IEIE Transactions on Smart Processing and Computing

Early SKIP Mode Decision for HEVC with Hierarchical Coding Structure

Xiem HoangVan[1,2]* Minh Dinh Bao[1,2], and Byeungwoo Jeon[3]

[1]Faculty of Electronics and Telecommunications, VNU - University of Engineering and Technology

[2]Vietnam National University, Hanoi, Vietnam

[3]Department of Electrical and Computer Engineering, Sungkyunkwan University, Korea

* Corresponding Author: Xiem HoangVan <xiemhoang@vnu.edu.vn>

Received: July 30, 2020; Revised August 26, 2020; Accepted September 14, 2020;

Abstract: To meet real-time video communications, an investigation of early SKIP mode decisions is still important in many practical applications of High-Efficiency Video Coding (HEVC) encoders. On the other hand, in most current state-of-the-art early SKIP mode decision methods, the temporal layer index (TId), which is commonly used in the hierarchical coding structure, has not been fully exploited. In this regard, this paper proposes a novel early SKIP mode decision method for a HEVC encoder using TId-based rate-distortion (RD) modeling. In the proposed method, two new SKIP mode-checking conditions were introduced based on statistical analysis on TId and the RD-cost correlation. The experimental results showed that the proposed early SKIP mode decision method outperformed the relevant state-of-the-art HEVC complexity reduction methods significantly, with an encoding time saving of approximately 47% while having a BD-rate loss of only 0.45%.

Keywords: High Efficiency Video Coding, SKIP mode early decision, temporal layer index, rate-distortion optimization

1. Introduction

The demand for High Definition (HD) and Ultra-HD (UHD) video in multimedia services, such as television broadcasting or video streaming, has increased considerably. To address this requirement, the High-Efficiency Video Coding (HEVC) standard was introduced by the Joint Collaborative Team on Video Coding (JCT-VC) in 2013 [1]. HEVC was designed based on the hybrid prediction and transform-based coding paradigm commonly adopted in previous video standards, such as H.262/MPEG-2 and H.264/AVC [2]. On the other hand, while H.264/AVC divides a video frame into macroblocks with a maximum size of 16×16 [3], HEVC can have a coding unit (CU) with larger sizes, e.g., 64×64. The flexibility of using variable and large block sizes is one of the main advances that help HEVC achieve a 50% bitrate saving compared to the H.264/AVC [1].

In HEVC, a hierarchical coding structure was commonly used in a Random Access (RA) configuration [1] to achieve high compression performance. In this case, the frame at a higher temporal layer index (TId) can access the decoded frames at lower TIds. For each coding frame, from the largest coding unit (CU) size (64×64), the quad-tree partition will recur until it reaches the smallest CU size (8×8) to configure suitably sized blocks and their optimal prediction modes. Each CU consists of a prediction unit (PU) [4] and a transform unit (TU) [4], which contains information on the predictions and transforms, respectively.

In HEVC inter coding, one PU can have various sizes, such as 2N×2N, N×N, 2N×N, N×2N,… while a TU can encode with a block size from 4×4 to 16×16. This hierarchical coding structure helps increase the accuracy of deciding the partitioning and prediction mode. To achieve the optimal coding, a rate-distortion optimization (RDO) process is applied to check every possible CU size and PU mode. This process is described in Fig.1. The RDO [5] cost is commonly calculated as follows:

$$J = D + \lambda \times R \qquad (1)$$

where J is the RD-cost. R is the bitrate based on current coding decisions. λ is the Lagrange multiplier corresponding to QP (Quantization Parameter), and D is

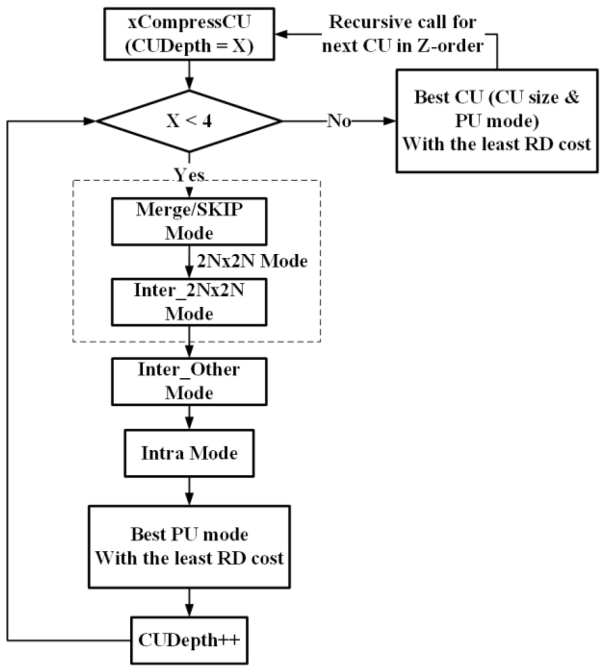


Figure 1. Optimal decision process of HEVC

the distortion associated with the current prediction mode. HEVC uses the sum of the squared error (SSE) between the current PU and its matching block to measure the distortion *D* [5].

As shown in Fig.1, although the optimal CU size and PU mode are reached (such as the CU size is 32×32 and PU mode is SKIP), HEVC still executes the checking process for the remaining cases. This process naturally leads to high computational complexity. Therefore, reducing the encoding time of HEVC is essential for real-time video communication applications.

