

Adaptive Inter-Mode Decision for HEVC Jointly Utilizing Inter-Level and Spatiotemporal Correlations

Liquan Shen, Zhaoyang Zhang, and Zhi Liu, *Member, IEEE*

Abstract—High Efficiency Video Coding (HEVC) adopts the quadtree structured coding unit (CU), which allows recursive splitting into four equally sized blocks. At each depth level, it enables SKIP mode, merge mode, inter $2N \times 2N$, inter $2N \times N$, inter $N \times 2N$, inter $2N \times nU$, inter $2N \times nD$, inter $nL \times 2N$, inter $nR \times 2N$, inter $N \times N$ (only available for the smallest CU), intra $2N \times 2N$, and intra $N \times N$ (only available for the smallest CU) in inter-frames. Similar to H.264/AVC, the mode decision process in HEVC is performed using all the possible depth levels (or CU sizes) and prediction modes to find the one with the least rate distortion (RD) cost using Lagrange multiplier. This achieves the highest coding efficiency, but leads to a very high computational complexity. Since the optimal prediction mode is highly content dependent, it is not efficient to use all the modes. In this paper, we propose a fast inter-mode decision algorithm for HEVC by jointly using the inter-level correlation of quadtree structure and the spatiotemporal correlation. There exist strong correlations of the prediction mode, the motion vector and RD cost between different depth levels and between spatially temporally adjacent CUs. We statistically analyze the prediction mode distribution at each depth level and the coding information correlation among the adjacent CUs. Based on the analysis results, three adaptive inter-mode decision strategies are proposed including early SKIP mode decision, prediction size correlation-based mode decision and RD cost correlation-based mode decision. Experimental results show that the proposed overall algorithm can save 49%–52% computational complexity on average with negligible loss of coding efficiency, exhibiting applicability to various types of video sequences.

Index Terms—High Efficiency Video Coding (HEVC), mode decision, motion estimation (ME).

I. INTRODUCTION

ISO-IEC/MPEG and ITU-T/VCEG recently formed the Joint Collaborative Team on Video Coding (JCT-VC)

Manuscript received September 17, 2013; revised December 22, 2013 and February 8, 2014; accepted March 20, 2014. Date of publication March 26, 2014; date of current version October 1, 2014. This work was supported in part by the Shanghai Rising-Star Program under Grant 11QA1402400, in part by the Innovation Program of Shanghai Municipal Education Commission under Grant 13ZZ069, in part by the National Natural Science Foundation of China under Grant 61422111, Grant 60832003, Grant 60902085, Grant 61171084, Grant 61172096, and Grant U1301257, and in part by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Programme under Grant 299202. This paper was recommended by Associate Editor S. Takamura.

L. Shen and Z. Zhang are with the Key Laboratory of Advanced Display and System Application, Ministry of Education, Shanghai University, Shanghai 200072, China (e-mail: jsslq@163.com; zhzyzhang@shu.edu.cn).

Z. Liu is with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China, and also with IRISA/INRIA-Rennes, Campus Universitaire de Beaulieu, Rennes 35042, France (e-mail: liuzhisjtu@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/TCSVT.2014.2313892

[1], [2]. The JCT-VC has finished the next-generation video coding standard, High Efficiency Video Coding (HEVC). The quadtree structured coding unit (CU) is adopted in HEVC, and this kind of structure completely breaks the normal procedure of 16×16 macroblock (MB) coding architecture in H.264/AVC [3]. In the HEVC test model (HM), pictures are first divided into slices and slices are divided into a sequence of treeblocks. A treeblock is a square block (64×64 pixels) of luma samples together with two corresponding blocks of chroma samples, whose concept is broadly analogous to that of MBs in previous standards. The CU is the basic unit of the splitting region used for inter/intra prediction. The CU concept allows treeblock recursive splitting into four equally sized blocks. This process generates a content-adaptive coding tree structure comprised of CU blocks that may be as large as a treeblock or as small as 8×8 pixels. The prediction unit (PU) is the basic unit used for carrying the information related to the prediction processes. In general, it is not restricted to be square in shape to facilitate partitioning, which matches the boundaries of real objects in the picture. Intra-coded CUs have only two PU types, $2N \times 2N$ and $N \times N$ (available for the smallest CU) but inter-coded CUs have all of 8 (or 7) PU types, $2N \times 2N$, $2N \times N$, $N \times 2N$, $N \times N$ (available for the smallest CU), $2N \times nU$, $2N \times nD$, $nL \times 2N$, and $nR \times 2N$ [4].

Fig. 1 shows the architectures of quadtree structured CUs and PU partitioning. Fig. 1(a) and (b) shows the CU splitting procedure and PU partitioning, respectively. For the current CU in depth level (X), the procedure shown in Fig. 1(a) will be performed, and the current CU will be divided into four sub-CUs coming to the next depth level (X+1). The procedure also can be conducted on each sub-CU. To determine the best CU size (or depth level), HM tests each possible depth level. Meanwhile, a CU can be divided into several prediction partitions, as shown in Fig. 1(b). Prediction modes for CUs in inter-frames include SKIP mode, merge mode, inter $2N \times 2N$, inter $2N \times N$, inter $N \times 2N$, inter $N \times N$ (available for the smallest CU), inter $2N \times nU$, inter $2N \times nD$, inter $nL \times 2N$, inter $nR \times 2N$, intra $2N \times 2N$, and intra $N \times N$ (available for the smallest CU). Similar to the test model of H.264/AVC, the mode decision process in HM is performed using all the possible prediction modes to find the one with the least rate distortion (RD) cost using Lagrange multiplier (more details can be found in [5]). Therefore, both depth level decision and prediction mode decision cost the major computational complexity of a HEVC encoder, which should be reduced for the implementation of a fast encoding.

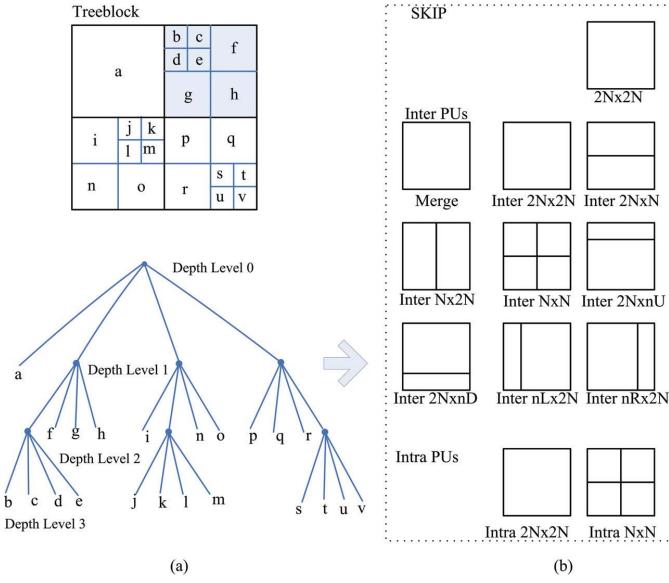


Fig. 1. CU splitting and PU partitioning. (a) CU splitting. (b) SKIP, inter PUs and intra PUs.

A number of fast algorithms [6]–[14] have been proposed to reduce the mode decision complexity for the previous video coding standard, H.264/AVC, achieving significant time saving with little loss of coding efficiency. Early termination (ET) strategies of mode decision based on all-zero block or SKIP mode detection are proposed in [6]–[9] to reduce the computation load of motion estimation (ME) in H.264/AVC. Methods introduced in [10] and [11] use frame difference and texture property to determine whether an MB belongs to homogeneous regions or not, and reduces the number of prediction modes. A highly efficient fast mode decision algorithm is proposed in [12] using multiphase nearest mean classification based on RD cost clustering. In our previous work, the motion homogeneity and inter-layer correlation are used to skip some unnecessary prediction modes of H.264/AVC [13], and its extension to H.264/SVC is presented in [14]. However, these methods are not directly applicable to the HEVC encoders, in which high computational complexity is intrinsically related to the use of new hierarchical quadtree structure. In addition, the previous algorithms are more favorably designed for the videos with a relatively lower spatial resolutions.

Recently, studies on reduction of computation complexity of HEVC encoders also have been reported, which can be roughly divided into three categories: training-based CU size decision, spatiotemporal correlation-based CU size prediction, and inter-level correlation-based CU size decision. Training-based CU size decision methods are proposed in [15]–[19]. CU splitting ET methods in [15]–[17] adopt the support vector machine (SVM) and the Bayesian decision rule to accelerate the CU depth decision process, the pyramid motion divergence (PMD) is used to help selecting CU size in [18], and mode filtering based on texture complexity analysis is used to reduce the number of candidates in [19]. These methods need to define threshold by training the data offline, and thus the threshold accuracy for the untrained video is a problem. Meanwhile,

property of each CU and the coding information from neighboring coded CUs are not adaptively exploited. Methods in the second category predict the best maximum depth level and the best prediction mode based on both spatial and temporal correlations observed in the coding tree depths. Complexity control methods of HEVC are proposed in [20] based on a decision algorithm that dynamically adjusts the depth range of CUs, and fast CU depth decision methods [21], [22] are proposed by reducing the depth search range based on the depth information correlation between the spatiotemporal adjacent coding tree units and the current coding tree unit. Adaptive CU depth range algorithms are proposed in [23] and [24] to reduce the candidate depth levels, and fast CU decision algorithms are proposed in [25] and [26] to accelerate the coding procedure in either frame level or CU level. A hierarchical complexity allocation scheme [27] is developed to allocate computational complexities among the frames based on linear programming. Meanwhile, candidate mode selection algorithms in [28] and [29] use the vicinity texture characteristics of the current encoding block. However, the property of the hierarchical quadtree structure in HEVC is not exploited in these methods. Methods in the third category based on analyzing coding information [including SKIP mode and coded_block_flag (CBF)] decide whether performing the remaining mode or CU size decision process. Zero transformed coefficient levels [30], [31] and SKIP mode [32], [33] are exploited to perform ET of CU size decision, and a fast CU size decision algorithm [34] uses RD cost of SKIP mode to terminate procedures of CU splitting. When the SKIP CUs and all-zero CUs are on the last level of quadtree or not existed, these methods become useless. Furthermore, coding information between different depth levels are not fully studied, and other correlation information including mode size, RD cost, and motion vector also can be exploited in the procedure of mode decision.

