Unimodal Stopping Model-Based Early SKIP Mode Decision for High-Efficiency Video Coding

Yue Li, Gaobo Yang, Yapei Zhu, Xiangling Ding, and Xingming Sun, *Senior Member, IEEE*

***Abstract*—High-efficiency video coding (HEVC) can greatly improve coding efficiency compared with the prior video coding standard H.264/AVC by adopting advanced hierarchical coding structures such as coding unit (CU), prediction unit (PU), and transform unit. For each CU, an exhaustive mode decision strategy is adopted to achieve the best rate distortion (RD) cost, which simultaneously results in enormous computational complexity. In this paper, an early SKIP mode decision algorithm is proposed for the HEVC encoder to speed up the process of mode decision. Each CU size is categorized into either rare used or frequent used by exploiting the correlation of CU depth, which is estimated from the temporally colocated CUs. For the rare-used CU size, the SKIP mode is directly selected as the optimal mode and the remaining mode decision process is early terminated. For the frequent-used CU size, a unimodal stopping model is designed for its early SKIP mode decision by exploiting both hierarchical mode structure and RD cost property. Experimental results show that the proposed early SKIP mode decision method achieves average 58.5% and 54.8% encoding time savings, while the Bjontegaard Delta bit rate only increases average 0.8% and 0.8% for various test sequences under the random access and the low delay B conditions, respectively.**

***Index Terms*—High efficiency video coding (HEVC), mode decision, SKIP mode, unimodal stopping model (USM).**

I. Introduction

THE increasing demand of high-resolution video services requires more transmission bandwidth due to the huge amount of video data. To address this demand, High Efficiency Video Coding (HEVC) has been standardized by Joint Collaborative Team on Video Coding (JCT-VC) as the most recent video coding standard [1]. To improve the video coding efficiency, HEVC adopts many advanced coding tools such as large block size, asymmetric motion partitioning (AMP), etc. Among them, hierarchical coding structure is very flexible in which each coding unit (CU) block have prediction unit (PU) blocks of symmetric and asymmetric sizes and transform unit (TU) blocks of quad-tree partitions. This flexible coding structure contributes a significant coding gain over the prior video coding standard H.264/AVC [2]. However, the coding tree unit (CTU) with 64 $\times$ 64 pixels should be recursively split into four CUs until the smallest CU (SCU) of 8 $\times$ 8 pixels, in which an exhaustive mode decision has to be computed for each CU in rate-distortion optimization (RDO) sense. Thus, the flexibility of hierarchical block partitioning imposes significant computational burden on HEVC encoder to seek the optimal combinations of CU, PU and TU sizes by RDO. This is actually a bottleneck of HEVC to hinder its real-time video applications. It is highly desirable to reduce the computational complexity of HEVC encoder while simultaneously maintaining the coding efficiency without great degradation of perceptual video quality [3].

During past decades, there are lots of fast mode decision (FMD) algorithms presented for H.264/AVC and its extensions [4]–[10]. They exploit many advanced techniques including RD cost correlation [4], [5], [7], [9], spatio-temporal correlations [6], optimal stopping model [8], macroblock position constraint model [10] and so on. In addition, several early SKIP/DIRECT mode decision algorithms have been proposed for multi-view video coding (MVC) by exploiting RD correlation, inter-view correlation and mode homogeneity [11]–[15]. However, these techniques can not be directly extended to HEVC due to its more flexible hierarchical structure of coding units.

In recent years, a few FMD algorithms were proposed to speed up the HEVC encoder [16]–[29]. They can be roughly divided into three types. The first type exploits some encoding parameters such as coded block flag (CBF), sample adaptive offset (SAO) and residual distortion [16]–[18]. For instance, a FMD approach was proposed by using the available CBFs of all luminance and chrominance components [16]. If the CBF value of current PU mode is equal to zero, the reminding mode decision process is early terminated. The second type speeds up the mode decision by exploiting RD cost or spatial-temporal and inter-level mode correlation [19]–[21]. The third type adopts hybrid model analysis strategy for FMD [22]–[29]. For example, an adaptive mode ordering strategy was proposed by exploiting both RD cost and bit cost statistics [22].

The SKIP mode is an efficient coding tool for HEVC [32]. It represents a coded block data without motion searching and

Manuscript received September 12, 2016; revised December 14, 2016 and February 3, 2017; accepted February 10, 2017. Date of publication February 15, 2017; date of current version June 15, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61572183, Grant 61379143, and Grant 61672222, in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20120161110014, in part by the Priority Academic Program Development of Jiangsu Higer Education Institutions, and in part by Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ivan V. Bajic.

Y. Li, G. Yang, and X. Ding are with the School of Information Science and Engineering, Hunan University, Changsha 410082, China (e-mail: yueli@hnu.edu.cn; yanggaobo@hnu.edu.cn; xianglingding@163.com).

Y. Zhu is with the Faculty of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421002, China (e-mail: zyp1016@hynu.edu.cn).

X. Sun is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nangjing 210044, China (e-mail: sunnudt@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2669863

residual coding since its transform coefficients are all quantized to zero. That is, the SKIP mode does not need time-consuming motion estimation (ME) [30], [31], [35]. Meanwhile, it is the dominant mode at high QPs (low bitrates), especially for static or texture-smooth regions. The most representative FMD schemes with early SKIP mode decision are summarized as follows. In [32], an early SKIP mode decision method was proposed by using the CBF and Motion Vector Difference (MVD) of Inter 2N $\times$ 2N mode. In [33], the RD correlation between the SKIP mode and the 2N $\times$ 2N Merge mode is exploited to select an adaptive threshold for early SKIP mode decision. However, its saving of encoding time is limited since it puts emphasis on coding quality. In [34], an early Merge mode decision scheme was proposed by exploiting the hierarchical depth correlation and motion estimation of Inter 2N $\times$ 2N mode. However, it only exploits the hierarchical depth correlation between depth level "0" and other levels. If the optimal mode of depth level "0" is not the Merge mode, it can not exploit the hierarchical depth correlation for other depth levels. In addition, for the larger CU sizes, which are rare-used in motion-complex regions, the early SKIP mode decision is not fully exploited as well [32]–[34].

We believe that there is still some space left for early SKIP mode decision to further reduce the computational complexity. Firstly, most existing early SKIP mode decision approaches only use the RD cost or the mode correlation between the SKIP and Inter 2N $\times$ 2N mode. That is, when the SKIP mode is the optimal mode, they do not efficiently early terminate the checking of other modes. Secondly, larger CU sizes for complex motion regions and hierarchical mode correlation are also not fully exploited. This motivates us to propose an early SKIP mode decision method for HEVC by exploiting CU depth level correlation, hierarchical mode structure and RD cost property. Specifically, the contributions are three-folds: 1) each CU size is firstly categorized into either rare-used or frequent-used by exploiting the CU depth level correlation, which is estimated from two temporally co-located CUs. 2) for rare-used CU size, the SKIP mode is directly selected as the optimal mode to maximize the reduction of computational cost. 3) for frequent-used CU size, a novel unimodal stopping model (USM) is established by considering both mode hierarchy and RD cost to early predict whether the SKIP mode is the optimal mode or not.

The rest of this paper is organized as follows. Motivations and some statistical analyses are given in Section II. Section III presents the proposed early SKIP mode decision algorithm. Experimental results are reported in Section IV, and conclusion is made in Section V.