In HEVC, the SKIP mode usually takes up a high probability of being selected as the best PU mode [6]. Because the residual bit cost of the SKIP mode is zero, this mode carries the lowest bit cost for coding. Another characteristic of SKIP mode is always checked at the first step of the RDO checking process. Therefore, considerable encoding time can be saved if the SKIP mode can be determined to be the best PU mode early and accurately because the remaining modes in the RDO checking process can be skipped. This work also brings a low bit cost after encoding the video. Therefore, an early SKIP detection approach has been actively investigated for HEVC complexity reduction [6-12]. To the best of the authors' knowledge, none of these methods exploit the coding frame position in a group of pictures (GOP), commonly represented by the TId. This paper discusses the impact of the TId in a HEVC hierarchical coding structure on the distribution of the SKIP mode. Two novel SKIP mode decision conditions are introduced based on the obtained relationship.

The remainder of this paper is organized as follows. First, the related work is detailed in Section 2. Section 3 outlines the proposed early SKIP mode decision method, and Section 4 presents the experimental results and the validity of the proposed method. Finally, this paper is concluded in Section 5.

2. Related Work

Many approaches have been introduced to reduce the encoding time of HEVC [6-22]. These studies can be divided into three main classes. The first class is making a fast decision for the best PU mode. Several methods in this approach have been implemented as a coding tool option in the HEVC reference software (HM), such as the Early Skip Decision (ESD) algorithm [6] and Coded block flag (CBF) Fast Method (CFM) [7]. In ESD [6], the information of the differential motion vector (DMV) and CBF are used to detect the SKIP mode early. The CFM method in [7] will terminate the remaining checking of the PU modes when all CBF values are zero.

The proposals reported thus far [8-10] constructed their algorithms based on the rate distortion (RD-cost) obtained from the reference frames. In references [8] and [9], the RD-cost of the SKIP mode was compared with a pre-defined threshold, and the SKIP mode was selected if its RD-cost was less than a certain threshold. The threshold was estimated based on the correlation between the current PU and its neighboring PU or based on the QP parameter, respectively. Hu et al. [10] also used a threshold to detect the SKIP mode early. In their study, the CU was partitioned into a high distinction region and a low distinction region. A Bayesian decision rule was used to find the threshold for each region. Proposal [11] divided the CUs into two categories, rarely used and frequently used. A direct method and a unimodal stopping model (USM) were then used to detect the SKIP mode for the two categories early, respectively. The temporal and spatial correlation between the current CU and its neighboring and its co-located CUs is used in reference [12]. This approach utilized the PU mode information of encoded CUs to find the suitable mode and wipe out the unlikely modes.

Pan et al. [13] proposed a method for selecting the merge mode when the prediction residuals were transformed and quantized into zeros, and the motion vector length values were zero, or the CU parent of the current CU used the merge mode as the best mode. In reference [14], the RD-cost distributions of the SKIP mode, the merge mode and Inter_2Nx2N mode for various CU depths were analyzed statistically to determine the early SKIP mode decision method.

The second approach is the early decision of the CU size. HM reference software has supported a fast option for this approach based on the CU early termination [15]. This method will terminate the quad-tree division when the current CU uses the SKIP mode. Several methods used machine learning to decide whether the current CU is divided into four sub-CUs or not [16-18]. In reference [16], a support vector machine (SVM) was used to train the model. After that, another study [17] trained a model based on multi-class learning, while the CNN method was used elsewhere [18]. In reference [19], the SAD value between current CU and its co-located CU was used to decide when SAD is less than a threshold calculated on a QP's function. The current CU will not be divided anymore. Shen et al. [20] proposed a fast decision for a current CU based on the CU size information of its neighbor CUs and its co-located CU.

Finally, the third approach determines both the CU size and PU mode. For example, the proposed method in reference [21] was used both for RD-cost and SAD value to compare with two thresholds, respectively. An early decision will be made when the threshold condition is satisfied. In [22], the RD-cost value of the SKIP mode and CU size information were modeled based on a Neyman-Pearson decision rule [23].

To the best of the authors' knowledge, the impact of the temporal layer index has not fully exploited in the relevant HEVC early SKIP mode decision process. To address this problem, the next section presents a novel early SKIP mode decision method designed specifically for the HEVC with a hierarchical coding structure.

3. Proposed Method

To exploit the TId impact in HEVC early SKIP mode decisions, this Section will first analyze the distribution of using the SKIP mode in HEVC for frames at several TIds. Two stopping conditions for early SKIP mode will then be introduced.

3.1 Observation and statistics

Fig. 2 gives an example of the HEVC with a hierarchical coding structure, i.e., the random access (RA) configuration. A group of pictures (GOP) of 16 is typically used [24]. The frame position can be represented by the picture order count (POC) and the temporal layer index (TId), where $TId = \{0; 1; 2; 3; 4\}$.

As shown in Fig. 2, the temporal distance (TD) between the coded picture and its references naturally depends on the TIds. A larger TD can be observed for the picture at a smaller TId and vice versa. This may directly affect the distribution of coding modes in HEVC encoding.

Therefore, it is reasonable to study the effects of the TId on the percentage of mode selection in a picture with different TIds. To address this, this study examined the distribution of the SKIP, Merge, Inter ME (motion estimation), and Intra modes commonly used in HEVC when coding four selected video sequences, *BlowingBubbles*, *BasketballPas,s* *BQMall*, and *RaceHorses* obtained in [25].