As shown in the previous researches, most of them only use the spatiotemporal correlation or SKIP mode correlation between different depth levels. There is still some room for further improvement in the mode decision process. To overcome these problems, this paper proposes an adaptive inter-mode decision for HEVC. The main idea of the proposed algorithm is that the coding information correlation between the depth levels and the spatiotemporal correlation among the adjacent CUs are jointly used to analyze motion and mode characteristics of the current CU and adjust mode decision strategies. It consists of three strategies: a novel early SKIP mode decision (ES-MD) method, a prediction size correlation-based mode decision (PSC-MD) method, and a RD cost correlation-based mode decision (RDC-MD) method. The features of the proposed algorithm are as follows. First, unlike the previously SKIP mode decision, ES-MD make decision based on SKIP mode correlation not only among spatially and temporally adjacent CUs, but also between inter-level. Second, PSC-MD reduces unnecessary prediction mode based on a mode complexity parameter, which is defined according to the mode context of spatially, temporally, and inter-level adjacent CUs. Third, RDC-MD uses the RDcost correlation between different depth levels and among the spatially and temporally adjacent CUs, and the RDcost values of the previously

TABLE I
PREDICTION MODE DISTRIBUTION IN INTER-FRAMES

Sequences	Depth Level	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra Modes
Shields (720×576)	0	54.8%	4.7%	4.8%	4.6%	17.8%	13.2%	0.1%
	1	71.2%	4.8%	2.6%	2.7%	8.2%	10.0%	0.4%
	2	81.9%	5.4%	2.4%	1.7%	3.0%	4.9%	0.7%
	3	89.4%	5.9%	1.9%	0.3%	1.0%	1.0%	0.7%
Basketball (720×576)	0	29.4%	13.4%	8.8%	11.8%	19.2%	15.5%	1.9%
	1	47.9%	13.0%	4.5%	4.7%	9.8%	17.7%	2.5%
	2	63.1%	12.8%	3.7%	3.9%	4.9%	9.0%	2.6%
	3	73.5%	13.5%	4.1%	1.0%	2.6%	1.8%	3.4%
RA-BQ (832×480)	0	40.9%	9.9%	4.4%	8.0%	15.3%	6.2%	15.2%
	1	54.6%	7.3%	3.8%	3.1%	6.7%	9.7%	14.8%
	2	67.1%	5.6%	3.5%	1.5%	3.1%	4.0%	15.2%
	3	76.1%	4.0%	2.5%	0.4%	1.4%	0.8%	14.6%
ShipCalendar (1280×720)	0	67.5%	4.0%	8.1%	7.0%	7.5%	5.9%	0.0%
	1	80.9%	5.3%	3.6%	2.2%	2.7%	5.3%	0.0%
	2	89.2%	5.0%	1.8%	0.9%	1.0%	2.1%	0.0%
	3	93.8%	4.3%	0.8%	0.2%	0.3%	0.7%	0.0%
StockholmPan (1280×720)	0	65.8%	2.0%	6.8%	8.3%	9.2%	7.8%	0.1%
	1	79.6%	2.9%	3.7%	3.2%	3.3%	7.1%	0.1%
	2	88.2%	3.8%	2.1%	1.4%	1.9%	2.3%	0.3%
	3	93.2%	4.2%	1.1%	0.4%	0.4%	0.5%	0.3%
Parkrun (1920×1088)	0	31.2%	12.8%	5.9%	17.1%	12.8%	17.6%	2.7%
	1	42.5%	17.7%	4.3%	7.2%	8.4%	17.5%	2.4%
	2	52.2%	20.4%	4.2%	6.6%	3.8%	10.2%	2.5%
	3	61.1%	24.8%	5.4%	1.9%	1.3%	1.8%	3.8%
Fireworks (1920×1088)	0	26.3%	10.8%	7.2%	14.3%	11.7%	11.1%	18.7%
	1	36.2%	9.9%	4.3%	9.8%	7.4%	17.6%	14.7%
	2	44.5%	11.9%	6.2%	8.3%	4.6%	11.9%	12.5%
	3	55.0%	15.7%	8.6%	3.2%	1.9%	2.2%	13.4%
Average		62.7%	9.1%	4.3%	4.8%	6.1%	7.7%	5.1%

*Small inter modes include Inter N×N, Inter 2N×nU, Inter 2N×nD, Inter nL×2N, Inter nR×2N.

encoded CUs are employed to derive the adaptive threshold for mode decision. To the best of our knowledge, there has been almost no work in a direction similar to that of the proposed algorithm.

The rest of this paper is organized as follows. Section II analyzes the prediction mode distribution and the prediction mode correlation between different depth levels. The proposed adaptive inter-mode decision algorithm is detailed in Section III. Experimental results and conclusion are given in Sections IV and V, respectively.

II. OBSERVATION AND STATISTICAL ANALYSIS

In HM, the mode decision process uses Lagrange multiplier to find the one with the least RD cost. RD cost function (J_{mode}) used in HM is evaluated as

$$J_{\text{mode}} = (\text{SSE}_{\text{luma}} + \omega_{\text{chroma}} \cdot \text{SSE}_{\text{chroma}}) + \lambda_{\text{mode}} \cdot B_{\text{mode}} \quad (1)$$

where B_{mode} is the bitrate cost dependent on each decision case, SSE is the average difference between the current CU and the matching block, ω_{chroma} is the weighing factor for chroma component, and λ is the Lagrange multiplier. Choosing a large mode size (or CU size) means that a small number of bits are required to signal the choice of motion vector(s) and the type of prediction mode, but the residual between the

current CU and the matching block may contain a significant amount of energy in the region with complex motion or rich texture. Choosing a small mode size (or CU size) may give a smaller SSE after ME, but requires a larger number of bits to signal the motion vector(s) and type of prediction. For regions with homogenous motion or static background, it is sufficient to represent motion at a coarse level, and coding using larger block sizes will not result in a much larger residual. For regions with complex motion and regions containing motions from different objects, it is suitable to represent motion at a fine level, and smaller block sizes are more likely to be selected as the optimal coding mode. If we can exploit CU motion or texture complexity to determine those CUs that have a higher probability to be coded using larger block sizes, we can skip the time-consuming process of computing RD costs on smaller block sizes for a high percentage of CUs, and thus significantly reduce the computation complexity of the whole coding process. The proposed inter-mode decision algorithm is based on CU properties evaluated by entire coding information from adjacent CUs including parent CUs in the upper depth level, spatially adjacent CUs and temporally adjacent CUs.

The prediction mode distribution and the prediction mode correlation between different depth levels are analyzed using the HM 10.0 encoder in Tables I and II and Fig. 2.

TABLE II
CONDITIONAL PROBABILITY OF THE PREDICTION MODE AT DEPTH LEVEL X(%)

Depth level : X-1	Basketball							StockholmPan						
	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra Modes	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra Modes
SKIP Mode	95.4	1.3	0.7	0.6	1	0.9	0.1	98.5	0.4	0.3	0.3	0.3	0.2	0.0
Merge Mode	40.9	39.1	4.5	1.8	4.2	5.4	4.2	52.8	30.8	4.4	1.2	1.9	2.8	6.1
Inter 2N×2N	52.5	18.5	12.8	2.5	4.3	5.6	3.8	74.5	9.1	8.7	2.0	1.7	3.3	0.7
Inter 2N×N	36.4	19.2	14.2	5	9.2	10.6	5.5	62.2	9.4	10.3	4.3	3.6	9.2	0.9
Inter N×2N	38.5	18.2	11.8	5.2	8.9	11.6	5.8	72.8	8.5	7.0	3.2	3.5	4.2	0.8
Small inter modes	38.8	18.6	8.3	4.2	12.3	13.7	4.1	52.4	11.1	6.1	3.9	14.8	10.9	0.9
Intra Modes	7.9	7.3	8	2.8	5.6	4.2	64.2	29.1	19.2	4.5	1.3	2.5	1.9	41.6

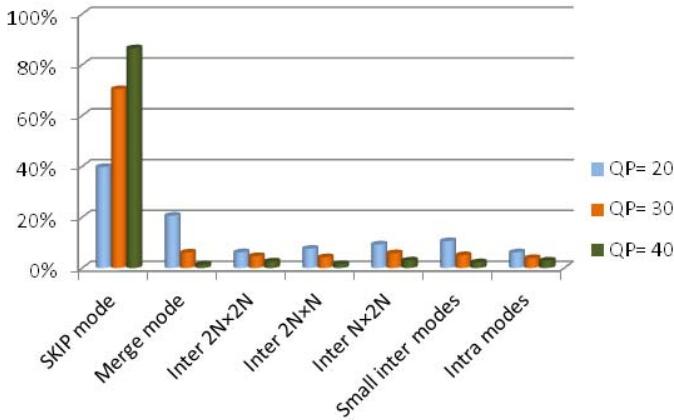


Fig. 2. Prediction mode distribution using different QPs.

