increasing requirement for scalable high efficiency video coding (SHVC) encouraged the research for next generation coding standard. A specification and reference software about scalable extension of H.265/HEVC have been developed by the joint video team (JVT) for SHVC[1][2]. The objective of scalable video coding is to enable the generation of a unique bitstream that can adapt to various bit-rate, transmission channel and display capabilities. Taking into account that the complexity of HEVC codec is higher than other existing standard codecs, its scalable extension is expected to be simpler. However, the computation complexity is increased by introducing some new modes.

The HEVC standard inherits the well-known blockbased hybrid coding architecture of H.264/AVC[1]. SHVC as well as HEVC employs a Quad-tree Based Partitioning[2]. The partitioning of inter prediction is supported by Coding Unit (CU) sizes from 64x64 down to 8x8 samples. All partitions for inter prediction are represented by motion parameters obtained from three coding modes, namely motion estimation (ME mode), merge mode, or skip mode. Moreover, ILRP (Inter Layer Reference Prediction) mode is the prediction mode that has been newly implemented in SHVC. Finally, CU size is determined by comparing RDcost (Rate Distortion-cost). Inter mode of H.265/HEVC adopted (1)ME that transmit difference motion vector. (2)merge mode(skip mode) that does not transmit to this. ME predict current frame using coded frame. Disadvantage of this mode is the exahsted time for motion search. The merge mode selects the most suitable motion information from the neighboring CU. Then merge mode transfer index to indicate only motion information candidate to decoder. The skip mode is used as the information of the 64x64 block, so skip mode has lowest complexity with other modes.
ME uses BL that is up-sampled to the coding of EL as the reference frame. In order to achieve high coding efficiency, it enables the prediction between different resolutions.

The proposed algorithm dedicated to random access mode(RAM). Fig. 1 shows a schematic diagram of RAM. Similar to the temporal scalability feature in the H.264/MPEG-4 AVC scalable video coding (SVC) extension, RAM is implemented to HEVC. RAM can generally be enabled by restricting motion compensated prediction to reference pictures with a temporal layer less than or equal to the temporal layer of the picture to be predicted. RAM provides not only temporal scalability but also an improved coding efficiency compared to classical B picture structure. RAM have coding high efficiency by using pictures future as well as picture past. On the other hand, RAM induces delay by replacing the coding sequence. Group of picture(GOP) is constituent units of the sequence that composed of a plurality pictures. A GOP is consisted with eight frame.

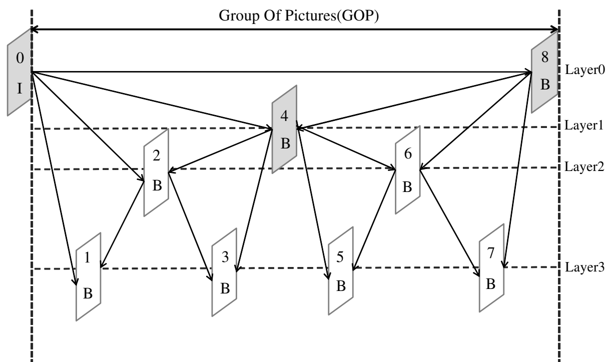


Figure 1. Random access mode structure

To realize a display method corresponding to the request of the terminal, SHVC is composed of base layer (BL) and enhancement layer (EL). However, there is a problem that the computation complexity increased by addition of the new mode. Therefore, our study objective is the computation complexity reduction. In the next section, we verify the ratio of mode selection in EL.

2. Verification

2.1 Activity ratio of merge mode

In previous work, it is proved that the coding efficiency can be improved by using merge mode effectively [3]. This previous work also show the fact that merge mode is selected with high probability. Therefore, since merge mode has lower complexity compared to traditional ME, the computation complexity can be reduced by selecting merge mode more efficiently. To efficiently achieve low complexity, we evaluate the probability that the merge mode and skip mode are used as best mode. Fig. 2 show the activity ratio of merge mode and skip mode in one of the sequence.