II. Motivation and Statistical Analyses

A. *Distributions and Correlation of CU Depth Levels*

Fig. 1 is the quad-tree coding structure of CU. By recursively partitioning of CTU into SCUs, the hierarchical depth varies from 0 to 3 to adapt to diverse video contents. That is, videos with diverse motion activities have different distribution of CU sizes (or depth levels). For the CUs in static or texture-smooth regions, lower CU depth levels are always selected. For the CUs in motion-complex regions, they usually select larger CU depth levels. To characterize the distribution of CU depth levels, three typical video sequences with different motion activities and spatial resolutions are tested under the conditions summarized in Table I. BasketballPass (416 $\times$ 240) sequence has fast motion, BQMall (832 $\times$ 480) sequence has a moderate motion and Johnny (1280 $\times$ 720) sequence has relatively slow motion. Fig. 2 compares their distributions of CU depth levels. From it, we observe that different videos have quite different distributions of CU depth levels. The possibility of selecting depth 2 and depth 3 is about 10% for slow motion video, whereas it is about 34% for videos with medium or fast motions. This implies that depth 0 and depth 1 are not the optimal depth levels for those CUs in motion-complex regions. If we can early decide whether a CU depth level is suitable or not for different regions, and the exhaustive mode decision processing can be avoided by early skipping the rest levels to reduce the computational complexity.

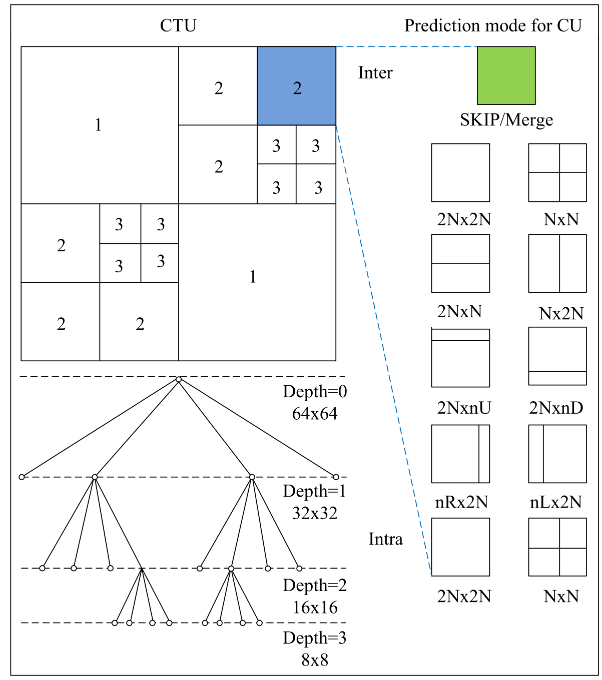


Fig. 1. CU and PU structures of HEVC encoders.

TABLE I
Test Conditions

Search range	64
CTU size	64
Depth level range	From "0" to "3"
Configuration	Random Access (RA)
Number of encoded frames	61
Basis quantization parameter (QP)	20, 25, 30, 35

In addition, video sequences have strong temporal correlations among neighboring frames. Fig. 3 shows the example of the reference structure of random access in which the group of picture (GOP) size is 8. To analyze the temporal correlation, the probability of depth level difference $P_{DD=i}$ is defined as follows:

$$P_{DD=i} = \frac{N_{DD=i}}{\sum_{i=0}^{3} N_{DD=i}} \qquad (1)$$

$$DD = |d_{cur} - d_{pre}|, DD \in \{0,1,2,3\} \qquad (2)$$

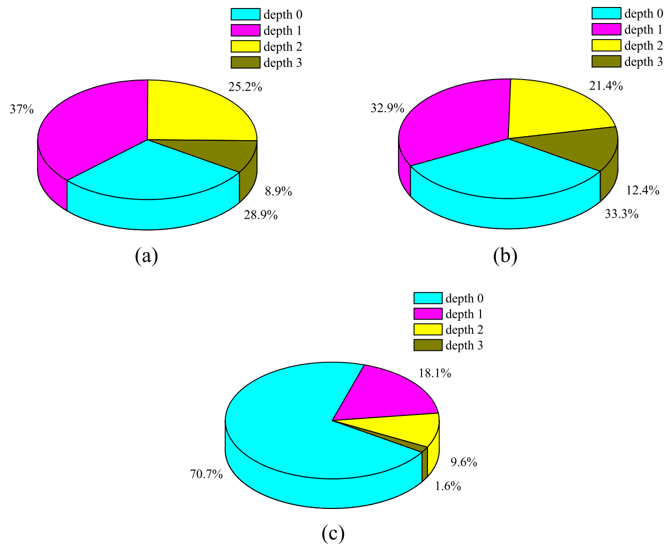


Fig. 2. CU depth levels distribution of three video sequences: (a) Basketballpass, (b) BQMall, and (c) Johnny.

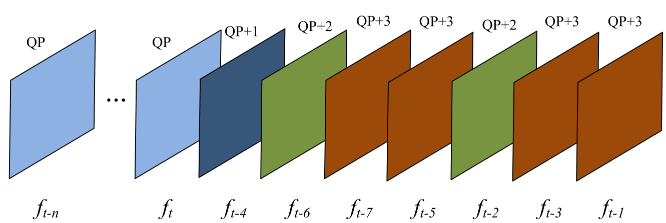


Fig. 3. Reference structure of random access.

where d_{cur} and d_{pre} are the depth levels of the current CU and its co-located CU in previously encoded frame, respectively. Suppose current frame and the temporally co-located frames have the same QP values. If f_{t-2} is the current frame whose QP value is QP+2, the QP value of the frame f_{t-6}, which is defined as the nearest neighboring encoded frame, is also equal to QP+2. DD is the depth level difference between the current CU and its nearest co-located CU. $N_{DD=i}$ represents the number of CUs whose DD is equal to i under the test conditions summarized in Table I. The statistical results of $P_{DD=i}$ are summarized in Table II. From it, $P_{DD=0}$ is about 76.2% on average, while the sum of $P_{DD=0}$ and $P_{DD=1}$ is up to 96.2%. Therefore, we can conclude from Table II that since the depth level correlation between current CU and its temporally co-located CU is extremely high, the CU depth levels of the co-located CUs in encoded frame should be exploited to predict the depth levels of the CUs in current frame.

B. Distributions and RD Cost of PU Mode

HEVC encoder supports eleven prediction modes including SKIP/Merge, Inter_2N × 2N, Inter_N × 2N, Inter_2N × N, Inter_N × N (available for the minimum CU), Inter_2N × nU, Inter_2N × nD, Intern_L × 2N, Intern_R × 2N, Intra_2N × 2N and Intra_N × N (available for the minimum CU). Among them, the SKIP mode is always performed before checking other modes in the process of mode decision. This mechanism is similar with the SKIP mode decision of H.264/AVC, which does not need computation-intensive ME to obtain motion vector and residual signal. That is, the SKIP mode of HEVC can obtain desirable coding gain with low computational complexity. It is reasonable to early determine whether the SKIP mode is the optimal mode or not for a CU.

TABLE II
PROBABILITY OF DEPTH LEVEL DIFFERENCE (DD) BETWEEN TWO ADJACENT FRAMES

Sequences	QP	$DD=0$	$DD=1$	$DD=2$	$DD=3$	$DD\leq1$
Basketballpass (416 × 240)	20	74.8	21.8	2.9	0.5	96.6
	25	78.8	17.5	3.2	0.5	96.3
	30	79.6	16.8	3.2	0.4	96.4
	35	81.6	15.7	2.4	0.3	97.3
	ave.	78.7	18.0	2.9	0.4	96.7
BQMall (832 × 480)	20	56.4	37.0	6.1	0.4	93.4
	25	63.9	29.6	5.8	0.6	93.5
	30	68.4	25.4	5.5	0.6	93.8
	35	71.9	22.7	4.9	0.5	94.6
	ave.	65.2	28.7	5.6	0.5	93.9
Johnny (1280 × 720)	20	69.1	26.5	4.1	0.3	95.6
	25	84.5	13.9	1.5	0.1	98.4
	30	90.6	8.4	0.9	0.1	99.0
	35	94.4	4.8	0.6	0.1	99.2
	ave.	84.6	13.4	1.8	0.2	98.0
ave.		76.2	20.0	3.4	0.4	**96.2**

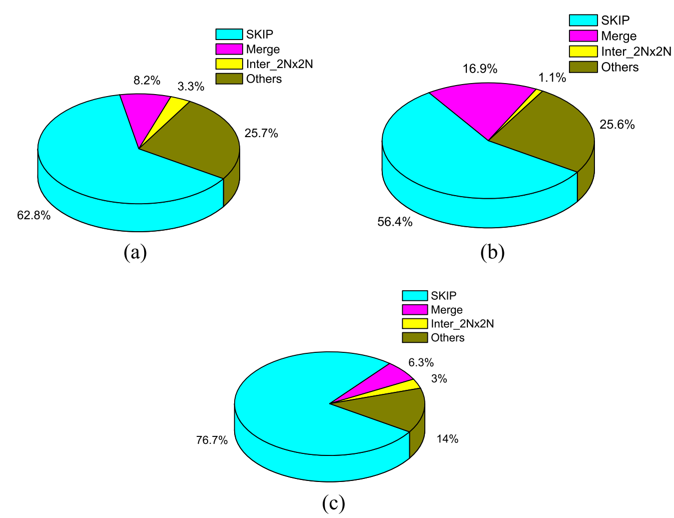


Fig. 4. Mode distribution of three video sequences: (a) Basketballpass, (b) BQMall, and (c) Johnny.