We encode seven sequences with different motion activities and spatial resolutions of 720×576 , 832×480 , 1280×720 , and 1920×1088 . Among these test sequences, *Basketball* and *Fireworks* are with a large global/local motion, while *Shields*, *ShipCalendar*, *StockholmPan*, and *Parkrun* are with a medium local motion or a smooth motion. *RA-BQ* is a scene-cut sequence, which is generated by cascading two sequences (*RaceHorses* and *BQMall*) for scene change. It contains multiple scenes and every 10 frames has a scene cut frame. Test conditions are set as follows: fast encoder decision (FEN), fast decision for Merge RD cost (FDM) and the search mode enhanced predictive zonal search (EPZS) enabled, the hierarchical B frame structure with a group (GOP) length of 8; quantization parameter (QP) is chosen with 20, 30, and 40; search range of ME is configured with 64; the number of frames equals to 60, treeblock size equals to 64 pixels and depth range is from 0 to 3; Table I shows the mode distribution of CUs in inter-frames at each depth level. Fig. 2 shows prediction mode distribution under different QPs.

Most of the CUs in inter-frames choose SKIP mode as the best mode, and the percentage distribution is similar at each depth level. The average percentage of SKIP mode CUs is 63%, while the total average percentage of CUs with other modes is about 37% in Table I, because stationary regions and homogeneous regions prevail in natural video sequences and these two types of regions mostly select SKIP mode.

About 68%–94% of CUs are usually coded as SKIP mode for the low motion sequences such as *ShipCalendar*. Although basketball contains large areas of fast movement, about 29%–74% of CUs still choose SKIP mode. The proportion of SKIP mode CUs increases at each depth level in Fig. 2 as the QP increases. If we can decide in advance for a CU whether the optimal prediction mode is SKIP mode or not, the wasteful process of variable size ME can be omitted, and thus a huge amount of computation can be saved.

With similar video characteristic, the prediction mode of a CU at the depth level X is strongly related to that of its parent CU at the depth level X-1. To characterize this correlation, we show the conditional probability of the prediction mode at the depth level X for two typical sequences in Table II, where *StockholmPan* is with low and simple motion and *Basketball* is with high and complex motion. It can be seen that the prediction mode of the parent CU can be a good reference for the prediction of the coded mode for the current CU. For example, when the parent CU is coded with the large inter-modes (SKIP mode, merge mode, and inter $2N \times 2N$), the current CU is unlikely to choose other inter-modes including inter $2N \times N$, inter $N \times 2N$, and inter $N \times N$. When the parent CU uses small inter-modes, the current CU tends to increase the percentage of small inter-modes. When the best mode at the depth level X-1 is intra-modes, most of best modes at the current depth level are intra-modes.

Meanwhile, spatially adjacent CUs usually hold similar textures in natural videos. Therefore, the optimal prediction mode of the current CU may have a strong correlation with its spatially adjacent CUs. Besides, the temporal similarity between the successive frames or treeblocks exists in natural videos, and the optimal prediction mode of the current CU does not exhibit a wide variation from that of the corresponding CUs in the reference frames.

To use the inter-level correlation and the spatiotemporal correlation, we define a set of predictors, Ω , as

$$\Omega = \{CU_1, CU_2, CU_3, CU_4, CU_5, CU_6, CU_7, CU_8\}. \quad (2)$$

As shown in Fig. 3, CU_1 , CU_2 , CU_3 , and CU_4 denote the spatially adjacent CUs of the current CU (CU_0), CU_5 , and CU_6 denote temporally adjacent CUs located at the same position as CU_0 in the reference frames, and CU_7 and CU_8 denote the parent CUs at the upper depth levels.

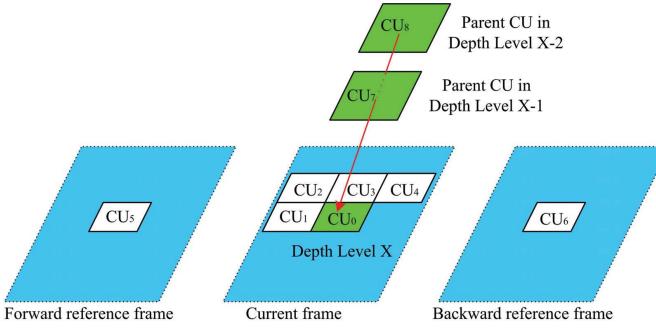


Fig. 3. Predictors and the current CU.

III. PROPOSED ADAPTIVE INTER-MODE DECISION ALGORITHM

A. Novel ES-MD

SKIP mode provides a good coding performance and requires a little computational complexity, where the motion vector predictor is adopted for the current MB (or CU) to generate the predicted block [35]. Meanwhile, SKIP mode is the dominant mode at low bitrates (high QPs) in HEVC, and the distribution is similar to that in the previous video coding standard, H.264/AVC. Once SKIP mode can be predecided, the variable-sized ME computation of a CU can be entirely skipped. Usually, the decision to use SKIP mode is delayed until the RD costs of all other modes (inter- and intra-modes) have been calculated and SKIP mode is found to have the minimum RD cost. Based on this consideration, we first propose a novel ES-MD strategy for HEVC to avoid the whole variable-sized ME process as well as evaluations on intra-modes.

To analyze the SKIP mode correlation, we define four types of CUs, I, II, III, and IV: I represents the CU whose spatially adjacent CUs (including CU₁, CU₂, CU₃, and CU₄ in Ω) and co-located CUs in the reference frames (CU₅ and CU₆) all choose SKIP mode; II represents the CU whose parent CUs in the two upper depth levels (CU₇ and CU₈) both choose SKIP mode as the optimal mode; III represents the CU whose parent CU in the upper depth level (CU₇) and spatially adjacent CUs (or CU₇ and temporally adjacent CUs) all choose SKIP mode as the optimal mode; and IV represents the remaining CUs of the combinations of I, II, and III. By exploiting the exhaustive mode decision in HM 10.0 under the aforementioned test conditions in Section II, we calculate the SKIP mode distribution in Table III for these four types of CUs using video sequences with different frame rates.

As shown in Table III, the probabilities of SKIP mode CUs for the types of I, II, and III are more than 99%, with the maximum of 99.9% in RA-BQ, and the minimum of 96.7% in Parkrun. The probabilities of SKIP mode CUs are higher for types of I, II in high frame rate sequences due to strong temporal correlation between the frames. However, the average probability of SKIP mode CUs is less than 50% for the type IV. Therefore, for types of I, II, and III, it is reasonable to consider SKIP mode as the only candidate mode. If the encoder can decide SKIP mode for CUs in types I, II, and III at the early stage, then variable-sized ME can be skipped and coding time can be reduced dramatically.

B. Prediction Size Correlation-Based Mode Decision

At each depth level, various prediction mode sizes are used in the prediction procedures. Large sizes are always chosen for CUs in the homogeneous region, and small sizes are chosen for CUs with active motion or rich texture [36]. Based on this observation, all the prediction modes available in HEVC are classified into four activity classes, and the mode-weight factor can be assigned based on the complexity of each mode, as summarized in Table IV. A mode complexity (MC) parameter of the current CU can be defined according to the mode context of the available predictors in Ω as

$$MC = \frac{\sum_{i=1}^N w_i \cdot k_i \cdot \alpha_i}{\sum_{i=1}^N k_i \cdot \alpha_i} \quad (3)$$

where N is the number of CUs equal to 8 and w_i is the mode-weight factor. Only the prediction modes of those available adjacent CUs in Ω will be used. Hence, k_i is set to 1, when CU _{i} is available; otherwise, k_i is set to 0. α_i is the CU-weight factor, which is assigned to the adjacent CUs according to their correlation from the current CU. The stronger correlation between the neighboring CU and the current CU, the larger the weight should be assigned. Experiments are conducted to compute the correlation degree of neighboring CUs and the current CU. Test sequences and test conditions in Section II are adopted, and the results are shown in Table V. Meanwhile, eight CU-weight factors also have an additional property, $\sum_{i=1}^8 \alpha_i = 1$. Based on correlation, the CU-weight factor for each adjacent CU is summarized in Table VI. It should be noted that, when the selected optimal depth level of a spatially or temporally adjacent CU is smaller than the current depth level, its optimal mode size is considered to be $2N \times 2N$; when the selected optimal depth level of a spatially or temporally adjacent CU is larger than the current depth level, its optimal mode is considered to be inter $N \times N$ or intra $N \times N$.

Generally, the more complexity the CU has, the larger value of the mode complexity will be. Based on the mode complexity, each CU is classified as simple mode, normal mode or complex mode as

$$\begin{cases} MC < T_{r0} & CU \in \text{simple mode} \\ T_{r0} \leq MC < T_{r1} & CU \in \text{normal mode} \\ MC \geq T_{r1} & CU \in \text{complex mode} \end{cases} \quad (4)$$

where T_{r0} is equal to the mode-weight factor of inter $2N \times 2N$ or intra $2N \times 2N$, i.e., 1, and T_{r1} is equal to the mode-weight factor of inter $N \times N$, i.e., 4.