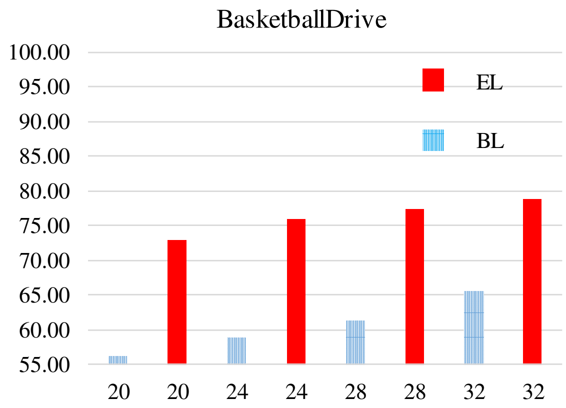


Figure 2. Activity ratio of merge mode and skip mode

From the results shown in Fig. 2, merge mode and skip mode are used over 70% in EL. Moreover, there are similar rate in other sequences. Accordingly, we can achieve low complexity by early selecting merge mode and skip mode before the mode decision by RDcost.

2.2 Temporal and spatial correlation evaluation

In this work, to efficiently select the optimal mode, the proposed algorithm uses the temporal and spatial correlations between frames and layers. Fig. 3 illustrates an example of the temporal and spatial correlations for a scalable video codec with two quality layers.

In our proposed algorithm, ME and merge mode is selected when temporal correlation is strong. On the other hand, if spatial correlation is stronger ILRP mode will be selected. It can be considered as a typical situation on a scene change. To evaluate the correlation between continuous frames, the proposed algorithm evaluate the temporal correlation by using the sum of squared errors (SSE).

Generally, the best mode in SHVC is selected by a evaluation function which calculates the RDcost. In this work, a SSE based evaluation function is used to evaluate the strength of the correlation.

$$SSE = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (Error(x,y))^2 \quad (1)$$

(x, y) :pixel coordinates, $Error(x, y)$: pixel difference at (x,y), N :block size

SSE is used as the threshold value to judge the correlation direction.

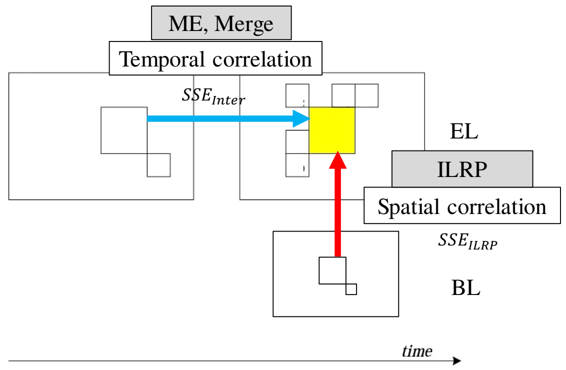


Figure 3. Illustration of the temporal and spatial correlation

2.3 Distribution of SSE

It is neccesary to get the distribution of SSE values to make use of the SSE to reduces the candidate mode. In our proposed algorithm, the three kinds of SSE is calculated when 64x64 block is coded. SSE_{Inter} is define as the value which is calculated in inter mode. SSE_{ILRP} is define as the value which is calculated in ILRP mode. SSE_{Skip} is define as the value which is calculated in skip mode. The difference value between SSE_{Inter} and SSE_{ILRP} is used to calculate the correlation strength of temporal and spatial. We use an average value of SSE_{diff} of one GOP. Fig. 4 shows the average value of SSE_{diff} of 5 selected sequences by histogram.

$$SSE_{diff} = SSE_{Inter} - SSE_{ILRP} \quad (2)$$

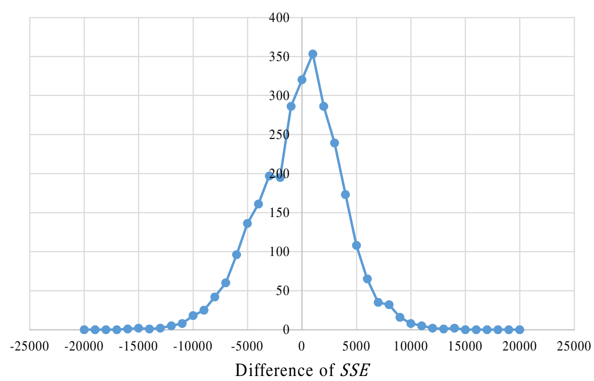


Figure 4. Distribution of SSE_{diff}

From the evaluation results, it is clear that when when the SSE_{diff} value is large the spatial correlation is strong than temporal correlation. On the other hands, temporal correlation is stronger when the SSE_{diff} value is small.