We further analyze the probability of SKIP mode when it is selected as the optimal mode. Three video sequences "Basketballpass", "BQMall" and "Johnny" are tested with the same conditions summarized in Table I. Fig. 4 reports their mode distributions. From it, we observe that most CUs select the SKIP mode as the optimal mode. For video sequences with fast or medium motion such as "Basketballpass" and "BQMall", the percentages are about 63% and 56%, respectively. For video sequence with low motion such as "Johnny", up to 76.7% CUs are encoded with the SKIP mode. The inherent reason is that static and texture-homogeneous regions prevail in natural videos, which are most likely to choose the SKIP mode as the optimal mode.

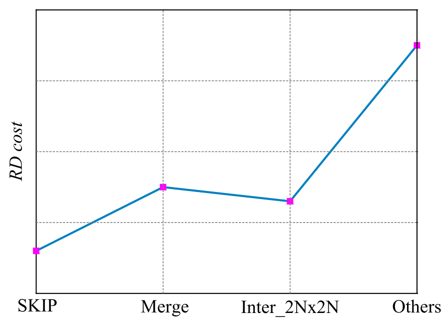


Fig. 5. Relationship between RD cost (J_m) and mode under SKIP mode is the optimal mode.

Thus, if we can design a stopping model to for early SKIP mode decision, it will early terminate the mode decision process to save coding time.

HEVC adopts an exhaustive mode decision strategy for each CU via RDO, and the RD cost function (J_m) is computed as

$$J_m = SSE + \lambda \cdot B_{mode} \tag{3}$$

where SSE is the sum of squared error (SSE) between the current CU and its reconstructed CU, B_{mode} is the bit cost and λ is the Lagrangian multiplier. To achieve the least RD cost for each CU, the optimal mode decision function can be defined as

$$\hat{m} = \underset{m \in \mathrm{M}}{\text{argmin}} \{J_m\} \tag{4}$$

where J_m is the RD cost of the mode m, which is computed by (3). M represents all candidate modes, and $\hat{m}$ is the optimal mode to be selected for the current CU. From (4), we know that if mode m is selected as the optimal mode, J_m has the minimum value of RD cost. If the SKIP mode is the optimal mode, there exists one situation as shown in Fig. 5, which depicts the relationship between the RD cost J_m and the mode m. From it, we observe that compared with the SKIP mode, there is a significant increase of RD costs for the Merge mode, the Inter_2N $\times$ 2N and others modes. Therefore, we can firstly check the SKIP mode for a CU, and then check the Merge mode and the Inter_2N $\times$ 2N mode. If the RD costs of the Merge mode and the Inter_2N $\times$ 2N mode increase as the tendency shown in Fig. 5, the others modes can be skipped to reduce the encoding time.

III. Proposed Early SKIP Mode Decision Algorithm

A. *Early SKIP Mode Decision for Rare-Used CU Sizes*

Motivated by the analysis in Section II-A, each CU size is firstly categorized into rare-used or frequent-used based on the depth level predicted for the current CU. For the rare-used CU size, the SKIP mode is directly selected as the optimal mode to avoid unnecessary variable-size ME process. From Table II, there are only 76.2% CUs on average whose depth level is equal to the temporally co-located and encoded CUs. If the depth level of co-located CU is directly used to predict the depth level of the current CU, the prediction error is not negligible. Moreover, the kind of prediction error might also be accumulated. To reduce the prediction error, two co-located CUs of nearest neighboring frames are exploited to predict the depth level of the current CU, as shown in Fig. 6. Let D_0^s and D_1^s be the maximum depth levels of two previously neighboring CUs in the intervals of t_{pre0} and t_{pre1}, respectively. The predicted depth level D_p^s of current CU size is calculated as

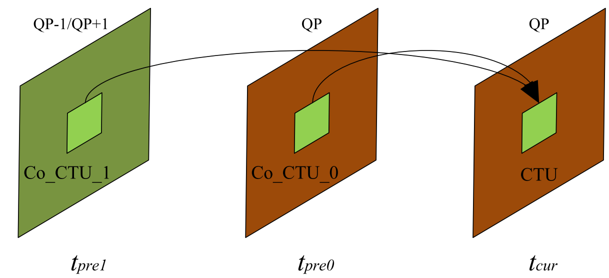


Fig. 6. Correlation between current CU and temporally co-located CUs.

$$D_p^s = w \cdot D_0^s + (1 - w) \cdot D_1^s \tag{5}$$

$$s = \begin{cases} 0, & \text{CU size with } 64 \times 64 \\ 1, & \text{CU size with } 32 \times 32 \end{cases} \tag{6}$$

where s denotes the CU size and w is the weighting factor. In general, the closer is the neighboring CU to current CU, the larger w should be assigned. That is, w has a strong temporal correlation, which is defined as follows:

$$w = \frac{|t_{cur} - t_{pre1}|}{|t_{cur} - t_{pre0}| + |t_{cur} - t_{pre1}|} \tag{7}$$

where t_{cur}, t_{pre0} and t_{pre1} are the current interval and two neighboring encoded intervals, respectively. And the nearest neighboring frames in the intervals t_{pre0} and t_{pre1} are encoded with QP and QP-1, respectively. Especially, if it does not exist the encoded frame with QP-1, the encoded frame with QP+1 is chosen as the neighboring frame. The basic unit of intervals is assumed to be 1 here. In terms of D_p^s, current CU is classified as either frequent-used or rare-used as follows:

$$\text{CU} \in \begin{cases} \text{rare-used}, & D_p^s > s + TH \\ \text{frequent-used}, & \text{otherwise} \end{cases} \tag{8}$$

where s is size of the current CU and TH is a threshold to be adjusted. In this paper, we only consider the CU sizes of 64 $\times$ 64 and 32 $\times$ 32 for the current CU to decide whether they are frequent-used or rare-used, and the other CU sizes are defined as frequent-used. Since the probability that the absolute depth level difference is less than or equal to 1 is more than 96%, the threshold TH is set as 1. Let D_p^0 and D_p^1 be the prediction depth of current CU with size 64 $\times$ 64 and 32 $\times$ 32, respectively. When s is equal to 0 and D_p^0 is bigger than 1, the CU size with 64 $\times$ 64 is defined as rare-used. Similarly, if s is equal to 1 and D_p^1 is bigger than 2, the CU size with 32 $\times$ 32 is also defined as rare-used. For a CU with complex motion, the larger CU sizes is seldom selected as the optimal CU sizes. If the SKIP mode is early determined for those CUs in the regions with complex motion, it can save the coding time by skipping variable-sized ME.