For a CU with simple mode, the area covered by the current CU and its adjacent CUs usually contains slow-motion content or homogeneous texture, and the optimal prediction modes of its adjacent CUs are usually within inter $2N \times 2N$, intra $2N \times 2N$, merge mode, and SKIP mode. Therefore, a CU with simple mode only needs ME on size of $2N \times 2N$ and performing intra $2N \times 2N$ prediction. For a CU with complex mode, most of its adjacent CUs choose small-size inter-modes or intra $N \times N$, and thus it is not necessary to perform ME

TABLE III
STATISTICAL ANALYSIS OF SKIP MODE DISTRIBUTION FOR EACH TYPE OF CUs IN INTER-FRAMES

Sequences	Frames per second (FPS)=15								FPS=30							
	I		II		III		IV		I		II		III		IV	
	SKIP mode (%)	Other Modes (%)	SKIP Mode (%)	Other modes (%)	SKIP Mode (%)	Other Modes (%)	SKIP Mode (%)	Other Modes (%)	SKIP mode (%)	Other Modes (%)	SKIP Mode (%)	Other modes (%)	SKIP Mode (%)	Other Modes (%)	SKIP Mode (%)	Other Modes (%)
Shields	98.7	1.3	99.5	0.5	99.2	0.8	58.0	42.0	99.3	0.7	99.5	0.5	99.3	0.7	59.9	40.1
Basketball	98.6	1.4	99.4	0.6	99.2	0.8	36.2	63.8	98.9	1.1	99.5	0.5	99.3	0.7	43.3	56.7
RA-BQ	99.9	0.1	99.6	0.4	99.3	0.7	28.9	71.1	99.8	0.2	99.6	0.4	99.4	0.6	37.9	62.1
ShipCalendar	99.0	1.0	99.6	0.4	99.3	0.7	60.6	39.4	99.5	0.5	99.5	0.5	99.4	0.6	67.7	32.3
StockholmPan	99.5	0.5	99.8	0.2	99.7	0.3	65.4	34.6	99.6	0.4	99.8	0.2	99.6	0.4	66.3	33.7
Parkrun	96.7	3.3	98.0	2.0	99.1	0.9	33.9	66.1	97.5	2.5	98.6	1.4	98.3	1.7	37.7	62.3
Fireworks	99.5	0.5	99.6	0.4	99.1	0.9	25.2	74.8	99.7	0.3	99.6	0.4	99.0	1.0	27.2	72.8
Average	98.8	1.2	99.3	0.7	99.3	0.7	44.0	56.0	99.2	0.8	99.4	0.6	99.2	0.8	48.6	51.4

Note: low frame rate (15 fps) sequences are achieved by temporal downsampling of the original ones

TABLE IV
MODE-WEIGHT FACTORS ASSIGNED TO EACH PREDICTION MODE

Class	Mode	Motion activity	Mode-weight factor
1	SKIP mode	Motionless or homogeneous motion,	0
2	Inter $2N \times 2N$, Merge mode, Intra $2N \times 2N$	Slow motion or homogeneous texture	1
3	Inter $2N \times N$, Inter $N \times 2N$	Moderate motion	2
4	Inter $N \times N$, Inter $2N \times nU$, Inter $2N \times nD$, Inter $nR \times 2N$ and Inter $nL \times 2N$, Intra $N \times N$	Fast motion, Highly-textured region or new object	4

TABLE V
CORRELATION BETWEEN THE CURRENT CU AND ITS NEIGHBORING CUS

Index (i) in Fig. 3	1	2	3	4	5	6	7	8
correlation	0.76	0.71	0.78	0.69	0.55	0.54	0.75	0.66

TABLE VI
CU-WEIGHT FACTORS ASSIGNED TO CUS IN Ω

Index (i) in Fig. 3	1	2	3	4	5	6	7	8
α_i	0.2	0.1	0.2	0.1	0.05	0.05	0.2	0.1

TABLE VII
CANDIDATE MODES FOR EACH TYPE OF CUS

CU type	Candidate modes
CUs with simple mode	SKIP mode, Merge mode, Inter $2N \times 2N$ and Intra $2N \times 2N$
CUs with complex mode	Small inter modes, Intra $N \times N$ SKIP mode and Merge mode
CUs with normal mode	SKIP mode, all inter modes and intra modes

on sizes of $2N \times 2N$, $2N \times N$, and $N \times 2N$. Based on the above analysis, the candidate prediction modes that will be tested using RD optimization for each CU are summarized in Table VII.

To verify the validity of candidate modes shown in Table VII, extensive experiments have been conducted by exploiting the exhaustive mode decision under the aforementioned test sequences and test conditions in Section II. It can

be observed from Tables VIII and IX that for CUs with simple mode, the average probabilities of choosing SKIP mode, merge mode, inter $2N \times 2N$, and intra $2N \times 2N$ are 92.0%, 3.9%, 1.7%, and 0.9%, respectively, and the average probability of remaining mode is not greater than 0.5%. The percentage of optimal modes that are covered by the selected candidate modes reaches about 98.5%. It is higher for sequences in high frame rate. For CUs with complex mode, the average probabilities of choosing inter $2N \times 2N$ and intra $2N \times 2N$ are 1.2% and 0.6%, respectively, and that of inter $2N \times N$ and inter $N \times 2N$ is about 1%. The total probabilities of small inter-modes and intra $N \times N$ are about 89.2%. The percentage of optimal modes that are covered by the selected candidate modes can reach about 96%. For the *ShipCalendar* and *StockholmPan* sequence, the percentage of intra-modes for CUs with complex mode is very low, i.e., 0.0% because it contains a large area having smooth motion or with the static background, where inter-modes are more likely to be chosen.

TABLE VIII
STATISTICAL ANALYSIS OF PREDICTION MODE DISTRIBUTIONS FOR DIFFERENT TYPES OF CUs (F/S = 30) (%)

	CUs in simple mode region								CUs in complex mode region							
	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra 2N×2N	Intra N×N	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra 2N×2N	Intra N×N
Shields	96.6	1.9	0.7	0.1	0.3	0.3	0.1	0.0	8.6	0.3	0.9	0.9	4.5	84.9	0.0	0.0
Basketball	94.0	3.4	1.1	0.2	0.5	0.5	0.2	0.0	4.4	0.6	0.4	0.1	0.6	89.4	0.0	4.5
RA-BQ	94.3	2.3	1.2	0.2	0.5	0.4	1.2	0.1	5.8	0.4	0.0	0.3	1.5	52.2	1.5	38.2
ShipCalendar	97.0	1.7	0.6	0.1	0.2	0.4	0.0	0.0	11.1	2.0	0.9	0.3	1.1	84.5	0.0	0.0
StockholmPan	97.7	1.3	0.5	0.1	0.2	0.2	0.1	0.0	12.9	0.2	2.3	1.1	1.1	82.5	0.0	0.0
Parkrun	86.3	10.4	1.3	0.5	0.4	0.9	0.2	0.1	5.4	1.2	1.0	1.1	0.6	87.0	0.1	3.6
Fireworks	89.1	2.2	3.8	1.3	0.7	0.7	1.8	0.3	5.4	0.1	0.3	0.3	0.1	77.4	1.2	15.1
Average	93.6	3.3	1.3	0.4	0.4	0.5	0.5	0.1	7.7	0.7	0.8	0.6	1.3	79.7	0.4	8.8

When the selected optimal depth level of a CU is larger than the current depth level, its optimal mode is considered to be Inter N×N or Intra N×N in Tables 8-9.

TABLE IX
STATISTICAL ANALYSIS OF PREDICTION MODE DISTRIBUTIONS FOR DIFFERENT TYPES OF CUs (F/S = 15) (%)

	CUs in simple mode region								CUs in complex mode region							
	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra 2N×2N	Intra N×N	SKIP Mode	Merge mode	Inter 2N×2N	Inter 2N×N	Inter N×2N	Small inter modes	Intra 2N×2N	Intra N×N
Shields	95.0	2.6	1.2	0.2	0.4	0.4	0.2	0.0	6.7	0.3	1.3	0.0	3.1	82.4	0.0	6.2
Basketball	89.7	5.8	2.1	0.4	0.7	0.6	0.6	0.1	2.6	0.5	1.0	0.9	1.0	79.7	0.2	14.1
RA-BQ	86.5	3.5	3.1	0.3	0.8	0.4	5.2	0.3	1.0	0.3	1.5	0.5	1.4	44.9	2.2	48.1
ShipCalendar	94.0	2.6	2.0	0.2	0.4	0.3	0.4	0.0	3.4	0.2	3.1	1.7	4.6	86.6	0.0	0.5
StockholmPan	97.4	1.3	0.7	0.2	0.2	0.2	0.1	0.0	8.1	0.3	2.1	1.4	1.5	85.7	0.0	0.9
Parkrun	81.5	12.9	2.3	0.7	0.6	1.1	0.5	0.2	6.9	1.8	1.3	1.5	1.1	74.4	0.8	12.2
Fireworks	89.4	1.7	3.7	1.3	0.7	0.6	2.2	0.5	3.4	0.3	0.5	0.6	0.3	60.4	2.0	32.6
Average	90.5	4.4	2.1	0.5	0.5	0.5	1.3	0.2	4.6	0.5	1.5	0.9	1.8	73.4	0.7	16.4

C. RD Cost Correlation-Based Mode Decision

In HM, the mode decision process exhaustively searches the best mode from candidate modes according to RD optimization scheme. It is analyzed in [37] and [38] that the majority of best prediction modes after mode decision in Inter-frames are usually with large size modes such as SKIP mode, merge mode, and inter 2N × 2N, which implies that small-sized ME and intra prediction are unnecessary in most of the cases. So, it is better to have a proper ET strategy in the midway of fast mode decision algorithm. Basically, ET strategies are all threshold-based, and most of the ET threshold determination algorithms are based on the RD cost prediction. Meanwhile, a mass amount of inter-level, spatial, and temporal correlations exist in the coding procedure of HEVC. This leads to the fact that, the resultant RD cost information of the current CU is intimately related to that of its adjacent CUs. Thus, the RD cost of the current CU ($\text{RDcost}_{\text{pre}}$) can be predicted using the minimal RD cost values from the available predictors in Ω as

$$\text{Rdcost}_{\text{pre}} = \frac{\sum_{i=1}^6 \alpha_i \cdot k_i \cdot \text{Rdcost}_i + \alpha_7 \cdot k_7 \cdot \text{Rdcost}_7 / 4 + \alpha_8 \cdot k_8 \cdot \text{Rdcost}_8 / 16}{\sum_{i=1}^8 k_i \cdot \alpha_i} \quad (5)$$

where Rdcost_i are the RD cost of CU_i , and Rdcost_7 and Rdcost_8 are the RD costs of parent CUs in the two upper depth levels. α_i and k_i are defined as the same as those in (3).