3. Propose Algorithm

3.1 Adaptive mode selection

The proposed algorithm is used to the adaptive mode selection by determining the threshold value which is evaluated in section 2.1 and 2.3. Candidate mode less than 64x64 block is reduced by SSE_{diff} before best mode is determined by RD-cost. As Fig. 2 shows, ME process account for about 30%. Therefore, in this work the threshold is determined by this evaluation result. Since ME is applied when temporal correlation is very high, in this work ME is applied when SSE_{diff} is less than -3000. Similarly, since merge mode is performed to about 70%, merge mode is performed when SSE_{diff} is less than 5000 and bigger than -3000. When SSE_{diff} is more than 5000, ILRP mode is performed because spatial correlation is very strong.

3.2 Mode selection by motion vector

In the proposed algorithm, the candidate mode is reduced by using the motion vector as well as SSE. Generally if the motion vector is large, finding similar block in reference frame is difficult. Thus the motion vector can indicate the performance of the prediction. On the other hand, if the background as less moving image is encoded, the motion vector is zero. If the motion vector is zero in the 64x64 block, ME with small block size doesn't need to be processed. In this case, ME with small block size is applied. Therefore, our proposed algorithm is combined with SSE and the motion vector to improve the prediction.

3.3 Prediction mode in RAM

As shown in Fig. 1, RAM is a hierarchical structure consisted of layer0 to layer3. We have tried to implement the proposed adaptive mode selection algorithm to all layers. As a result, bitrate significantly increased in layer0 and layer1. In layer0 and layer1, we also find that small block size are often selected due to the larger temporal distance. On the other hand, in layer2 and layer3 ME is processed using close reference frames and motion vector is often small. Therefore, we always perform ME to layer0 and layer1 and adaptively select mode using the correlation between frames in layer2 and layer3.

3.4 Overall process

The flowchart of overall proposed algorithm is shown in Fig. 5.

(1) When 64x64 block is encoded at each mode (skip mode, inter mode, ILRP mode), and calculated the SSE.

(2) The motion vector of each mode in 64x64 block is saved.

(3) Difference of SSE between the inter mode (ME and merge mode) and ILRP mode is calculated.

(4) If SSE_{diff} is less than 5000 and SSE_{ILRP} not equal to SSE_{Skip}, ILRP mode and skip mode are selected, go to (8). Otherwise, go to (5).

(5) If SSE_{diff} is more than -3000 and $AllMv$ not equal to 0, merge mode and skip mode are selected, go to (8) Otherwise, go to (6).

(6) If SSE_{diff} is less than -3000 and $AllMv$ not equal to 0, all of mode are selected, go to (8) Otherwise, go to (7).

(7) If SSE_{ILRP} equal SSE_{Skip} in (4) or $AllMv$ equal to 0 in (5),(6), skip mode is selected, go to (8).

(8) RDcost is calculated using the mode determined from (4) to (7).

(9) Next to 64x64 block, go to (1).

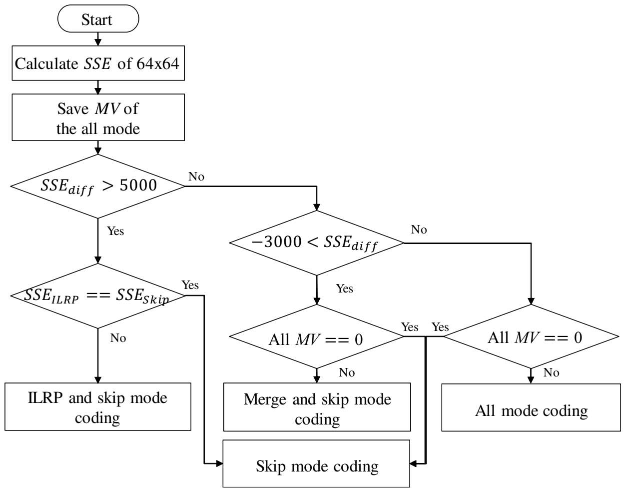


Figure 5. Proposed flowchart

4. Simulation Results

The proposed algorithm is implemented in reference software SHM6.0[2]. In this simulation, two-layer spatial scalability model with the resolution Half HD and Full HD is used. The computation complexity reduction is evaluated with different QP of 20, 24, 28, 32. Time saving is defined as