From the obtained results shown in Figs. 3 and 4, the following conclusions can be made:

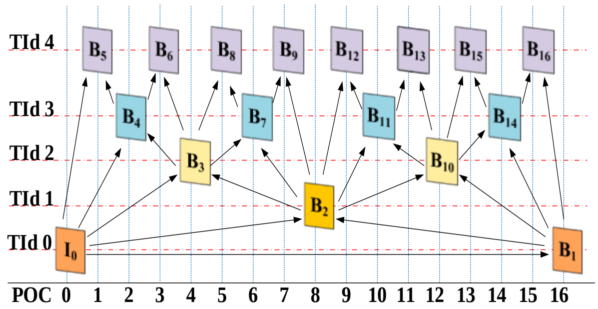


Figure 2. HEVC hierarchical coding structure

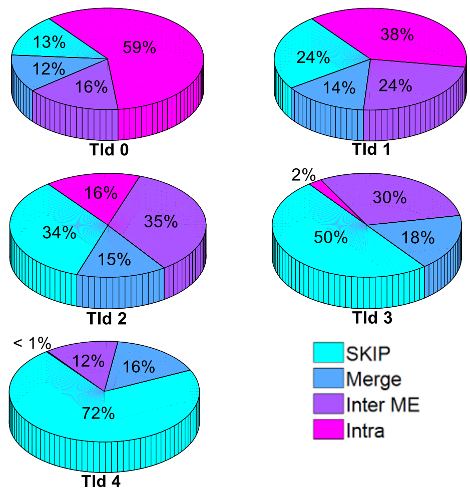


Figure 3. Mode distribution for various TIds, BasketballPass QP = 27

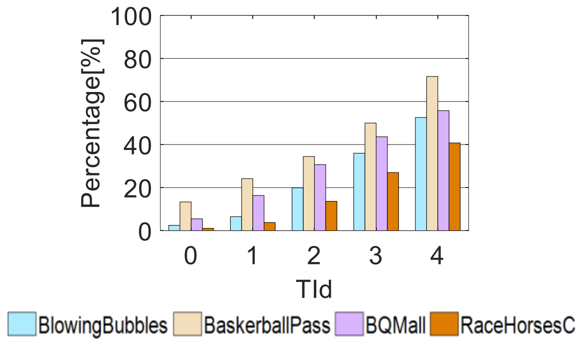


Figure 4. Percentage of selecting SKIP mode in various TIds

- The percentage of SKIP mode increases consistently with increasing TId value. A higher TId indicates a higher percentage of selecting SKIP as the best mode;
- For frames with TId > 2, the SKIP mode is frequently selected as the best mode;
- The percentage of using SKIP mode also varies with the content of the video. Hence, besides the TId information, the content of the video strongly affects the percentage of selecting the SKIP mode. In this case, videos with simple motion like *BasketballPass* will likely achieve many SKIP modes.

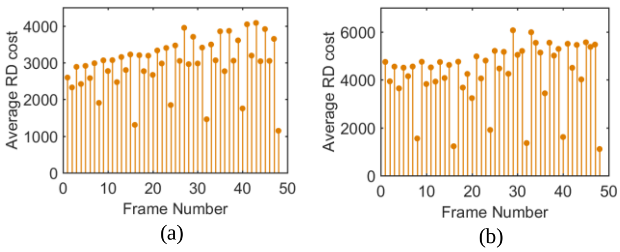


Figure 5. The average RD-cost of the SKIP mode in each frame in sequence (a) BQMall, (b) Racehorses

In addition, according to reference [10], the average RD-cost difference between SKIP mode and other PU modes is large, and the SKIP mode can be distinguished by its RD-cost. The average RD-cost of SKIP mode is almost equal in frames of the same level of TId. This observation was confirmed in Fig.5.

From the above observation, it is reasonable to design an early SKIP mode decision algorithm for a HEVC based on the TId and the RD-cost of reference pictures. The proposed algorithm will naturally follow the optimal RDO process in HEVC but with some modifications, as detailed in the next sub-section.

3.2 Early SKIP mode conditions

Fig. 6 illustrates the proposed fast HEVC inter coding algorithm where two early SKIP mode conditions are included. Details of these conditions are presented as follows.

First, as specified in the RDO process (see Fig.1), the SKIP mode can be selected for coding a CU if the corresponding RD-cost is smaller than the other modes, i.e., Merge, Inter ME, and Intra. On the other hand, this process will require considerable time to check all the coding possibilities. To avoid this, this paper proposes to select the SKIP early as the best mode if the two following conditions are satisfied:

Condition 1: The best mode will be SKIP when its RD-cost is smaller than the RD-cost of the Merge mode and the estimated RD-cost function for the remaining modes, J_{TId}, which can be modeled by a function of TId and RD-cost values of all blocks encoded with SKIP mode in the reference picture, J_{TId-1}. This means that

$$(J^{SKIP}_{Current} < J^{Merge}_{Current})\&(J^{SKIP}_{Current} < f(TId, J_{TId-1})) \quad (1)$$

Under this condition, to achieve a highly accurate $f(TId, J_{TId-1})$, this study examined the correlation between the average RD-cost of SKIP mode of the frame in the TId level (J_{TId}) and the RD-cost of its associated reference in TId-1 level (J_{TId-1}) as well as the TId values. Fig.7 presents the RD-cost correlation and TId influence for the samples obtained from *BlowingBubbles*, *BasketballPass*, *BQMall*, and *RaceHorses*.