The threshold is determined based on the minimum value among $\text{RDcost}_{\text{pre}}$ and RD costs of spatial neighboring predictors in Ω

$$\text{Thr} = \lambda \cdot \min \{ \text{RDcost}_1, \text{RDcost}_2, \text{RDcost}_3, \text{RDcost}_4, \text{RDcost}_{\text{pre}} \} \quad (6)$$

where λ is the adjustment parameter. To set the value of λ , we should consider the signaling information CBF in the HEVC syntax, which specifies the nonzero transform coefficient levels. CBF will be generated after checking each prediction mode. If CBF of a test mode equals to zero, the current CU is an all-zero block. In this condition, the adjustment parameter is reset with $1.25 \cdot \lambda$. When the minimal RD cost value is smaller than Thr , the mode decision procedure is terminated in the proposed method, and small-sized ME and intra prediction are skipped.

By exploiting the exhaustive mode decision in HM 10.0 encoder under the aforementioned test conditions and test sequences in Section II, we investigate the relationship between λ and the average accuracy of ET/ the percentage of early terminated CUs in Fig. 4(a) and the relationship between λ and the coding efficiency loss in Fig. 4(b). The accuracy of ET (ρ) is defined as, $S/T \times 100\%$, where T denotes

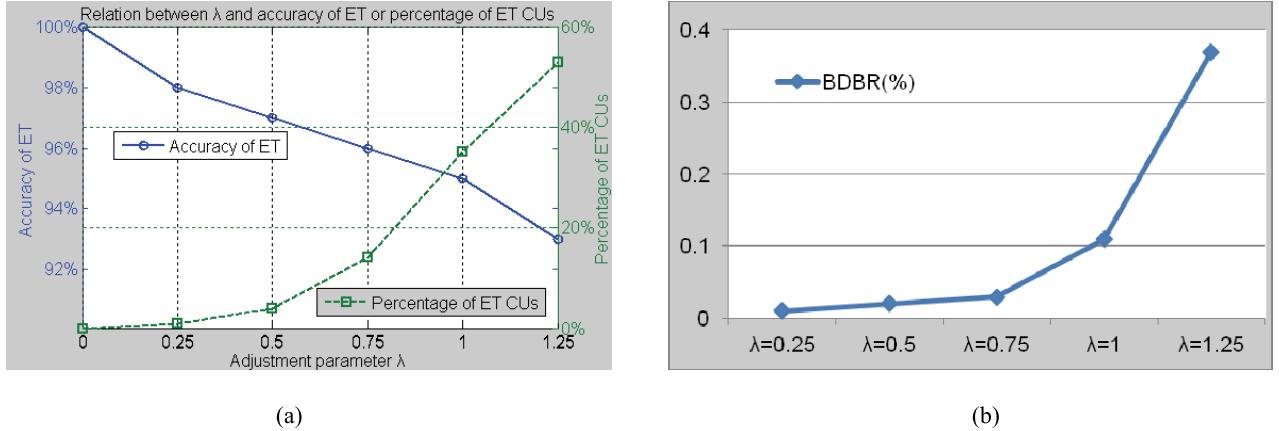


Fig. 4. Relation between λ and the accuracy of ET, the percentage of early terminated CUs and increase of BDBR (QPs are with 10, 20, 30 and 40). (a) Relation between λ and accuracy of ET, the percentage of ET CUs. (b) Relation between λ and BDBR.

the total number of early terminated CUs, and S denotes the number of CUs having the same optimal mode chosen by exhaustive mode decision among T CUs. Coding efficiency loss is measured by bit rate increase based on the Bjøntegaard method [39]. When λ becomes larger, both the mode decision complexity and the accuracy of ET decline. Meanwhile, the loss of coding efficiency will increase. The selected value of λ should keep coding efficiency and accuracy of ET while reducing the complexity greatly. As shown in Fig. 4(a), when $\lambda < 1$, the percentage of early terminated CUs increases considerably with the increase of λ . Meanwhile, the ET accuracy decreases relatively slowly, and Bjøntegaard delta bit rate (BDBR) increase is smaller in Fig. 4(b); when $\lambda > 1$, the ET accuracy decreases faster with λ value increasing, and thus, it leads to a significant increase of BDBR. We can also see that when the value of λ is selected to be 1, a large complexity saving can be achieved with a lower increase of BDBR (about 0.1%). More than 34.3% CUs can be early terminated without performing exhaustive mode decision. Thus, the λ value of 1 is employed in our work.

D. Overall Algorithm

Based on the aforementioned analysis, including strategies of ES-MD, PSC-MD, and RDC-MD, the proposed adaptive inter-mode decision algorithm for HEVC is summarized as follows.

Step 1: Start mode decision for CUs in inter-frames.

Step 2: Derive coding information of parent CUs in the upper depth levels and spatially/temporally adjacent CUs.

Step 3: Test SKIP and merge modes. If the current CU meets one of the three ES-MD conditions in Section III-A, go to Step 7.

Step 4: Compute MC using (3). When the current CU is with simple mode, the candidate mode is inter $2N \times 2N$ and intra $2N \times 2N$; when the current CU is with complex mode, the candidate modes include inter $N \times N$ (available for the smallest CUs), inter $2N \times nU$, inter $2N \times nD$, inter $nL \times 2N$, inter $nR \times 2N$, and intra $N \times N$ (available for the smallest CUs); otherwise, the candidate modes include all prediction modes.

Step 5: Compute Thr for ET using (5) and (6).

Step 6: Loop each candidate inter- and intra-mode.

Step 6.1: Perform ME and get RD cost and CBF of the current mode;

Step 6.2: If the RD cost value is smaller than Thr, terminate the mode decision procedure and go to Step 7.

End Loop

Step 7: Determine the best prediction mode for the current CU.

IV. EXPERIMENTAL RESULTS

A. Test Conditions

To evaluate the performance of the proposed mode decision algorithm, the proposed algorithm is implemented on the recent HEVC reference software (HM 10.0). Two coding constraint conditions are defined in the Call for HEVC Proposals [40]: constraint set 1 (CS1) and constraint set 2 (CS2). CS1 (a random access case) corresponds to a broadcast scenario with a maximum GOP size of 8 and CS2 (a low delay case) corresponds to a low-delay scenario with no picture reordering. The above two constraint conditions restricting coding parameters and the temporal coding structure are used in all tests. Coding treeblock has a fixed size of 64×64 pixels (for luma) and a maximum depth level of 4, resulting in a minimum CU size of 8×8 pixels. The initial search range is 64 in both horizontal and vertical directions. FEN, FDM and the search mode EPZS enable. The proposed algorithm is evaluated with four QPs of 22, 27, 32, and 37 using test sequences recommended by JCT-VC with five resolutions ($416 \times 240/832 \times 480/1280 \times 720/1920 \times 1080/2560 \times 1600$). Note that, the seven test sequences, which are used to verify the validity of the proposed algorithm in Section III, are not used as test sequences. The experimental results are presented in Tables X–XIII and Figs. 5–10, where coding efficiency is measured with Peak signal-to-noise ratio (PSNR) and bitrate and computational complexity is measured with the consumed encoding time. BDBR (%) are used to represent the average coding efficiency differences [39], and DT (%) represents coding time change in percentage. Positive and negative values represent increments and decrements, respectively. Code built with C++ compiler using the function

TABLE X

RESULTS OF THE PROPOSED ES-MD, PSC-MD, AND RDC-MD COMPARED WITH HM ENCODERS UNDER RANDOM-ACCESS CONDITION

	Sequences	FPS	ES-MD		PSC-MD		RDC-MD	
			BDBR (%)	DT (%)	BDBR (%)	DT (%)	BDBR (%)	DT (%)
Class A (2560×1600)	Traffic	30	0.1	-41.4	0.7	-36.8	0.2	-11.8
	PeopleOnStreet	30	0.1	-16.7	0.7	-27.1	0.3	-11.9
Class B (1080p)	BasketballDrive	50	-0.1	-35.1	0.3	-35.3	0.6	-17.2
	BQTerrace	60	0.2	-46.7	0.4	-40.4	0.4	-18.2
	Cactus	50	0.2	-40.4	0.6	-37.5	0.4	-18.0
	Kimono1	24	-0.1	-26.5	0.5	-31.0	0.1	-14.3
	ParkScene	24	0.1	-42.3	0.6	-38.7	0.1	-17.2
Class C (WVGA)	RaceHorses	30	0.1	-15.8	0.7	-25.8	0.4	-15.2
	BasketballDrill	50	-0.2	-32.1	0.3	-34.4	0.1	-18.4
	BQMall	60	0.1	-40.2	0.7	-38.0	0.2	-19.4
	PartyScene	50	0.1	-28.1	0.3	-30.7	0.3	-17.5
Class D (WQVGA)	RaceHorses	30	0.0	-14.9	0.9	-22.9	0.2	-18.3
	BasketballPass	50	0.1	-40.5	0.1	-35.5	-0.1	-16.6
	BlowingBubbles	50	0.2	-28.7	0.5	-29.9	0.2	-18.3
	BQSquare	60	0.0	-42.6	0.1	-37.8	0.0	-20.2
Class F	SlideShow	30	0.6	-48.1	1.0	-44.2	-0.3	-15.3
	SlideEditing	20	2.7	-67.1	-0.5	-51.7	-0.4	-18.3
	ChinaSpeed	30	0.1	-25.8	0.6	-30.9	0.3	-14.2
	BasketballDrillText	50	-0.2	-32.5	0.1	-34.3	0.0	-17.8
Average			0.21	-35.03	0.45	-34.89	0.16	-16.74