$$TS = \frac{T_{proposed} - T_{SHM}}{T_{SHM}} \times 100(\%) \quad (3)$$

where T_{SHM} is the encoding time of the original reference software and $T_{proposed}$ is that of the proposed algorithm. Bitrate and PSNR of proposed algorithm are compared with original reference software and defined as

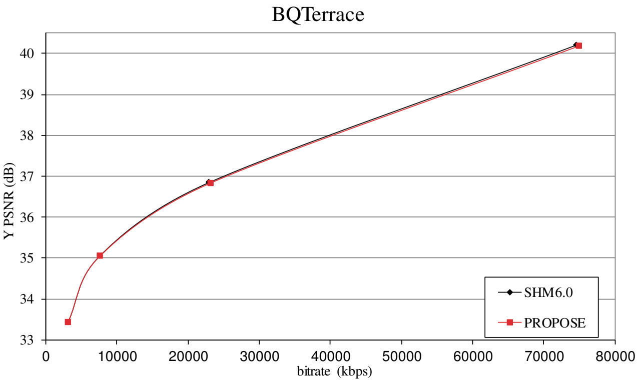
$$\Delta Bitrate = \frac{bitrate_{proposed} - bitrate_{SHM}}{bitrate_{SHM}} \times 100(\%) \quad (4)$$

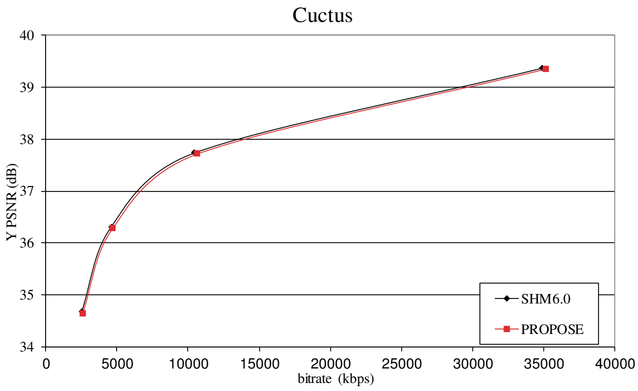
$$\Delta PSNR = PSNR_{proposed} - PSNR_{SHM}(dB) \quad (5)$$

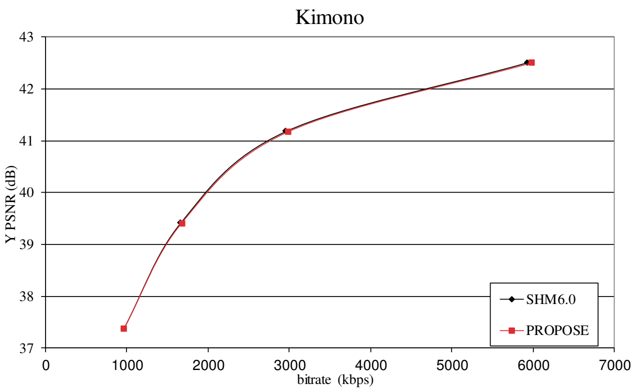
The simulation results are shown in Table 1. The coding efficiency is evaluated by R-D curve (Fig. 6). As the simulation result shows, the proposed algorithm can achieve 20% computation complexity reduction of total encoding time. As shown in fig. 6, the coding efficiency is reduced in high QP. The proposed algorithm has been proposed for Half HD and Full HD, Low QP. Because FullHD is typically used in SHVC. We must propose the threshold value at can correspond to various QP value and resolutions, such as 4k and 8k in the future work.