As shown in Fig.7, $\log(J_{TId})$ can be estimated from $\log(J_{TId-1})$ and TId through a plane equation. This means that

$$\log(J_{TId}) = \alpha \times \log(J_{TId-1}) + \beta \times TId + \gamma \quad (2)$$

Here, α, β, and γ are plane parameters that can be obtained practically using the first 50 frames extracted from the above sequences. In this paper, (α, β, γ) was chosen as follows: $\alpha = 1.04$; $\beta = -0.0457$; $\gamma = 0.2041$.

In this case, to find a functional form $f(TId, J_{TId-1})$, both sides of (2) can be rewritten as a power of base 1. This means that

$$J_{TId} = K \times J_{TId-1}{}^{\alpha} \quad (3)$$

Here, K is given as

$$K = 10^{\beta \times TId + \gamma} \quad (4)$$

Therefore, with (α, β, γ) chosen above, $f(TId, J_{TId-1})$ can be written as

$$f(TId, J_{TId-1}) = (1.6 \times 0.9^{TId}) \times J_{TId-1}{}^{1.04} \quad (5)$$

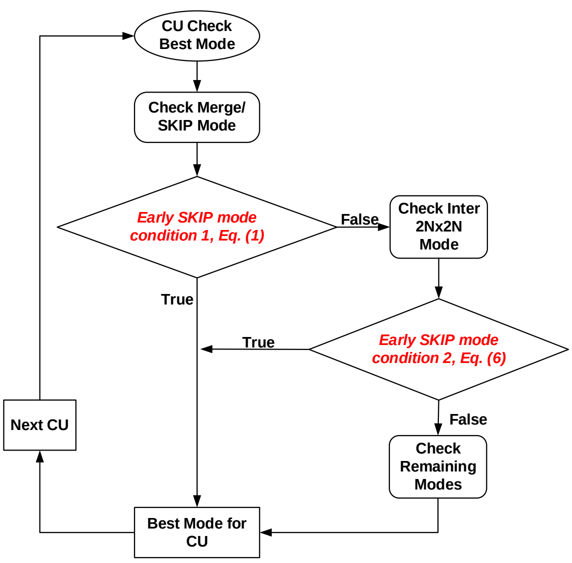


Figure 6. Proposed fast HEVC inter coding (CU: Coding Unit)

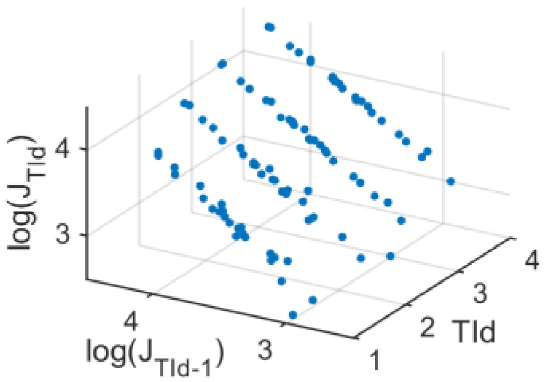


Figure 7. Illustration of RD-cost (J) correlation and TId for several test samples

Table 1. Specification of test video sequences

Class	Sequences	Frame Rate (Hz)	Frame to be encoded
Class A (2560×1600)	Traffic	30	150
	PeopleOnStreet	30	150
Class B (1920×1080)	BasketballDrive	50	500
	BQTerrace	60	600
	Cactus	50	500
	Kimono	24	240
	Parkscene	24	240
Class C (832 × 480)	BasketballDrill	50	500
	BQMall	60	600
	PartyScene	50	500
	RaceHorses	30	300
Class D (416 × 240)	BasketballPass	50	500
	BlowingBubbles	50	500
	BQSquare	60	600
	RaceHorses	30	300
Class E (1280×720)	FourPeople	60	600
	KristenAndSara	60	600

Condition 2: As discussed above, for TId > 2, the SKIP mode was highly selected as the best mode. To explore this fact, this paper proposes to select the SKIP early as the best mode if the coding block belonging to the frames at TId > 2 and the RD cost of the SKIP mode is smaller than that of the 2N×2N. This means that

$$(J_{Current}^{SKIP} < J_{Current}^{2N\times 2N})\&(TId > 2) \qquad (6)$$

Under this condition, when the RD-cost of the SKIP mode is smaller than the RD-cost of 2N×2N ME mode, and the TId is larger than two, a SKIP mode will be selected as the best mode. The next section will discuss the accuracy of each early SKIP condition and the overall encoding time-saving achievement.

4. Performance Evaluation

This Section presents the overall assessment of the proposed early SKIP mode decision method for HEVC. First, the test conditions are described in detail. The performance of the overall method regarding the contribution of each early SKIP condition to the overall performance is analyzed.

4.1. Test conditions

To verify the validity of the proposed method, the method was implemented on top of the HEVC test model (HM) 16.20 reference software [26]. Test video sequences were selected following the HEVC common test condition [25], as summarized in Table 1. The RA [25] with four QPs (22, 27, 32, 37) was used for the coding configuration. The Bjontegaard Delta Bitrate (BD-Rate) [27] and Time-Saving (TS) metrics were used to assess the proposed method. Here, TS was calculated as follows:

$$TS = \frac{T_o - T_p}{T_o} \times 100\% \qquad (7)$$

where T_o and T_p are the total encoding times of the anchor reference and the proposed approach integrated into HM, respectively. The negative BD-Rate represents the better coding efficiency of the proposed method compared to its corresponding anchor and vice versa. The positive TS value represents the time savings achieved by the proposed method when compared with its corresponding anchor and vice versa.