TABLE XI

RESULTS OF THE PROPOSED ES-MD, PSC-MD, AND RDC-MD COMPARED WITH HM ENCODERS UNDER LOW-DELAY CONDITION

	Sequences	FPS	ES-MD		PSC-MD		RDC-MD	
			BDBR (%)	DT (%)	BDBR (%)	DT (%)	BDBR (%)	DT (%)
Class B (1080p)	BasketballDrive	50	0.2	-22.3	0.4	-28.5	0.0	-11.1
	BQTerrace	60	0.3	-44.6	0.3	-37.6	0.2	-13.7
	Cactus	50	0.1	-35.5	0.3	-36.5	0.3	-17.9
	Kimono1	24	0.2	-22.3	0.5	-31.4	0.2	-15.8
	ParkScene	24	0.5	-38.6	0.6	-36.8	0.3	-17.9
Class C (WVGA)	RaceHorses	30	0.2	-13.5	0.8	-27.3	0.3	-15.0
	BasketballDrill	50	0.2	-27.5	0.3	-32.6	0.2	-16.8
	BQMall	60	0.4	-37.3	0.8	-36.5	0.3	-17.4
	PartyScene	50	0.4	-25.6	0.6	-34.1	0.6	-22.1
Class D (WQVGA)	RaceHorses	30	0.3	-10.9	0.9	-23.9	0.5	-15.0
	BasketballPass	50	0.8	-43.3	0.6	-39.0	0.4	-23.6
	BlowingBubbles	50	0.2	-23.1	1.0	-27.0	0.3	-17.2
	BQSquare	60	0.1	-30.4	0.6	-30.5	0.1	-17.6
Class E (720p)	FourPeople	60	0.4	-65.1	0.5	-50.0	0.1	-21.0
	Johnny	60	0.1	-63.0	0.5	-49.6	0.1	-22.8
	KristenAndSara	60	0.5	-61.2	0.7	-48.1	-0.2	-21.8
Class F	SlideShow	30	0.2	-44.0	0.2	-40.3	-0.3	-14.1
	SlideEditing	20	-1.1	-73.6	-0.8	-52.2	0.2	-18.3
	ChinaSpeed	30	0.0	-23.9	0.8	-30.8	0.1	-14.8
	BasketballDrillText	50	0.3	-28.3	0.6	-33.6	0.4	-17.2
Average			0.21	-36.70	0.52	-36.31	0.21	-17.55

clock () to measure the run time. Simulations were run on an Intel 2.40 GHz Core2 Duo processor with 4-GB random access memory. The operating system was Windows XP SP2.

B. Results of the Individual Methods Compared With HM Encoders

The individual performance evaluation results of the proposed strategies, i.e., ES-MD, PSC-MD, RDC-MD under

the low-delay and random-access configurations are shown in Tables X and XI. It can be observed that three proposed strategies can greatly reduce the coding time with similar coding efficiency for all sequences. As far as the proposed ES-MD strategy concerned, about 35% and 37% coding time has been reduced in two configurations with the maximum of 74% and the minimum of 11%. Meanwhile, the coding efficiency almost has no loss in terms of BDBR

TABLE XII
RESULTS OF PROPOSED OVERALL ALGORITHM TO HM ENCODERS

		Random-Access			Low-Delay	
Sequences		FPS	BDBR (%)	DT (%)	BDBR (%)	DT (%)
Class A (2560×1600)	Traffic	30	0.9	-59.3	--	--
	PeopleOnStreet	30	1.1	-38.3	--	--
Class B (1080p)	BasketballDrive	50	0.6	-48.7	1.0	-47.8
	BQTerrace	60	0.7	-58.8	0.9	-57.6
	Cactus	50	0.9	-54.7	0.0	-52.7
	Kimono1	24	0.7	-42.8	0.8	-42.7
	ParkScene	24	1.1	-56.4	1.2	-54.9
Class C (WVGA)	RaceHorses	30	1.0	-34.6	1.0	-36.6
	BasketballDrill	50	0.1	-46.3	0.7	-45.2
	BQMall	60	0.9	-52.2	1.2	-52.2
	PartyScene	50	0.8	-42.7	1.0	-45.6
Class D (WQVGA)	RaceHorses	30	0.9	-31.5	1.6	-32.9
	BasketballPass	50	0.4	-51.4	1.3	-49.6
	BlowingBubbles	50	0.6	-44.0	1.8	-39.9
	BQSquare	60	0.5	-53.5	1.0	-45.8
Class E (720p)	FourPeople	60	--	--	1.0	-72.3
	Johnny	60	--	--	1.0	-71.5
	KristenAndSara	60	--	--	1.4	-69.3
Class F	SlideShow	30	0.9	-57.8	-0.5	-54.5
	SlideEditing	20	-0.9	-73.1	-0.8	-76.6
	ChinaSpeed	30	1.0	-40.6	1.2	-41.6
	BasketballDrillText	50	0.4	-46.6	0.8	-45.9
Average			0.68	-49.12	0.88	-51.80

TABLE XIII
RESULTS OF PROPOSED OVERALL ALGORITHM FOR SCENES CUT SEQUENCES

		Random-Access			Low-Delay	
Sequences	Picture size	FPS	BDBR (%)	DT (%)	BDBR (%)	DT (%)
BasketballPass_BlowingBubbles	416×240	25	0.8	-34.1	1.0	-42.0
		50	0.4	-44.6	1.2	-42.7
PartyScene_BasketballDrill	832×480	25	0.4	-34.7	0.4	-41.4
		50	0.5	-43.1	0.7	-42.9
Kimonol_ParkScene	1920×1080	12	0.6	-38.1	0.6	-45.6
		24	0.5	-47.7	0.5	-46.4
Cactus_BasketballDrive	1920×1080	12	0.5	-42.4	0.6	-47.1
		24	0.6	-51.4	0.6	-50.4
PeopleOnStreet_Traffic	2560×1600	15	0.8	-42.1	0.6	-39.5
		30	0.9	-44.2	1.0	-45.7
Average			0.60	-42.24	0.71	-44.38

Note: low frame rate sequences are achieved by temporal downsampling of the original ones

(0.21% increment). The above result analysis indicates that the proposed ES-MD strategy can efficiently reduce the encoding time while keeping the same RD performance as the original encoder. As shown in Table VIII, the proposed PSC-MD strategy can reduce coding time by 35%–36% on average. This coding time reduction is particularly high for low-activity sequences such as *Traffic*, *BQTerrace*, and *SlideEditing*, but is still evident for high-activity sequences such as *PeopleOnStreet* and *RaceHorses*. On the other side, the loss of coding efficiency is negligible (with 0.44%–0.52% BDBR increase). This result indicates that the proposed PSC-MD approach can efficiently skip unnecessary modes in mode decision. For the proposed adaptive RDC-MD approach, 17%–18% coding time has been reduced in Tables X and XI. It also shows a consistent gain

in coding speed for different activity sequences with the least gain of 11% in *BasketballDrive* and the most gain of 24% in *BasketballPass*. The average increase of BDBR is 0.18%, which is negligible. Therefore, the proposed RDC-MD strategy keeps the RD performance of the original HM encoder, while reducing the computational complexity of the coding process considerably.

C. Results of the Overall Algorithm Compared With HM Encoders

In the following, we analyze the experimental results of the proposed overall algorithm, which incorporates ES-MD, PSC-MD, and RDC-MD. Table XII shows the performance evaluation results of the proposed overall algorithm. The proposed algorithm can greatly reduce the encoding time for