TABLE III
MODE TYPES AND THEIR CORRESPONDING MODES

Types N	1	2	3	4
mode	SKIP	Merge	Inter_2N × 2N	Others

B. Unimodal Stopping Model (USM) Based Early SKIP Mode Decision

1) Optimal Stopping Model for Mode Decision: In [38], an optimal stopping problem is formulated by a sequence of random variables and a sequence of real-valued reward functions. A decision-maker sequentially checks these variables to obtain the observed value and finds a time to stop by maximizing the expected reward. Moreover, an optimal stopping model is introduced for the early mode decision of H.264/SVC [8]. Firstly, the probabilities of all N candidate modes are ranked in a descending order

$$p_i \geq p_j \ \forall i, j \in [1, N], i < j. \tag{9}$$

Then, the optimal stop mode is defined as

$$K_* = \max\{K_\alpha, K_\beta\} \tag{10}$$

where

$$K_\alpha = \min\left\{k \geq 1 : \sum_{i=1}^{k} p_i \sum_{j=k}^{N} \frac{1}{\sum_{r=1}^{j} p_r} > \tau - k\right\} \tag{11}$$

$$K_\beta = \min\left\{k \geq 1 : p_{k+1} \sum_{j=k+1}^{N} \frac{1}{\sum_{r=1}^{j} p_r} \leq 1\right\} \tag{12}$$

where $\tau \epsilon [N, N+1]$ is a constraint threshold. A larger τ means a better decision performance. However, the optimal stopping model of H.264/SVC in [8] are not suitable for the early mode decision of HEVC because HEVC supports quite different coding units. In this paper, an optimal stopping model is proposed for the early SKIP mode decision by considering the characteristic of coding units and different coding modes for HEVC.

2) Unimodal Property: Table III summarizes the mode types and the corresponding modes. From (3) and (4), since J_m only has one minimum value, J_m can be regarded as a unimodal function of m. Thus, J_m is written as

$$\exists m, m^* \in [1, N], J_m - J_{m^*} > 0 \tag{13}$$

where N represents the type of mode, m^* is the optimal mode type and m is the mode type except m^*.

According to (13), if m^* is selected as the optimal mode, the RD costs of other modes must be larger than that of the mode m^*. That is, the RD cost of the mode m^* (J_{m^*}) is the unimodal point. However, the unimodal point is found by an exhaustive mode decision process. To reduce the computational complexity, we exploit the hierarchical mode structure and the RD cost property to establish a unimodal stopping model (USM), so as to early predict whether the RD cost of the mode (J_{m^*}) is the minimum value or not.

a) RD cost-based USM: If the SKIP mode is chosen as the optimal mode, the RD cost can be expressed by

$$\forall m \in [2, N], J_m - J_1 > 0 \tag{14}$$

where J_m is the RD cost of mode m. Fig. 7(a) and 7(b) show two cases of the relationship between the RD cost and the mode type when the SKIP mode is selected as the optimal mode. Please note that there is a strong linear relationship between J_m and m, which can be expressed by

$$J_m = a \cdot m + c \tag{15}$$

where the symbols a and c are the variable parameters. If the SKIP mode is the optimal mode, a is a positive value and J_m is a monotonic increasing trend function of m. In other words, if (15) is confirmed to be a monotonic increasing trend function, the SKIP mode will be selected as the optimal mode. Thus, the USM is established by combining (14) and (15) for early SKIP mode decision. To achieve a good trade-off between time reduction and coding performance, the RD costs of three modes including SKIP, Merge and Inter2N × 2N are firstly checked and used to verify whether the USM is met or not. Then, the condition of early SKIP mode decision is rewritten as

$$\forall m \in [2, 3], J_m - J_1 > 0. \tag{16}$$

b) Hierarchical mode and RD cost property-based USM: It is claimed that there is a hierarchical correlation for the Merge mode [34], [39]. If the Merge mode is selected as the optimal mode for the parent CU, its children CUs are also more likely to be encoded with the Merge mode. Similarly, if the parent CU is encoded with the SKIP mode, its children CUs have a high probability to be encoded with the SKIP mode. Thus, the condition of early SKIP mode decision in (16) should further consider the hierarchical correlation of mode among parent CU and its children CUs. We only consider the mode information of the parent CU whose optimal mode is the SKIP mode. Then, we define the SKIP mode of the parent CU as P_SKIP to distinguish from the SKIP mode of the children CU. As shown in Fig. 7(c), if the optimal mode of the parent CU is SKIP mode, and the RD cost of the SKIP mode of the children CU is less than that of the Merge mode, the mode decision for the children CU can be early terminated. Thus, (16) is rewritten as

$$\text{P\_SKIP} = \text{SKIP mode} \ \&\& J_2 - J_1 > 0. \tag{17}$$

C. The Proposed Overall Algorithm

In Subsections *A* and *B*, two early SKIP mode decision strategies are presented for rare-used and frequent-used CUs, respectively. Fig. 8 is the overall flowchart of the proposed early SKIP mode decision algorithm for HEVC. The main steps are summarized as follows.

1) Start mode decision for CUs in the inter-frames.
2) Compute D_p^s using (5). If the current CU size is rare-used, go to step 3, otherwise go to step 4.
3) Check the SKIP mode only, and the other modes are skipped.
4) If the parent CU of current CU is encoded with the SKIP mode, go to step 5, otherwise go to step 6.

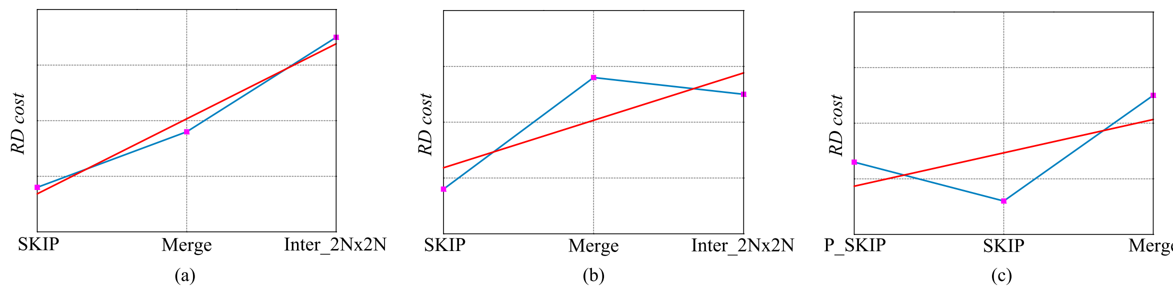


Fig. 7. Relationship between the RD cost and the type of prediction modes when the SKIP mode is optimal prediction mode.

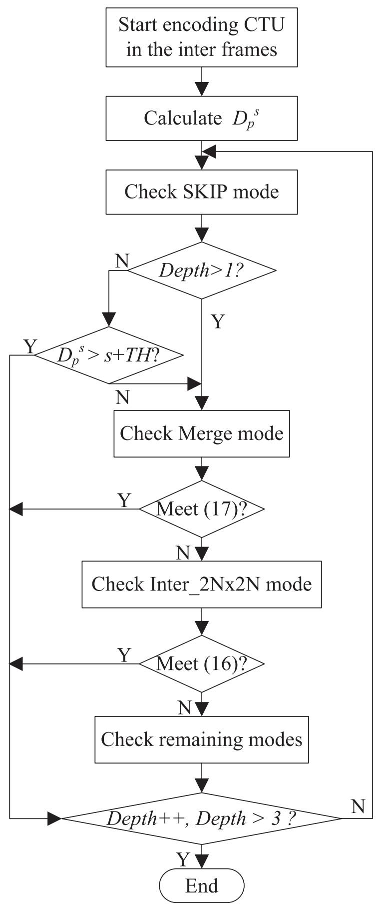


Fig. 8. Flowchart of the proposed early SKIP mode decision method.

5) Judge between the SKIP mode and the Merge mode. If the current CU meets (17), the SKIP mode is early determined and the other modes can be skipped. Otherwise, check the rest modes.
6) Make decision among the SKIP, Merge and Inter2N $\times$ 2N modes. If the RD cost of the current CU meets (16), the SKIP mode can be early determined and the rest modes are skipped. Otherwise, check the rest modes.
7) Determine the optimal mode for the current CU.