To evaluate the contribution of the proposed approach, several state-of-the-art HEVC complexity reduction methods were also used to compare with the proposed method. The methods selected included ESD [6], the combination of ESD and CFM [7] (ESD+CFM [6,7]), and the method reported by Kim et al. [14]. Because this proposal used RD-cost as a condition for Early SKIP mode detection, it is called the RD-cost Method (RDM). Because the ESD and CFM have already been implemented as a fast tool mode in HM 16.20 reference software, RDM was also implemented in HM 16.20 to make fair comparisons.

4.2. Performance of the overall method

Table 2 lists the comparison results when the HM 16.20 reference software is used as the anchor for all complexity reduction algorithms. Table 3 shows the comparison results between the proposed method with relevant benchmarks, particularly the ESD [6], ESD+CFM [6,7], and RDM [14], respectively.

As shown in Table 2, the proposed method can achieve better encoding time savings than the relevant benchmarks. On average, the TS by the proposed method was approximately 47.22%, while the TS values for ESD [6], ESD+CFM [6,7], and RDM [14] were 33.93%, 35.72%, and 35.08%, respectively. Compared to the RD performance, the average BD-Rate value of the proposed method was 0.45%, while the ESD [6] method had an average BR-Rate of 0.22%. The BD-Rate of ESD+CFM [6,7] and RDM [14] were 0.56% and 0.40%, respectively. This result indicates that the proposed approach can provide good savings in encoding time without affecting the coding efficiency significantly. The contribution of this trading can be significant for real-time applications.

The results in Table 2 also suggest that the TS of the proposed method is better with the videos having slow changes and smoother regions than the other ones. In class A, while the Traffic video (with more smooth regions) could save approximately 58.97% encoding time, the other sequence *PeopleOnStreet* (with more complex regions) achieved only 32.03% time-saving. In class B, the sequence with slow motion (*BQTerrace*) showed a complexity reduction of 55.85%, whereas *BasketballDrive* with high activity had only a 45.64% saving of computation time. In particular, the screen content videos in class E could save approximately 70% encoding time.

Table 2. Performance comparison with the anchor of HM 16.20

Class	Sequences	ESD [6]		ESD+CFM [6,7]		RDM [14]		Proposed method	
		TS (%)	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*
A	Traffic	43.55	0.16	44.83	0.55	43.16	0.18	58.97	0.36
	PeopleOnStreet	22.96	0.32	25.53	0.80	24.47	0.53	32.03	0.49
B	BasketballDrive	32.50	0.23	32.71	0.51	34.08	0.22	45.64	0.44
	BQTerrace	41.12	0.25	42.56	0.40	40.39	0.14	55.85	0.36
	Cactus	34.88	0.22	35.96	0.58	33.52	0.24	50.80	0.45
	Kimono	35.03	0.20	36.29	0.42	35.59	0.14	47.74	0.28
	Parkscene	40.13	0.27	41.64	0.57	40.12	0.40	55.48	0.42
C	BasketballDrill	27.76	0.22	29.26	0.44	30.50	0.00	37.09	0.09
	BQMall	35.42	0.33	37.34	0.69	36.26	0.29	47.51	0.28
	PartyScene	27.96	0.13	29.66	0.47	30.49	0.51	39.86	0.44
	RaceHorses	19.64	0.19	22.59	0.64	24.22	0.70	28.65	0.62
D	BasketballPass	24.19	0.32	27.46	0.58	26.03	0.66	34.64	0.43
	BlowingBubbles	30.41	0.17	32.50	0.50	32.52	0.60	43.75	0.65
	BQSquare	38.15	0.34	40.58	0.63	38.20	1.10	53.42	1.07
	RaceHorses	19.45	0.23	23.30	0.77	28.77	0.86	29.42	0.81
E	FourPeople	51.52	0.01	52.34	0.43	48.78	0.02	70.11	0.19
	KristenAndSara	52.18	0.20	52.60	0.52	49.32	0.12	71.75	0.23
	Average	***33.93***	***0.22***	***35.72***	***0.56***	***35.08***	***0.40***	***47.22***	***0.45***

Table 3. Performance comparison of the proposed method with existing algorithm as anchor