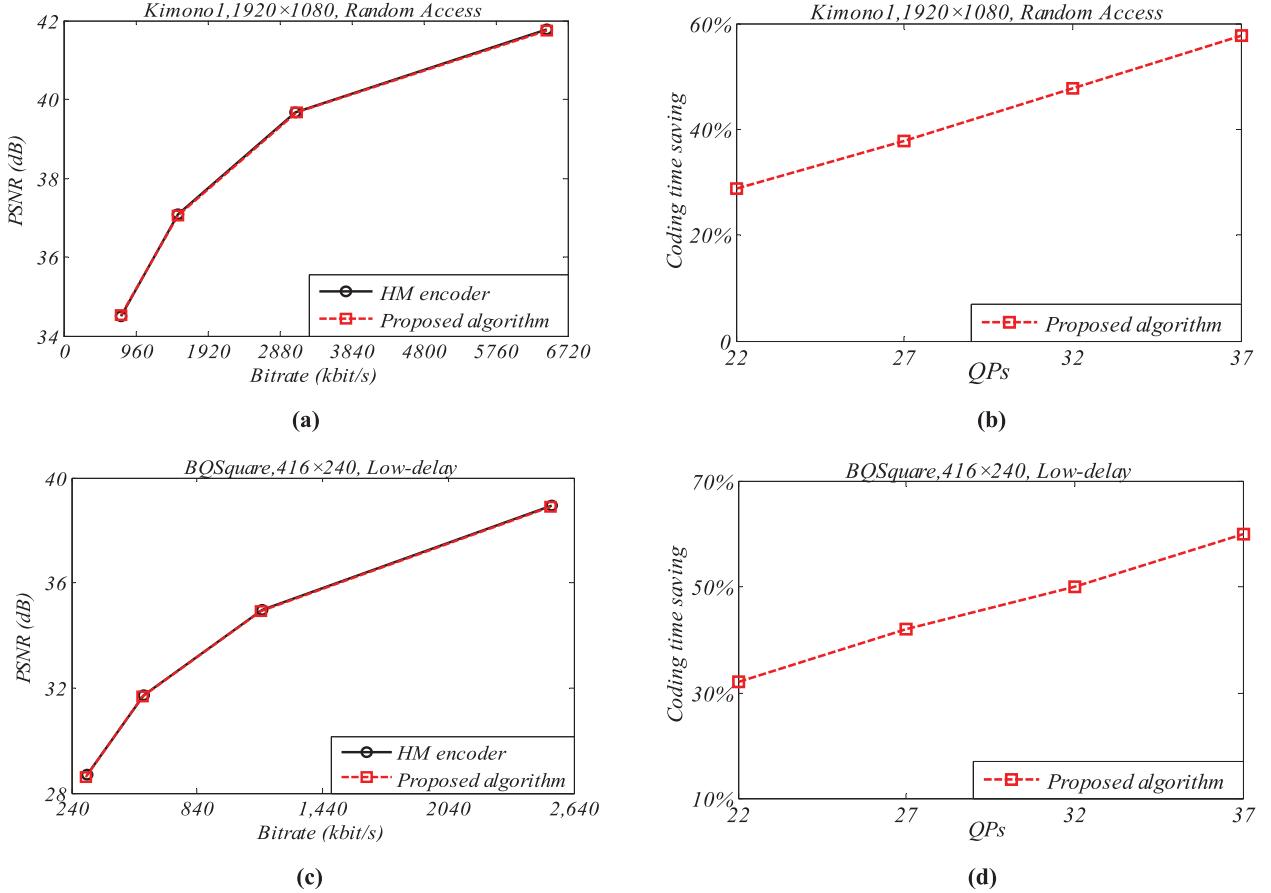


Fig. 5. Experimental result of *Kimono1* (1920×1080 , 24Hz) and *BQSquare* (416×240 , 60 Hz) under different QPs (22, 27, 32, and 37). (a) RD curves of *Kimono1*. (b) Time saving curve of *Kimono1*. (c) RD curves of *BQSquare*. (d) Time saving curve of *BQSquare*.



Fig. 6. Comparisons of the subjective qualities of the 30th frame of *RaceHorses* sequence (416×240). (a) HM 10.0, low-delay (PSNR: 28.15 dB, coding time: 8 s). (b) Our algorithm (PSNR: 28.06 dB, coding time: 4 s, speed up of 2).

all sequences with similar coding efficiency compared as the original HM encoder. As shown in Table XII, the proposed algorithm can reduce coding time by 49% and 52% in random-access and low-delay configurations, respectively, with the maximum of 77% in *SlideEditing* and the minimum of 32% in *RaceHorses* (416×240). For slow motion sequences like *SlideEditing*, *Traffic*, and *BQTerrace*, the proposed algorithm saves more than 58% encoding time. The computation reduction is particularly high because the exhaustive mode decision procedures of a significant number of CUs are not processed by the encoder. For high activity sequences like *RaceHorses* and *PeopleOnStreet*, the proposed algorithm also can reduce about 32%–38% coding time. Meanwhile, the coding efficiency loss is negligible in Table XII, where the average BDBR increment is 0.68%–0.88%. Therefore, the proposed algorithm



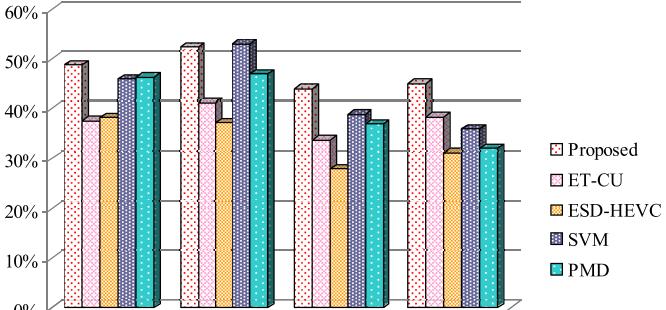
Fig. 7. Comparisons of the subjective qualities of the 30th frame of *BQMall* sequence (832×480). (a) HM10.0, low-delay (PSNR: 33.04 dB, coding time: 27 s). (b) Our algorithm (PSNR: 33.02 dB, coding time: 13 s, speed up of 2.1).



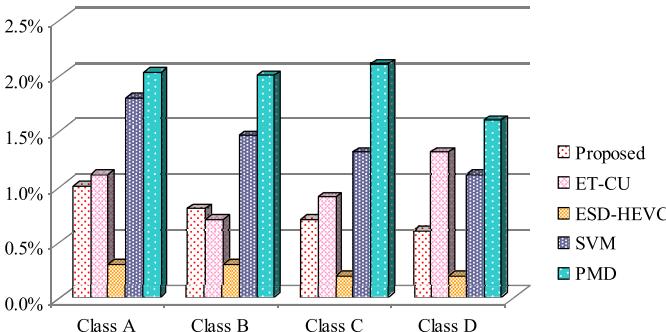
Fig. 8. Comparisons of the subjective qualities of the 30th frame of *Cactus* sequence (1920×1080). (a) HM10.0, random-access (PSNR: 36.89 dB, coding time: 108 s). (b) Our algorithm (PSNR: 36.87 dB, coding time: 54 s, speed up of 2).

can efficiently reduce the encoding time while keeping nearly the same RD performance as the original HM encoder.

Fig. 5 shows the experimental results of the overall algorithm under different QPs (22, 27, 32, and 37) for two



(a)



(b)

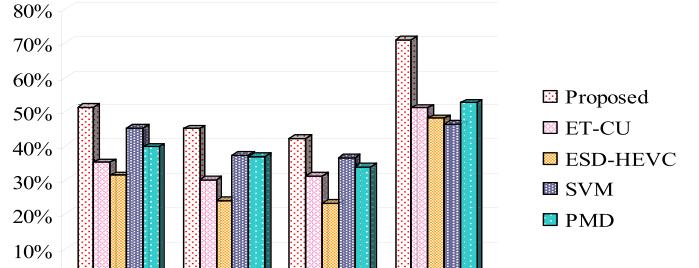
Fig. 9. Comparison results of proposed algorithm and previous works in the random-access configuration. (a) Coding time saving. (b) BDBR increase.

typical sequences *Kimono1* and *BQSquare*. We can see from Fig. 5 that the proposed algorithm can achieve the consistent time saving over a large bitrate range, with almost no PSNR loss and no bitrate increments.

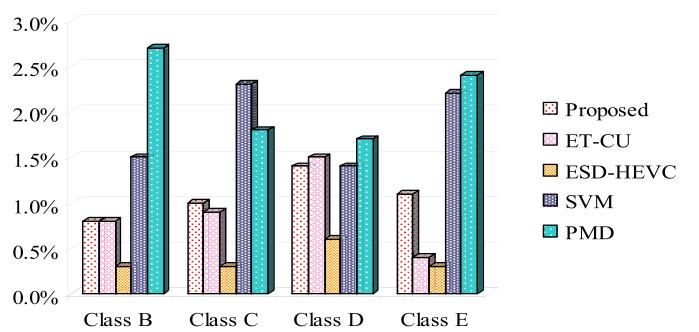
We also show performance of the proposed algorithm using scene-cut sequences with different frame rates in Table XIII. Scene-cut sequences in Table XIII are generated by cascading two sequences for scene change. They contain multiple scenes and every 10 frames has a scene cut frame. We can see that the proposed algorithm can reduce 42% and 44% coding time for scene-cut sequences under random-access and low-delay configurations, respectively. The coding efficiency loss is negligible with 0.6% and 0.7% BDBR increase. It also can be seen that the higher frame rate sequences achieve a larger coding time reduction since temporal correlation between the frames is high.

D. Subjective Results of the Overall Algorithm Compared With HM Encoders

Figs. 6–8 show the 30th frame of *RaceHorses*, *BQMall*, and *Cactus* video sequences to compare subjective quality of the HM 10.0 encoder and the proposed algorithm. It can be observed that the tail part of the horse marked by the red circle in Fig. 6(b) is more blurred than that in Fig. 6(a). It is the spot where the image quality of the proposed algorithm is obviously lower than the HM 10.0 encoder. In terms of the objective quality, the PSNR of the proposed algorithm is



(a)



(b)

Fig. 10. Comparison results of proposed algorithm and previous works in the low-delay configuration. (a) Coding time saving. (b) BDBR increase.

slightly lower than that of the HM 10.0 encoder; however, the proposed algorithm reduces the coding time by 50%. It can be observed from Figs. 7 and 8 that there is no significant flaw of image quality for the proposed algorithm from the perspective of subjective quality. The proposed scheme and the HM encoder have similar subjective qualities, but the coding speed of the proposed algorithm is increased significantly.

E. Performance Comparison With the State-of-the-Art Fast Algorithms

In addition to the HM encoder, we use four methods for an objective comparison of the coding performance in Figs. 9 and 10. These are CU splitting ET based on SVM (SVM) [15], CU selection method based on PMD [18], ET of CU encoding (ET-CU) [31] and early SKIP detection for HEVC (ESD-HEVC) [32], which are well-known efficient and fast algorithms for HEVC. ESD-HEVC performs good RD performance in Figs. 9 and 10, but its computation reduction is poor. ET-CU, SVM, and PMD have good computation reduction at the cost of RD performance. Among these four previous works, SVM achieves the largest coding time saving, and PMD yields large coding efficiency degradation in terms of BDBR. For five classes (A, B, C, D, and E) of test sequences, our algorithm achieves 40%–70% total encoding time reduction as well as appropriate RD performance with only 0.5%–1.4% BDBR increase on average. Compared with ESD-HEVC, the proposed algorithm can achieve better

performance on time saving. About 15%–23% encoding time can be further reduced. Meanwhile, the average coding efficiency loss is negligible, less than 0.5% BDBR increase. The proposed algorithm achieves about 6%–23% coding time saving in low-delay configuration (2%–13% in random-access configuration) compared with ET-CU, SVM, and PMD with a better RD performance. Moreover, the proposed algorithm is particularly prominent for the video sequences in Class E with more than 70% coding time saving while other four methods only obtain 46%–52% time saving. The above experimental results indicate the proposed scheme based on inter-level, spatial, and temporal analysis is efficient for all video sequences and consistently outperforms the recent approaches for HEVC.