IV. Experimental Results and Analyses

A. Test Conditions

To verify the validity of the proposed early SKIP mode decision approach, it is integrated into the HEVC reference software HM12.0 for experiments. The common test conditions of HEVC are adopted under the random access (RA) and low-delay B (LD B) configurations, respectively [36]. The search range is 64 for both horizontal and vertical directions. The size of CTU is 64 $\times$ 64, and correspondingly its depth is 4. QPs are 22, 27, 32, and 37. Fast encoder decision, fast decision for Merge, and transform skip are enabled. The performance is evaluated in terms of Bjonteggrad Delta BitRate (BDBR) [37] and Time Saving (TS). TS is defined as follows:

$$TS = \frac{T_{HM} - T_{Proposed}}{T_{HM}} \times 100\% \qquad (18)$$

where T_{HM} and $T_{Proposed}$ are the total encoding times of the original reference software HM and the proposed approach integrated into HM, respectively.

B. Results of Individual Algorithms Compared With Original HM

Table IV reports the experimental results of the proposed early SKIP mode decision method. Since two strategies are adopted for rare-used CUs and frequent-used CUs, which are denoted by Proposed1 and Proposed2 (actually USM), respectively. Therefore, their contributions to early SKIP mode decision are also separately provided for observation. From Table IV, the Proposed1 method achieves 11.1% total encoding time saving with 0.3% BDBR increase on average. Especially for video sequences with fast motions, it greatly reduces the total encoding time with only negligible BDBR loss compared with videos with slow motions. The reasons are as follows. For video sequences with fast motions, they usually have more complex motion regions in which large CUs are more likely to be the rare-used CUs. Thus, the SKIP mode can be early terminated for them. On contrary, for slow motion videos with more static background regions, there are less rare-used CUs of large CU sizes, which limits the encoding time saving. For example, the encoding time is reduced only about 1% for Johnny sequence with low motion. From Table IV, we can also observe that the USM based early SKIP mode decision method achieves about 47% encoding time reduction with only 0.6% BDBR increase on average. Moreover, it achieves much more encoding time reduction for slow motion videos, simply because there are more static regions which are more likely to choose the SKIP mode as the optimal mode. The proposed overall approach totally achieves 58.5% encoding time saving with 0.8% BDBR increase on average, respectively. That is, it effectively reduces the encoding time for video sequences with diverse motions.

TABLE IV
PERFORMANCE OF THE PROPOSED EARLY SKIP MODE DECISION UNDER THE RA CONDITION

		Proposed1		Proposed2		Proposed	
Resolution	Sequences	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A	PeopleOnstreet	0.3	22.7	0.6	27.1	0.8	50.0
(2560 × 1600)	Traffic	0.4	8.8	0.4	59.0	0.7	66.7
Class B	Kimono	0.3	4.5	0.4	47.9	0.7	52.7
(1920 × 1080)	ParkScene	0.3	9.5	0.5	55.1	0.8	63.5
	Cactus	0.2	8.8	0.4	48.9	0.6	57.6
	BasketballDrive	0.5	8.1	0.6	44.5	1.1	52.8
	BQTerrace	0.2	8.3	1.0	57.0	1.1	64.2
Class C	BasketballDrill	0.5	13.2	0.4	37.0	0.8	50.1
(832 × 480)	BQMall	0.4	12.5	0.7	47.4	1.0	59.3
	PartyScene	0.2	16.2	0.7	39.7	0.7	54.9
	RaceHorseC	0.4	17.4	0.8	26.9	1.1	44.4
Class D	BasketballPass	0.2	18.2	0.5	31.5	0.8	49.7
(416 × 240)	BQSquare	0.1	9.8	0.9	51.9	0.8	61.0
	Blowingbubbles	0.1	14.3	0.7	42.3	0.7	55.8
	RaceHorses	0.3	20.8	1.0	25.5	1.3	46.1
Class E	FourPeople	0.2	3.8	0.1	69.9	0.3	73.6
(1280 × 720)	Johnny	0.2	1.2	0.2	75.2	0.4	76.6
	KristenAndSara	0.2	2.2	0.3	71.0	0.5	73.4
	Average	0.3	11.1	0.6	47.7	0.8	58.5

TABLE V
COMPARISONS AMONG THE PROPOSED METHOD AND EXISTING WORKS UNDER THE RA CONDITION

		PanTB [34]		ESD [32]		ESD+CBF+ECU [16], [29], [32]		Proposed	
Resolution	Sequences	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A	PeopleOnstreet	0.4	21.5	0.4	20.8	2.4	32.5	0.8	50.0
(2560 × 1600)	Traffic	0.3	45.6	0.2	41.5	2.4	64.4	0.7	66.7
Class B	Kimono	0.2	35.0	0.3	35.1	1.5	52.3	0.7	52.7
(1920 × 1080)	ParkScene	0.3	41.7	0.3	39.8	1.9	60.2	0.8	63.5
	Cactus	1.6	41.4	0.3	35.1	2.6	53.4	0.6	57.6
	BasketballDrive	0.4	32.7	0.3	32.8	1.5	48.9	1.1	52.8
	BQTerrace	0.6	44.7	0.3	41.3	2.4	62.0	1.1	64.2
Class C	BasketballDrill	0.6	31.3	0.3	28.0	1.3	42.2	0.8	50.1
(832 × 480)	BQMall	0.8	38.8	0.4	34.4	2.6	52.7	1.0	59.3
	PartyScene	0.7	33.2	0.3	29.4	2.0	44.5	0.7	54.9
	RaceHorseC	0.6	20.9	0.3	20.3	2.0	31.4	1.1	44.4
Class D	BasketballPass	1.4	26.4	0.2	23.6	2.1	36.1	0.8	49.7
(416 × 240)	BQSquare	1.2	45.2	0.3	38.5	1.8	57.3	0.8	61.0
	Blowingbubbles	0.8	35.2	0.2	30.4	2.0	46.7	0.7	55.8
	RaceHorses	0.7	20.3	0.4	18.9	2.5	30.3	1.3	46.1
Class E	FourPeople	0.1	62.3	0.1	50.3	0.9	75.9	0.3	73.6
(1280 × 720)	Johnny	0.6	64.0	0.1	53.9	0.6	81.1	0.4	76.6
	KristenAndSara	0.4	58.7	0.2	50.8	0.9	76.8	0.5	73.4
	Average	0.7	38.8	0.3	34.7	1.9	52.7	0.8	58.5

C. Comparisons Among the Proposed Approaches and the State-of-the-Art Algorithms

To verify the validity of the proposed approach, six state-of-the-art algorithms are selected as benchmarks for comparisons under the RA and LD B conditions, respectively. The selected benchmarks include PanTB [34], ESD [32], combination of ESD+CBF+ECU [16], [29], [32], ShenTCSVT [19], ShenTMM [23] and AhnTCSVT [17]. To make fair comparisons, all these existing algorithms and the proposed approach are compared with the original HM12.0. Please note that LeeTB [33] is not selected for comparison because it puts emphasis on coding quality, and its reduction of computational complexity is almost equal to that of ESD [32]. Tables V and VI summarizes the experimental results of PanTB [34], ESD [32], ESD+CBF+ECU [16], [29], [32] and the proposed algorithm under the RA and LD B conditions, respectively. From Table V, we can observe that PanTB [34] achieves 38.8% encoding time saving with 0.7% BDBR increase on average. Compared with the PanTB [34] method, the proposed method further reduces 19.7% encoding time while maintaining similar RD performance. The reasons