Class	Sequences	Anchor ESD [6]		Anchor ESD+CFM [6,7]		Anchor RDM [14]	
		TS (%)	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*
A	Traffic	28.84	0.20	27.18	-0.18	29.33	0.18
	PeopleOnStreet	12.19	0.17	9.03	-0.31	10.46	-0.04
B	BasketballDrive	20.04	0.21	19.77	-0.07	18.14	0.22
	BQTerrace	28.53	0.11	26.73	-0.04	29.58	0.22
	Cactus	25.41	0.23	24.14	-0.13	26.96	0.21
	Kimono	23.46	0.07	21.94	-0.14	20.44	0.14
	Parkscene	31.35	0.15	29.58	0.15	27.10	0.01
C	BasketballDrill	18.97	-0.12	17.22	-0.34	8.29	0.09
	BQMall	28.64	-0.05	26.49	-0.41	18.49	-0.01
	PartyScene	25.05	0.31	23.21	-0.03	14.56	-0.07
	RaceHorses	16.45	0.42	13.21	-0.02	6.38	-0.08
D	BasketballPass	16.00	0.11	12.10	-0.15	11.57	-0.23
	BlowingBubbles	22.95	0.48	20.51	0.15	17.67	0.05
	BQSquare	30.44	0.73	27.50	0.43	27.44	-0.05
	RaceHorses	14.28	0.58	9.85	0.04	8.52	-0.05
E	FourPeople	26.53	0.17	25.26	-0.24	30.44	0.17
	KristenAndSara	29.97	0.03	29.35	-0.29	33.99	0.11
	Average	***23.48***	***0.22***	***21.36***	***-0.09***	***19.96***	***0.05***

This confirms that the distribution of SKIP mode in the smooth and simple activity regions is higher than that of complex or highly motion regions. Moreover, the proposed method provides good TS for all content videos. The worst case in *RaceHorses* sequence still achieved 29% saving time.

The results in Table 3 provide more information on the contribution of the proposed algorithm. Although the saving time of this algorithm was always better than the three existing state-of-art methods, the RD performance of the proposed method was no worse than the other methods. Compared with ESD [6], the proposed method achieved better coding efficiency in two sequences, *BasketballDrill* and *BQMall*. With the Kimono sequence, the BD-Rate value of only 0.07% was negligible. Compared to RDM [14], only four sequences in class B (except *ParkScene*) and sequence Traffic in class A showed poorer RD performance. Compared to ESD+CFM [6, 7], the coding efficiency was better in most test sequences (except for *ParkScene*, *BlowingBubbles*, and *BQSquare*). The direct comparisons in RD performance explain why the proposed method achieves more time saving while the coding efficiency loss was less than or similar to the others.

4.3. Performance of individual early SKIP mode condition

Table 4. Conditional probabilities that SKIP is the optimal mode when the condition 1 is satisfied

QP	*Prob* {RDO selects the SKIP mode | (1) happens}									Avg
	a	b	c	d	e	f	g	h	i	
22	0.9575	0.9772	0.9551	0.9944	0.9812	0.9882	0.9764	0.9801	0.9552	0.9739
27	0.9677	0.9889	0.9691	0.9983	0.9935	0.9959	0.9882	0.9922	0.9707	0.9850
32	0.9846	0.9933	0.9809	0.9993	0.9966	0.9985	0.9946	0.9964	0.9828	0.9919
37	0.9943	0.9960	0.9908	0.9996	0.9983	0.9993	0.9979	0.9984	0.9914	0.9962
Avg	0.9760	0.9889	0.9740	0.9979	0.9924	0.9955	0.9893	0.9918	0.9750	0.9867

Table 5. Conditional probabilities that SKIP is the optimal mode when the condition 2 is satisfied

QP	*Prob* {RDO selects the SKIP mode | (6) happens}									Avg
	a	b	c	d	e	f	g	h	i	
22	0.9819	0.9907	0.9814	0.9966	0.9874	0.9930	0.9884	0.9888	0.9822	0.9878
27	0.9851	0.9931	0.9876	0.9986	0.9946	0.9954	0.9935	0.9947	0.9860	0.9921
32	0.9954	0.9943	0.9923	0.9991	0.9957	0.9969	0.9960	0.9968	0.9882	0.9950
37	0.9989	0.9955	0.9953	0.9995	0.9976	0.9984	0.9980	0.9983	0.9888	0.9967
Avg	0.9903	0.9934	0.9891	0.9984	0.9938	0.9959	0.9940	0.9947	0.9863	0.9929

a: BQSquare, b: BasketballDrill, c: PartyScenel, d: FourPeople, e: Cactus, f: Kimono, g: ParkScene, h: Traffic, i: PeopleOnStreet

Table 6. Performance of individual proposed algorithm compared with HM 16.20

Class	Sequences	Condition 1 only		Condition 2 only		Condition 1 & Condition2	
		TS (%)	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*	*TS (%)*	*BD-Rate (%)*
A	Traffic	50.96	0.30	37.77	0.05	58.97	0.36
	PeopleOnStreet	26.22	0.35	19.88	0.16	32.03	0.49
B	BasketballDrive	34.20	0.13	26.15	0.15	45.64	0.44
	BQTerrace	47.60	0.22	36.50	0.13	55.85	0.36
	Cactus	38.17	0.15	25.97	0.10	50.80	0.45
	Kimono	37.81	0.10	31.70	0.09	47.74	0.28
	Parkscene	45.51	0.32	36.73	0.10	55.48	0.42
C	BasketballDrill	29.16	0.09	23.17	0.10	37.09	0.09
	BQMall	37.37	0.16	29.11	0.10	47.51	0.28
	PartyScene	31.56	0.31	21.20	0.05	39.86	0.44
	RaceHorses	23.61	0.29	12.66	0.23	28.65	0.62
D	BasketballPass	27.45	0.43	21.14	0.23	34.64	0.43
	BlowingBubbles	35.02	0.52	28.41	-0.02	43.75	0.65
	BQSquare	44.41	1.09	34.09	0.10	53.42	1.07
	RaceHorses	23.39	0.50	18.40	0.17	29.42	0.81
E	FourPeople	60.99	0.17	47.43	0.09	70.11	0.19
	KristenAndSara	63.19	0.16	44.17	0.17	71.75	0.23
	Average	***38.63***	***0.31***	***29.09***	***0.12***	***47.22***	***0.45***