V. CONCLUSION

In this paper, we propose an adaptive inter-mode decision algorithm to reduce the computational complexity of the HEVC encoder, which includes three strategies, i.e., ES-MD, PSC-MD, and RDC-MD. The results of comparative experiments demonstrate that the proposed algorithm can effectively reduces the computational load by 49%–52% on average as compared with the HM 10.0 encoder, while only incurring a negligible loss of coding efficiency (0.68%–0.88% BDBR increase). Furthermore, it consistently outperforms well-known efficient and fast algorithms for HEVC such as SVM [15], PMD [18], ET-CU [31], and ESD-HEVC) [32], with about an additional 2%–23% encoding time saving.

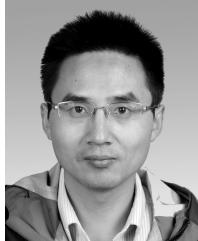
ACKNOWLEDGMENT

The authors would like to thank the associate editor and reviewers for their valuable comments and Dr. C. Zhou for helping to check the writing in English.

REFERENCES

- [1] G. J. Sullivan and J. R. Ohm, "Recent developments in standardization of high efficiency video coding (HEVC)," *Proc. SPIE*, vol. 7798, pp. 7798–7730, Aug. 2010.
- [2] G. J. Sullivan and T. Wiegand, "Video compression—from concepts to the H.264/AVC standard," *Proc. IEEE*, vol. 93, no. 1, pp. 18–31, Jan. 2005.
- [3] M. Karczewicz *et al.*, "A hybrid video coder based on extended macroblock sizes, improved interpolation, and flexible motion representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1698–1708, Dec. 2010.
- [4] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1685–1696, Dec. 2012.
- [5] *High Efficiency Video Coding (HEVC) Test Model 10 (HM 10) Encoder Description*, ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document JCTVC-L1002, Geneva, Switzerland, Jan. 2013.
- [6] Y. Lee and Y. Lin, "Zero-block mode decision algorithm for H.264/AVC," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 524–533, Mar. 2009.
- [7] H. Zeng, K. Ma, and C. Cai, "Fast mode decision for multiview video coding using mode correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1659–1666, Nov. 2011.
- [8] H. Wang, S. Kwong, and C. Kok, "An efficient mode decision algorithm for H.264/AVC encoding optimization," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 882–888, Jun. 2007.
- [9] A. C. W. Yu, G. R. Martin, and H. Park, "Fast inter-mode selection in the H.264/AVC standard using a hierarchical decision process," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 186–195, Feb. 2008.
- [10] D. Wu *et al.*, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 953–958, Jul. 2005.
- [11] H. Zeng, K. Ma, and C. Cai, "Hierarchical intra mode decision for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 907–912, Jun. 2010.
- [12] Y. Sun and J. 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.
- [13] L. Shen, Z. Liu, Z. Zhang, and X. Shi, "Fast inter mode decision using spatial property of motion field," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1208–1214, Oct. 2008.
- [14] L. Shen, Y. Sun, Z. Liu, and Z. Zhang, "Efficient SKIP mode detection for coarse grain quality scalable video coding," *IEEE Signal Process. Lett.*, vol. 17, no. 10, pp. 887–890, Oct. 2010.
- [15] X. Shen and L. Yu, "CU splitting early termination based on weighted SVM," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, pp. 1–11, 2013.
- [16] X. Shen, L. Yu, and J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in *Proc. PCS*, May 2012, pp. 453–456.
- [17] S. Cho and M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1555–1564, Sep. 2013.
- [18] J. Xiong, H. Li, Q. Wu, and F. Meng, "A fast HEVC inter CU selection method based on pyramid motion divergence," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 559–564, Feb. 2014.
- [19] H. Sun, D. Zhou, and S. Goto, "A low-complexity HEVC intra prediction algorithm based on level and mode filtering," in *Proc. IEEE ICME*, Jul. 2012, pp. 1085–1090.
- [20] G. Correa, P. Assuncao, L. Agostini, and L. A. da Silva Cruz, "Complexity control of HEVC through quadtree depth estimation," in *Proc. IEEE EUROCON*, Jul. 2013, pp. 81–86.
- [21] C. Zhou, F. Zhou, and Y. Chen, "Spatio-temporal correlation-based fast coding unit depth decision for high efficiency video coding," *J. Electron. Imag.*, vol. 22, no. 4, p. 043001, Oct. 2013.
- [22] Y. Zhang, H. Wang, and Z. Li, "Fast coding unit depth decision algorithm for interframe coding in HEVC," in *Proc. DCC*, Mar. 2013, pp. 53–62.
- [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] L. Shen, Z. Zhang, and P. An, "Fast CU size decision and mode decision algorithm for HEVC intra coding," *IEEE Trans. Consum. Electron.*, vol. 59, no. 1, pp. 207–213, Feb. 2013.
- [25] X. Li, J. An, X. Guo, and S. Lei, *Adaptive CU Depth Range, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, document JCTVC-E090, Geneva, Switzerland, Mar. 2011.
- [26] K. Choi, S. H. Park, and E. S. Jang, *Coding Tree Pruning Based CU Early Termination, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, document JCTVC-F092, Turin, Italy, Jul. 2011.
- [27] 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.
- [28] G. Chen, Z. Liu, T. Ikenaga, and D. Wang, "Fast HEVC intra mode decision using matching edge detector and kernel density estimation alike histogram generation," in *Proc. IEEE ISCAS*, May 2013, pp. 53–56.
- [29] Q. Yu, Y. Rong, and Y. He, "Fast intra mode decision strategy for HEVC," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 500–504.
- [30] K. Lee, H. Lee, J. Kim, and Y. Choi, "A novel algorithm for zero block detection in high efficiency video coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1124–1134, Dec. 2013.
- [31] R. Gwon and Y. Lee, *Early Termination of CU Encoding to Reduce HEVC Complexity, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, document JCTVC-F045, Turin, Italy, Jul. 2011.
- [32] J. Yang, J. Kim, K. Won, H. Lee, and B. Jeon, *Early SKIP Detection for HEVC, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, document JCTVC-G543, Geneva, Switzerland, Nov. 2011.
- [33] J. Kim, J. Yang, K. Won, and B. Jeon, "Early determination of mode decision for HEVC," in *Proc. PCS*, May 2012, pp. 449–452.
- [34] J. Kim, S. Jeong, S. Cho, and J. Choi, "Adaptive coding unit early termination algorithm for HEVC," in *Proc. IEEE ICCE*, Jan. 2012, pp. 261–262.
- [35] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 Part-10 AVC/H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.

- [36] S. Kim, K. Konda, C. Park, C. Cho, and S. Ko, "Fast mode decision algorithm for inter-layer coding in scalable video coding," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1572–1580, Aug. 2009.
- [37] L. F. Chen, S. P. Yang, and Y. K. Lai, "Model-based early termination scheme for H.264/AVC inter-prediction," in *Proc. ICASSP*, Apr. 2009, pp. 597–600.
- [38] L. Shen, Z. Liu, P. An, R. Ma, and Z. Zhang, "Low-complexity mode decision for MVC," *IEEE Trans. Circuits System Video Technol.*, vol. 21, no. 6, pp. 837–843, Jun. 2011.
- [39] G. Bjontegaard, "Calculation of average PSNR difference between RD-curves," presented at the 13th VCEG-M33 Meeting, Austin, TX, USA, Apr. 2001.
- [40] F. Bossen, *Common Test Conditions and Software Reference Configurations, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, document JCTVC-L 1100, Geneva, Switzerland, Jan. 2013.



Liqian Shen received the B.S. degree in automation control from Henan Polytechnic University, Shanyang, China, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2001, 2005, and 2008, respectively. He was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, as a Visiting Professor from 2013 to 2014. He has been with the Faculty of the School of Communication and Information Engineering, Shanghai University, since 2008, where he is currently an Associate Professor. He has authored and co-authored more than 80 refereed technical papers in international journals and conferences in the field of video coding and image processing. He holds 10 patents in the areas of image/video coding and communications. His research interests include High Efficiency Video Coding, perceptual coding, video codec optimization, 3-DTV, and video quality assessment.



Zhaoyang Zhang received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1962. He was the Director with the Key Laboratory of Advanced Display and System Application, Ministry of Education, Shanghai University, Shanghai, China, where he is currently a Distinguished Professor with the School of Communication and Information Engineering. He has published more than 200 refereed technical papers and 10 books. In addition, he holds 20 patents in the areas of image/video coding and communications. His research interests include digital television, 2-D and 3-D video processing, image processing, and multimedia systems.



Zhi Liu (M'07) received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, and the Ph.D. degree from Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China, in 1999, 2002, and 2005, respectively. He was a Visiting Researcher with the SIROCCO Team, IRISA/INRIA-Rennes, Rennes, France, since 2012. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai. His research interests include saliency model, image/video segmentation, image/video retargeting, video coding, and multimedia communication. He has published more than 100 refereed technical papers in international journals and conferences. Dr. Liu was a TPC Member with ICME 2014, WIAMIS 2013, IWVP 2011, PCM 2010, and ISPACS 2010. He co-organized special sessions on visual attention, saliency models, and applications at WIAMIS 2013 and ICME 2014.