TABLE VI
COMPARISONS AMONG THE PROPOSED METHOD AND EXISTING WORKS UNDER LD B CONDITION

		PanTB [34]		ESD [32]		ESD+CBF+ECU [16], [29], [32]		Proposed	
Resolution	Sequences	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A	PeopleOnstreet	0.3	18.7	0.3	16.8	1.3	25.9	0.5	47.2
(2560 × 1600)	Traffic	0.5	41.6	0.3	37.5	1.5	56.5	0.9	63.5
Class B	Kimono	0.4	31.1	0.3	29.0	1.2	43.4	0.6	46.8
(1920 × 1080)	ParkScene	0.6	38.1	0.3	34.1	1.5	51.2	0.8	60.2
	Cactus	2.2	37.8	0.4	30.2	1.8	45.5	0.9	53.4
	BasketballDrive	0.3	30.1	0.4	28.4	1.0	42.1	0.7	49.2
	BQTerrace	0.6	40.8	0.3	36.0	1.3	54.4	0.9	60.3
Class C	BasketballDrill	0.6	27.2	0.4	23.7	1.0	35.2	0.8	47.0
(832 × 480)	BQMall	0.8	33.3	0.4	28.8	1.5	43.4	0.9	55.9
	PartyScene	0.8	26.4	0.4	22.6	1.1	33.6	0.7	47.9
	RaceHorseC	0.5	18.9	0.3	17.2	1.0	25.9	0.8	42.7
Class D	BasketballPass	1.1	23.1	0.4	19.7	1.1	30.0	0.7	48.3
(416 × 240)	BQSquare	1.3	33.1	0.5	28.3	1.2	42.3	0.8	50.4
	Blowingbubbles	1.0	28.7	0.4	23.7	1.4	35.5	0.8	50.1
	RaceHorses	0.5	17.3	0.4	15.3	1.1	23.5	0.7	44.1
Class E	FourPeople	0.6	57.9	0.4	46.5	1.3	70.3	0.8	71.9
(1280 × 720)	Johnny	1.5	62.2	0.6	51.5	1.3	77.2	0.8	76.0
	KristenAndSara	1.1	54.9	0.5	47.1	1.1	70.8	1.1	71.2
	Average	0.8	34.5	0.4	29.8	1.3	44.8	0.8	54.8

TABLE VII
COMPARISONS AMONG THE PROPOSED METHOD AND EXISTING WORKS UNDER RA CONDITION

		ShenTCSVT [19]		ShenTMM [23]		AhnTCSVT [17]		Proposed	
Resolution	Sequences	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A	PeopleOnstreet	0.2	42.5	4.0	23.3	0.9	26.9	0.8	50.0
(2560 × 1600)	Traffic	1.1	60.5	2.1	44.1	0.8	61.6	0.7	66.7
Class B	Kimono	1.0	47.3	0.4	31.2	1.3	58.2	0.7	52.7
(1920 × 1080)	ParkScene	0.9	45.2	1.0	33.7	1.2	52.6	0.8	63.5
	Cactus	1.0	42.1	3.2	40.6	2.8	56.8	0.6	57.6
	BasketballDrive	1.0	41.8	1.4	28.7	2.0	50.9	1.1	52.8
	BQTerrace	1.1	49.7	1.2	35.7	1.6	54.4	1.1	64.2
Class C	BasketballDrill	2.1	39.8	5.2	34.1	1.9	45.2	0.8	50.1
(832 × 480)	BQMall	1.6	40.0	2.9	28.8	2.2	48.6	1.0	59.3
	PartyScene	1.1	45.8	2.5	27.0	0.8	37.7	0.7	54.9
	RaceHorseC	1.8	38.5	2.0	17.7	2.2	33.9	1.1	44.4
Class D	BasketballPass	1.8	28.9	2.2	16.7	1.5	33.6	0.8	49.7
(416 × 240)	BQSquare	0.4	34.4	0.4	22.9	0.6	45.1	0.8	61.0
	Blowingbubbles	1.3	33.7	2.6	23.4	0.7	38.2	0.7	55.8
	RaceHorses	1.8	24.4	2.1	12.8	1.1	26.6	1.3	46.1
Class E	FourPeople	1.4	66.3	2.9	56.9	1.7	74.1	0.3	73.6
(1280 × 720)	Johnny	0.9	67.8	1.0	57.8	1.3	75.7	0.4	76.6
	KristenAndSara	1.3	62.5	2.7	51.8	1.2	73.1	0.5	73.4
	Average	1.2	45.1	2.2	32.6	1.4	49.6	0.8	58.5

behind this are two-folds. First, the proposed approach reduces more coding time for motion-complex regions, benefiting from the large CU sizes which are rare-used for fast SKIP mode decision. Second, the hierarchical modes at different depth levels are further exploited. Meanwhile, compared with ESD [32], the proposed method also achieves more encoding time saving with less BDBR loss. The ESD+CBF+ECU [16], [29], [32] method achieves 52.7% encoding time reduction with 1.9% BDBR increment on average. The proposed method also reduces more encoding time and achieve better RD performance. From Table VI, the proposed method also outperforms the existing similar algorithms including (PanTB [34], ESD [32] and ESD+CBF+ECU [16], [29], [32]) under LD B condition in terms of the balanced performance of both encoding time reduction and BDBR loss.

Tables VII and VIII compare the proposed algorithm with ShenTCSVT [19], ShenTMM [23] and AhnTCSVT [17] under the RA and LD B conditions, respectively. ShenTCSVT [19] is a fast mode decision algorithm, ShenTMM [23] is a fast CU size decision algorithm and AhnTCSVT [17] includes both early SKIP mode decision and fast CU size decision. The performances of these algorithms are reported by comparing them with the original HM12.0, respectively. From Table VII,

TABLE VIII
PERFORMANCE COMPARISON OF PROPOSED METHOD WITH RECENT WORKS UNDER LD B CONDITION

		ShenTCSVT [19]		ShenTMM [23]		AhnTCSVT [17]		Proposed	
Resolution	Sequences	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A	PeopleOnstreet	0.4	32.5	2.2	19.9	1.5	28.0	0.5	47.2
(2560 × 1600)	Traffic	1.0	54.7	2.0	41.3	2.0	41.7	0.9	63.5
Class B	Kimono	0.8	38.0	0.3	27.8	0.8	47.9	0.6	46.8
(1920 × 1080)	ParkScene	0.8	40.5	0.9	28.9	1.1	48.3	0.8	60.2
	Cactus	0.5	43.5	3.1	36.4	2.2	48.0	0.9	53.4
	BasketballDrive	0.8	42.3	1.0	25.7	1.2	42.6	0.7	49.2
	BQTerrace	1.3	42.7	1.8	32.2	0.3	47.5	0.9	60.3
Class C	BasketballDrill	1.9	44.0	1.3	38.5	2.0	37.0	0.8	47.0
(832 × 480)	BQMall	2.0	43.5	2.2	25.3	1.2	40.9	0.9	55.9
	PartyScene	1.0	40.0	3.3	25.1	0.4	33.0	0.7	47.9
	RaceHorseC	1.0	34.1	1.1	14.0	1.1	26.1	0.8	42.7
Class D	BasketballPass	1.5	35.4	1.6	14.8	1.2	27.2	0.7	48.3
(416 × 240)	BQSquare	0.4	36.0	0.4	44.3	0.2	38.8	0.8	50.4
	Blowingbubbles	2.2	39.7	2.4	22.0	0.3	31.5	0.8	50.1
	RaceHorses	0.7	30.8	1.0	12.9	0.7	21.2	0.7	44.1
Class E	FourPeople	1.1	64.0	2.4	47.8	1.6	65.6	0.8	71.9
(1280 × 720)	Johnny	1.3	66.9	2.0	56.0	−0.3	73.9	0.8	76.0
	KristenAndSara	1.4	60.9	2.8	48.5	0.5	69.6	1.1	71.2
	Average	1.1	43.9	1.8	31.2	1.0	42.7	0.8	54.8

TABLE IX
OVERALL PERFORMANCE OF THE DIFFERENT FAST METHODS

	RA		LD B	
Methods	BDBR (%)	TS (%)	BDBR (%)	TS (%)
Proposed	0.8	58.5	0.8	54.8
PanTB [34]	0.7	38.8	0.8	34.5
ESD [32]	0.3	34.7	0.4	29.8
ESD+CBF+ECU [16], [29], [32]	1.9	52.7	1.3	44.8
ShenTCSVT [19]	1.2	45.1	1.1	43.9
ShenTMM [23]	2.2	32.6	1.8	31.2
AhnTCSVT [17]	1.4	49.6	1.0	42.7

we can observe that ShenTCSVT [19] achieves the minimum encoding time saving of 24.4% with a BDBR increase of 1.8%. For ShenTMM [23], its minimum encoding time saving is only 12.8% with a BDBR increase of 2.1%. AhnTCSVT [17] can reduce the encoding time of at least 26.6% with a BDBR increase of 1.1%. However, the proposed approach achieves the minimum encoding time saving of 44.1% with the BDBR increase of 1.1%. That is, compared with ShenTCSVT [19], ShenTMM [23] and AhnTCSVT [17], the proposed approach can save extra encoding time for about 13.4%, 25.9% and 8.9% with less BDBR increases, respectively. The great performance gain mainly benefits from the fact that compared with three existing approaches, the proposed approach achieves much better early SKIP mode decision for video sequences with complex motions. From Table VIII, the proposed approach also achieves superior performances over three existing approaches under LD B condition.