Before assessing the compression efficiency and time-saving, the accuracy of each early SKIP mode condition was first considered. Tables 4 and 5 list the conditional probabilities that the skip mode is selected as the best mode by RDO when (1) and (6) are satisfied, respectively. Following the results shown in Table 4, the conditional probabilities were 0.9867 on average and always higher than 0.95 over all the test sequences and QPs. This result shows that the constant parameters chosen in model (5) and condition 1 were sufficiently reliable for the early SKIP decision. The results in Table 5 also confirm the reliability of condition 2 for an early decision. With an average probability of approximately 0.9929, which was always higher than 0.98 in the overall tests, condition 2 helped save encoding time. Condition 2 was also easy to integrate into HM.

To evaluate the usefulness of each condition in designing the proposed early skip mode decision method, each condition was first integrated separately in HM 16.20, and the individual performance was evaluated. Both conditions were then implemented to determine the performance of the combined conditions. Table 6 lists their performance.

Both conditions individually provide relatively good encoding time saving without sacrificing much of the compression performance (Table 6). On average, condition 1 achieved 38.63% time-saving, while condition 2 achieved 29.09%. In particular, the time savings in classes with a high spatial resolution (class A and class B) were approximately 40% and 30% under conditions 1 and 2, respectively. This is a good result for real-time applications. On the other hand, the BD-Rate values of conditions 1 and 2 individually were approximately 0.31% and 0.12%, respectively. Considering the high value of time-saving, this degree of coding performance reduction can be negligible in many practical applications.

Furthermore, the combination of both conditions for the overall proposed method does not have a conflict. The time savings of the overall algorithm are always higher than either condition and achieved 47.22% on average. As a consequence of the higher saving time, the BD-Rate was slightly higher, i.e., 0.45%. Nevertheless, this trading is still acceptable for many practical real-time applications.

5. Conclusions

This paper proposed a novel method to make a fast decision for choosing SKIP mode as the best mode. The proposed method included two new SKIP mode-checking conditions. While the first used a model constructed based on a TId parameter and the associated RD cost of the reference frame, the second used the information related directly to the TId parameter to make a skip decision. The experimental results showed that both SKIP mode conditions were good enough to solve the problem of computation time, and there was no conflict when the two conditions were combined for the proposed algorithm. The proposed algorithm saved 47% of the encoding time on average, with only a 0.45% increase in BD-Rate under a Random Access Main configuration. This result shows that the contribution of the proposed method is suitable for real-time applications.

Acknowledgments

This work has been supported by Vietnam National University, Hanoi (VNU), under Project No. QG.19.22

References

[1] G. Sullivan, J. Ohm, W. Han and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649 - 1668, Sept. 2012. Article (CrossRef Link)

[2] V. Sze, M. Budagavi and G. J. Sullivan, High Efficiency Video Coding (HEVC): Algorithms and Architectures, Springer, 2014. Article (CrossRef Link)

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Circuits and Systems for Video Tecnology*, vol. 13, no. 7, pp. 560-576, Jul. 2003. Article (CrossRef Link)

[4] I. K. Kim, J. Min, T. Lee, W.J. Han and J. Park, “Block Partitioning Structure in the HEVC Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1697-1706, Dec. 2012. Article (CrossRef Link)

[5] V. Sze, M. Budagavi and G. J. Sullivan, “R-D Optimization,” in *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, Springer, 2014, pp. 282-287. Article (CrossRef Link)

[6] J. Kim, J. Yang, K. Won and B. Jeon, “Early determination of mode decision for HEVC,” in *Picture Coding Symposium*, Krakow, Poland, May. 2012. Article (CrossRef Link)

[7] R. H. Gweon and Y. L. Lee, “Early Termination of CU Encoding to Reduce HEVC Complexity,” *ITU-T/ISO/IEC Joint Coll. Team Video*, Jul. 2011. Article (CrossRef Link)

[8] M. Kim, N. Ling, L. Song and Z. Gu, “Fast SKIP mode decision with rate-distortion optimization for high efficiency video coding,” in *Multimedia and Expo Workshops*, 2014. Article (CrossRef Link)

[9] H. Lee, H. J. Shim, Y. Park and B. Jeon, “Early Skip mode decision for HEVC encoder with emphasis on coding quality,” *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 388-397, Sept. 2015. Article (CrossRef Link)

[10] Q. Hu, Z. Shi, X. Zhang and Z. Gao, “Early SKIP mode decision based on Bayesian model for HEVC,” in *Visual Communications and Image Processing*, 2015. Article (CrossRef Link)

[11] Y. Li, G. Yang, Y. Zhu, X. Ding and X. Sun, “Unimodal Stopping Model-Based Early SKIP Mode Decision for High-Efficiency Video Coding,” *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1431 - 1441, Jul. 2017. Article (CrossRef Link)

[12] L. Shen, Z. Zhang and Z. Liu, “Adaptive inter-mode decision for HEVC jointly utilizing inter-level and spatiotemporal correlations,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 10, pp. 1709-1722, 2014. Article (CrossRef Link)