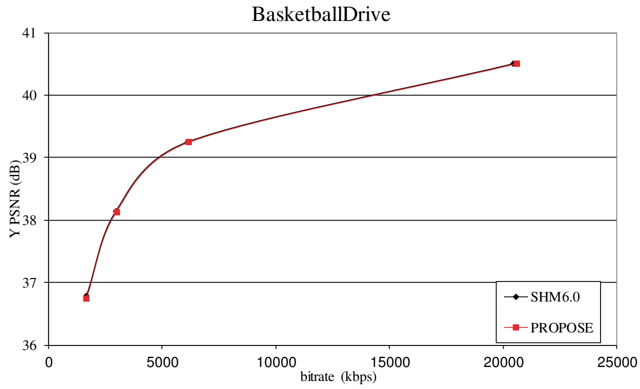
Table 1. Simulation Results

Sequence	QP=20 δ Bitrate (%)	QP=20 δ PSNR (dB)	QP=20 TS (%)	QP=24 δ Bitrate (%)	QP=24 δ PSNR (dB)	QP=24 TS (%)	QP=28 δ Bitrate (%)	QP=28 δ PSNR (dB)	QP=28 TS (%)	QP=32 δ Bitrate (%)	QP=32 δ PSNR (dB)	QP=32 TS (%)
BQTerrace	0.52	-0.02	-24.25	0.98	-0.01	-17.75	0.26	0.00	-11.97	0.39	0.00	-11.37
Cactus	0.70	-0.02	-27.02	1.69	-0.02	-27.79	1.27	-0.03	-31.46	0.78	-0.04	-29.81
Kimono	0.93	-0.01	-31.52	1.05	-0.01	-25.44	0.80	-0.02	-22.13	0.72	-0.01	-16.59
BasketballDrive	0.70	0.00	-29.60	0.92	-0.01	-20.92	0.78	-0.02	-17.05	0.81	-0.03	-17.51
ParkScene	0.75	-0.01	-18.19	0.63	-0.01	-18.19	0.67	-0.01	-13.08	0.55	-0.01	-13.16
Average	0.72	-0.01	-26.12	1.06	-0.01	-22.02	0.76	-0.02	-19.14	0.65	-0.02	-17.69









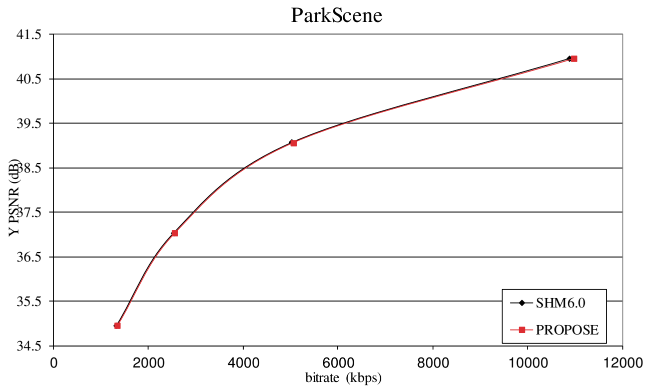


Figure 6. R-D curves: SHM6.0 vs. the proposed algorithm

5. Conclusion and Future Works

In this paper, we proposed algorithm to reduce the complexity by early selecting a coding mode candidate. As the simulation result shows, the proposed algorithm can achieve over 20% reference frames saving compared to the original SHVC algorithm. In our future work, we will improve the proposed algorithm to avoid deterioration of the coding efficiency.

References

[1] JCT-VC, High Efficiency Video Coding (HEVC) Test-Model 10 (HM10) Encoder Description, JCTVCL1002, 12th MeetingGeneva, CH,14.23 Jan. 2013.

[2] ISO/IEC and JTC1/SC29/WG11, Scalable HEVC (SHVC) Test Model6 (SHM 6), N14419, Apr. 2014.

[3] Xiem Hoangvan, Fernando Pereira, · Improving enhancement layer merge mode for HEVC scalable extension,” IEEE Picture Coding Symposium, 2015.

[4] H. R. Tohidypour, ”Adaptive search range method for spatial scalable HEVC,” · IEEE International Conference on Consumer Electronics (ICCE), 2014.