Table IX compares the performances among the proposed method, PanTB [34], ESD [32], ESD+CBF+ECU [16], [29], [32], ShenTCSVT [19], ShenTMM [23] and AhnTCSVT [17] under LD B and RA configurations, respectively. Apparently, the proposed method achieves more encoding time reduction with negligible RD performance loss compared with these state-of-the-art algorithms.

Fig. 9 reports the results of the proposed approach under different QPs (22, 27, 32, 37) for two typical video sequences including ParkScene and BasketballPass. As the QP increases, it still achieves more encoding time saving than the HM encoder, and it has almost no bitrate increase and no PSNR degradation.

TABLE X
CHARACTERISTICS OF THE SCENE CHANGE SEQUENCES

Sequences	Resolution	Frame rate	Encoding frames
BasketballPass_BlowingBubbles	416 × 240	50	960
PartyScene_BasketballDrill	832 × 480	50	960
Johnny_FourPeople	1280 × 720	60	1200
ParkScene_Kimono1	1920 × 1080	24	480
Traffic_PeopleOnStreet	2560 × 1600	30	300

D. Performance of the Proposed Method for Video Sequences With Scene Change

Video sequences with scene change are also tested. Table X summarizes five test video sequences, which are obtained by cascading two video sequences of different motion activities. For them, there is a scene change every 30 frames. For these five video sequences, the performance of the proposed method is reported in Table XI. From it, the proposed approach saves about 56% and 53% encoding time with only 0.7% and 0.6% BDBR increases under the RA and LD B configurations, respectively. Thus, the proposed method achieves desirable robustness

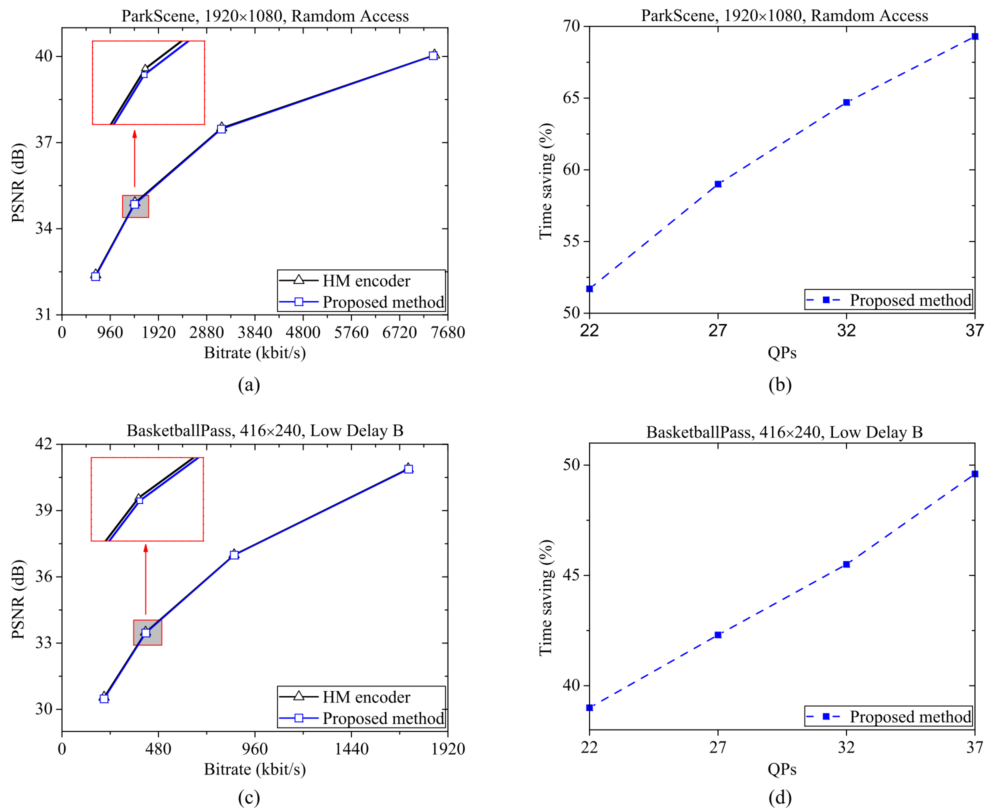


Fig. 9. Performance of *ParkScene* (1920 × 1080, 24 Hz) and *BasketballPass* (416 × 240, 50 Hz) under different QPs (22, 27, 32, 37). (a) RD curves of *ParkScene*. (b) Time saving of *ParkScene*. (c) RD curves of *BasketballPass*. (d) Time saving of *BasketballPass*.

TABLE XI
PERFORMANCE OF THE PROPOSED METHOD FOR SEQUENCES WITH SCENES CHANGE

Sequences	RA BDBR (%)	RA TS (%)	LD B BDBR (%)	LD B TS (%)
BasketballPass_BlowingBubbles	0.7	50.7	0.6	48.3
PartyScene_BasketballDrill	0.7	50.1	0.5	46.4
Johnny_FourPeople	0.3	69.7	0.4	67.9
ParkScene_Kimono1	0.8	55.8	0.7	51.4
Traffic_PeopleOnStreet	0.8	53.7	0.6	51.8
Average	0.7	56.0	0.6	53.2

because it also effectively reduces the total encoding time with negligible RD degradation.

V. CONCLUSION

This paper proposed an effective early SKIP mode decision approach for HEVC. The CU depth correlation is firstly exploited to categorize the CUs in inter-frames into either rare-used or frequent-used. For the rare-used CU size, the SKIP mode is directly selected as the optimal mode. For the frequent-used CU size, USM, which expresses the relationship between the SKIP mode and other modes, is established for early SKIP mode decision. Experimental results show that the proposed algorithm reduces 58.5% and 54.8% encoding time saving with 0.8% and 0.8% BDBR increase on average under RA and LD conditions, respectively.

ACKNOWLEDGMENT

The authors would like to thank Dr. Z. Q. Pan for providing the source codes [31], [34], [35] for experimental comparisons, and also the anonymous reviewers for their valuable comments.

REFERENCES

[1] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standardsłincluding high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[4] Y.-H. Sung and J.-C. Wang, "Fast mode decision for H.264/AVC based on rate-distortion clustering," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 693–702, Jun. 2012.

[5] B.-G. Kim, "Novel inter-mode decision algorithm based on macroblock (MB) tracking for the p-slice in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 273–279, Feb. 2008.

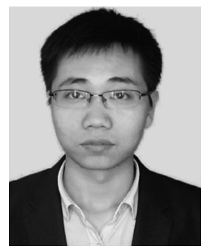
[6] L. Q. Shen, Z. Liu, P. An, R. Ma, and Z. Y. Zhang, "Low-complexity mode decision for MVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 837–843, Jun. 2011.

[7] C. H. Yeh *et al.*, "Fast mode decision algorithm through inter-view rate distortion prediction for multiview video coding system," *IEEE Trans. Ind. Inform.*, vol. 10, no. 1, pp. 594–603, Feb. 2014.

[8] T. S. Zhao, S. Kwong, H. L. Wang, and C.-C. J. kuo, "H.264/SVC mode decision based on optimal stopping theory," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2607–2618, May 2012.