[13] Z. Pan, S. Kwong, M. T. Sun and J. Lei, “Early MERGE Mode Decision Based on Motion Estimation and Hierarchical Depth Correlation for HEVC,” *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 405-412, Jun. 2014. Article (CrossRef Link)

[14] J. Kim and E. Izquierdo, “An Effective Strategy for Early Skip Mode Decision in HEVC,” in *IEEE International Conference on Image Processing*, Athens, Greece, Oct. 2018. Article (CrossRef Link)

[15] K. Choi, S. H. Park and E. S. Jang, “Coding Tree Pruning Based CU Early Termination,” ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) Document JCTVC-F092, Jul. 2011. Article (CrossRef Link)

[16] X. Shen and L. Yu, “CU splitting early termination based on weighted SVM,” *EURASIP Journal on Image and Video Processing*, 2013. Article (CrossRef Link)

[17] X. Gao, Z. Li and Y. Zhang, “A Fast HEVC Inter CU Size Decision Algorithm Based on Multi-Class Learning,” in *International Conference on Intelligent*

Human-Machine Systems and Cybernetics, Hangzhou, China, Aug. 2018. Article (CrossRef Link)

[18] Y. Li, Z. Liu, X. Ji and D. Wang, "CNN Based CU Partition Mode Decision Algorithm for HEVC Inter Coding," in *IEEE International Conference on Image Processing*, Athens, Greece, Oct. 2018. Article (CrossRef Link)

[19] X. Wang and Y. Xue, "Fast HEVC inter prediction algorithm based on spatio-temporal block information," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Cagliari, Italy, Jun. 2017. Article (CrossRef Link)

[20] L. Shen, Z. Liu, X. Zhang, W. Zhao and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 465-470, 2013. Article (CrossRef Link)

[21] J. Xiong, H. Li, F. Meng, Q. Wu and K. Ngan, "Fast HEVC inter CU decision based on latent SAD estimation," *IEEE Transactions on Multimedia*, vol. 17, no. 12, p. 2147–2159, 2015. Article (CrossRef Link)

[22] Q. Hu, X. Zhang, Z. Shi and Z. Gao, "Neyman-Pearson-Based Early Mode Decision for HEVC Encoding," *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 379-391, Mar. 2016. Article (CrossRef Link)

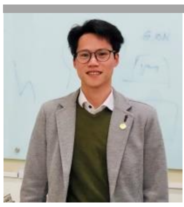
[23] J. Neyman and E. S. Pearson, "IX. On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London A*, vol. 231, no. 694-706, p. 289–337, Feb. 1933. Article (CrossRef Link)

[24] H. Schwarz, D. Marpe and T. Wiegand, "Hierarchical B pictures, Joint Video Team (JVT)," in *Document JVT-P014*, Poznan, Jul. 2005. Article (CrossRef Link)

[25] F. Bossen, "Common test conditions and software reference configuration," Document JCTVC-L1100, Geneva, 2013. Article (CrossRef Link)

[26] "HEVC reference software," [Online] Available: Article (CrossRef Link)

[27] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Document VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2001. Article (CrossRef Link)



Xiem HoangVan is the Deputy Head of the Department of Robotics Engineering, Faculty of Electronics and Telecommunications, Vietnam National University – University of Engineering and Technology (VNU-UET). He received his Ph.D. degree from *Lisbon University*, Portugal, in 2015, M.Sc. degree from *Sungkyunkwan University*, South Korea, in 2011, and B.E degree from *Hanoi University of Science and Technology*, Vietnam, in 2009, all in Electrical and Computer Engineering. His research interests are machine learning, image, and video communications.

Dr. Xiem has published about 50 papers on image and video processing and regularly reviews for many renowned IEEE, IET, and EURASIP journals and serves as a technical committee member for international conferences and funding agencies worldwide. He has received several technical awards for his contributions on image and video coding, including a *Best Paper Award* from the Picture Coding Symposium 2015 (Australia), a *Best Paper Award* from the International Workshop on Advanced Image Technology 2018 (Thailand), and a *Ph.D. award* from the Fraunhofer Portugal Challenge 2015, and an *Outstanding reviewer* award of the Elsevier Journal of Signal Processing: Image Communication. Dr. Xiem is a recipient of the prestigious *Golden Globe Award for Young Scientists (under 35 years old) in Science and Technology 2019* and the VNU Top young scientist award 2019.

Minh Dinh Bao received his BSc in 2020 from Vietnam National University (VNU). He is currently working in the Department of Robotics Engineering – University of Engineering and Technology (VNU-UET). His research interests are Image/Video Processing & Communication, and Multimedia Communication over Wired/Wireless Networks.

Byeungwoo Jeon received his BSc in 1985 and an MSc in 1987 from the Department of Electronics Engineering, Seoul National University, Seoul, Korea. He received his Ph.D. in 1992 from the School of Electrical Engineering at Purdue University, Indiana, in the United States. From 1993 to 1997, he was in the Signal Processing Laboratory at Samsung Electronics in Korea, where he worked on video compression algorithms, designing digital broadcasting satellite receivers, and other MPEG-related research for multimedia applications. Since September 1997, he has been with the Faculty of the School of Information and Communication Engineering, Sungkyunkwan University, Korea, where he is currently a professor. He currently serves as associate editors for *IEEE Transactions on Circuits and Systems for Video Technology* and *IEEE Transactions on Broadcasting*. His research interests include multimedia signal processing, video compression, statistical pattern recognition, and remote sensing.