[9] H. Q. Zeng, X. L. Wang, C. H. Cai, J. Chen, and Y. Zhang, "Fast multiview video coding using adaptive prediction structure and hierarchical mode decision," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1566–1578, Sep. 2014.

[10] Y. Li, G. B. Yang, Y. P. Zhu, C. Liu, and K. Liu, “Adaptive mode decision for multiview video coding based on macroblock position constraint model,” *J. Real-Time Image Process.*, vol. 12, pp. 575–582, Oct. 2016.
[11] L. Q. Shen, Z. Liu, T. Yan, Z. Y. Zhang, and P. An, “Early SKIP mode decision for MVC using inter-view correlation,” *Signal Process., Image Commun.*, vol. 25, no. 2, pp. 88–93, Feb. 2010.
[12] F. S. Wang, H. Q. Zeng, Q. H. Shen, and S. D. Du, “Efficient early direct mode decision for multi-view video coding,” *Signal Process. Image Commun.*, vol. 28, no. 7, pp. 736–744, Aug. 2013.
[13] Y. Zhang, S. Kwong, L. Xu, and G. Y. Jiang, “DIRECT mode early decision optimization based on rate distortion cost property and interview correlation,” *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 390–398, Jun. 2013.
[14] Z. Q. Pan, Y. Zhang, J. J. Lei, L. Xu, and X. M. Sun, “Early DIRECT mode decision based on all-zero block and rate distortion cost for multiview video coding,” *IET Image Process.*, vol. 10, no. 1, pp. 9–15, Jan. 2016.
[15] Y. Li, G. B. Yang, N. Chen, Y. P. Zhu, and X. L. Ding, “Early DIRECT mode decision for MVC using MB mode homogeneity and RD cost correlation,” *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 700–708, Sep. 2016.
[16] R. Gweon, Y. Lee, and J. Lim, “Early termination of cu encoding to reduce HEVC complexity,” JCT-VC doc. JCTVC-F045, Torino, Italy, USA, Jul. 2011.
[17] S. Ahn, B. Lee, and M. Kim, “A novel fast CU encoding scheme based on spatio-temporal encoding parameters for HEVC inter coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 422–435, Mar. 2015.
[18] J. Zhang, B. Li, and H. Li, “An efficient fast mode decision method for inter prediction in HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 8, pp. 422–435, Aug. 2016.
[19] L. Shen, Z. Zhang, and Z. Liu, “Adaptive inter-mode decision for HEVC jointly utilizing inter-level and spatiotemporal correlations,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1709–1722, Oct. 2014.
[20] W. Zhao, T. Onoye, and T. Song, “Hierarchical structure-based fast mode decision for H. 265/HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 10, pp. 1651–1664, Sep. 2015.
[21] Q. Hu, X. Zhang, Z. Shi, and Z. Gao, “Neyman–Pearson based early mode decision for HEVC encoding,” *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 379–391, Mar. 2016.
[22] S. H. Jung and H. Park, “A fast mode decision method in HEVC using adaptive ordering of modes,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 10, pp. 1846–1858, Oct. 2016.
[23] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, “An effective CU size decision method for HEVC encoders,” *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 465–470, Feb. 2013.
[24] J. Xiong *et al.*, “MRF-based fast HEVC inter CU decision with the variance of absolute differences,” *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2141–2153, Dec. 2014.
[25] J. Lee, S. Kim, K. Lim, and S. Lee, “A fast CU size decision algorithm for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 411–421, Mar. 2015.
[26] J. Xiong, H. Li, F. Meng, Q. Wu, and K. Ngan, “Fast HEVC inter CU decision based on latent SAD estimation,” *IEEE Trans. Multimedia*, vol. 12, no. 4, pp. 1121–1135, Dec. 2015.
[27] J. Vanne, M. Viitanen, and T. Hamalainen, “Efficient mode decision schemes for HEVC inter prediction,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1579–1593, Sep. 2014.
[28] T. Zhao, Z. Wang, and S. Kwong, “Flexible mode selection and complexity allocation in high efficiency video coding,” *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1135–1144, Dec. 2013.
[29] K. Choi, S. H. Park, and E. S. Jang, “Coding tree pruning based cu early termination,” JCT-VC doc. JCTVC-F092, Torino, Italy, Jul. 2011.
[30] C. Zhu, X. Lin, and L. P. Chau, “Hexagon-based search pattern for fast block motion estimation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 349–355, May 2002.
[31] Z. Q. Pan, Y. Zhang, and S. Kwong, “Efficient motion and disparity estimation optimization for low complexity multiview video coding,” *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.
[32] J. Yang, J. Kim, K. Won, and B. Jeon, “Early skip detection for HEVC,” JCT-VC doc. JCTVC-G543, Geneva, Switzerland, Nov. 2011.
[33] H. Lee, H. J. Shim, Y. Park, and B. Jeon, “Early Skip mode decision for HEVC encoder with emphasis on coding quality,” *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 388–397, Sep. 2015.
[34] 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 Trans. Broadcast.*, vol. 60, no. 2, pp. 405–412, Jun. 2014.
[35] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, “Fast motion estimation based on content property for low-complexity H.265/HEVC encoder,” *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 675–684, Sep. 2016.
[36] F. Bossen, “Common test conditions and software reference configurations,” JCT-VC doc. JCTVC-L1100, Jan. 2013.
[37] G. Bjontegaard, “Calculation of average PSNR differences between RD curves,” ITU-T SC16/Q6, VCEG-M33, Austin, TX, USA, Apr. 2001.
[38] T. S. Ferguson, “Optimal stopping and applications,” May 2012, accessed on: Sep. 2016. [Online]. Available: http://www.math.ucla.edu/tom/Stopping/Contents.html
[39] Z. Pan *et al.*, “Fast reference frame selection based on content similarity for low complexity HEVC encoder,” *J. Vis. Commun. Image Represent.*, vol. 40, Part B, pp. 516–524, Oct. 2016.

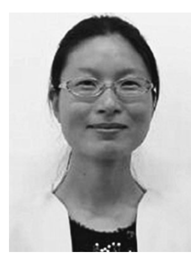


Yue Li received the B.S. degree from Hengyang Normal University, Hunan, China, in 2010, the M.S. degree from Central South University, Changsha, China, in 2013, and is currently working toward the Ph.D. degree at Hunan University, Changsha, China.

His current research interests include video coding and image processing.

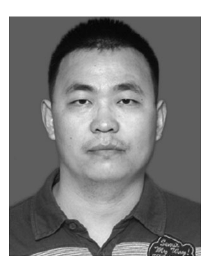
Gaobo Yang received the M.S. degree in information engineering from East China Jiaotong University, Nanchang, China, in 2001, and the Ph.D. degree in communication and information system from Shanghai University, Shanghai, China, in 2004.

He is a Professor with Hunan University, Changsha, China. He made an academic visit to University of Surrey, Surrey, U.K., from August 2010 to August 2011. He is the PI of several projects such as the Natural Science Foundation of China, Special Prophase Project on National Basic Research Program of China (973) and the program for New Century Excellent Talents in University. His research interests include image and video signal processing, digital media forensics.



Yapei Zhu received the B.S. degree from Hengyang Normal University, Hengyang, China, in 2010, and the M.S. degree from Ningbo University, Zhejiang, China, in 2013.

She is currently a Lecturer with Hengyang Normal University, Hengyang, China. Her research interests concentrate on image/video compression.



Xiangling Ding received the B.S. and M.S. degrees from Hunan Normal University, Hunan, China, in 2003 and 2006, respectively, and is currently working toward the Ph.D. degree at Hunan University, Changsha, China.

His current research interests include video information security and image processing.

Xingming Sun (M’07–SM’07) received the B.S. degree in mathematics from Hunan Normal University, Hunan, China, in 1984, the M.S. degree in computing science from the Dalian University of Science and Technology, Dalian, China, in 1988, and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2001.

He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include network and information security, digital watermarking, cloud computing security, and wireless network security.