Inter Prediction Complexity Reduction for HEVC based on Residuals Characteristics

Kanayah Saurty
Faculty of Information and
Communication Technologies
Université des Mascareignes
Pamplemousses, Mauritius

Pierre C. Catherine
School of Innovative
Technologies and Engineering
University of Technology,
La Tour Koenig, Mauritius

Krishnaraj M. S. Soyjaudah
Faculty of Engineering
University of Mauritius
Réduit, Mauritius

Abstract—**High Efficiency Video Coding (HEVC) or H.265 is currently the latest standard in video coding. While this new standard promises improved performance over the previous H.264/AVC standard, the complexity has drastically increased due to the various new improved tools added. The splitting of the 64×64 Largest Coding Unit (LCU) into smaller CU sizes forming a quad tree structure involves a significant number of operations and comparisons which imposes a high computational burden on the encoder. In addition, the improved Motion Estimation (ME) techniques used in HEVC inter prediction in order to ensure greater compression also contribute to the high encoding time. In this paper, a set of standard thresholds are identified based on the Mean Square (MS) of the residuals. These thresholds are used to terminate the CU splitting process and to skip some of the inter modes processing. In addition, CUs with large MS values are split at a very early stage. Experimental results show that the proposed method can effectively reduce the encoding time by 62.2% (70.8% for ME) on average, compared to HM 10, yielding a BD-Rate of only 1.14%.**

Keywords—HEVC; inter prediction; early termination scheme; complexity reduction; prediction residuals

Fig. 1. HEVC quad tree splitting of the LCU

I. INTRODUCTION

High Efficiency Video Coding (HEVC), produced by the Joint Collaborative Team on Video Coding (JCT-VC) consisting of ISO-IEC/MPEG and ITU-T/VCEG, is currently the latest standard in video image compression [32]. In fact, it makes use of the same hybrid approach, consisting of the intra/inter prediction coupled with 2-D transform coding, which all video compression technology have been using since H.261.

Video coding consists essentially of removing the maximum redundancy possible through intra and inter prediction. While intra prediction, which makes use of data from the same frame, produces a very high quality output, it is also accompanied by a high bitrate. Inter prediction exploits the redundancies among already decoded frames and contributes significantly towards the low bitrate of the compression algorithm.

Unlike its predecessor, H.264/AVC [35], which partitions the frame into 16×16 macroblocks, HEVC makes use of a block of size 64×64 pixels which is also known as the Largest Coding Unit (LCU). This large LCU size is advantageous for smooth regions of a picture although HEVC has to test all the various combinations of CU sizes. In fact, HEVC adopts a highly flexible quad-tree structure as shown in Fig. 1.
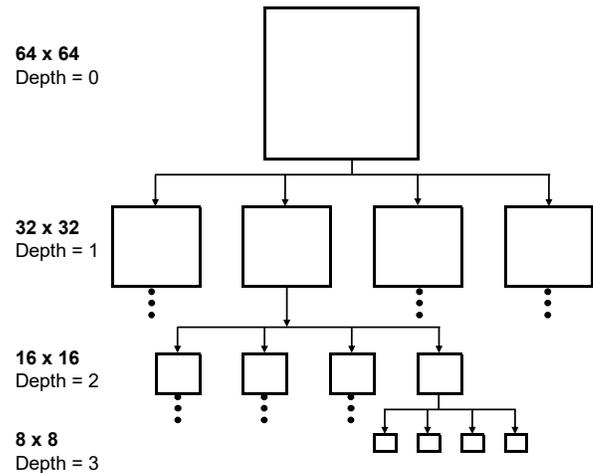
For each CU size, HEVC selects one out of many prediction possibilities resulting in the testing of various Prediction Units (PUs) and Transform Units (TUs) prior to selecting the best one based on the Rate Distortion (RD) cost. In addition, using a bottom up approach in the quad-tree structure, the four aggregated children CUs are compared with their parent to check the possibility of combining them as one CU only. The larger the CU size, the higher will be the compression as the prediction information sent will be less than those required by the four children. This process applies to both the intra and the inter prediction, although the PU partitioning is different for these two modes.

A high proportion of the encoding time is however devoted to inter-prediction during the encoding phase. Inter frame prediction alone consumes from 60% to 96% of the total encoding time [26]. In [13], inter prediction is shown to consume 84% of the encoding time while only 3% is taken by intra prediction for High Definition (HD) sequences.

Consequently, for practical applications such as high-resolution video services and real-time processing, HEVC still requires a significant complexity reduction while maintaining the high coding performance. Several approaches have been proposed to reduce the encoding complexity for both intra-picture and inter-picture prediction. In intra-prediction, com-

plexity reduction is achieved by reducing the number of modes used out of the maximum of 35 and by terminating the CU splitting process as early as possible [5], [6], [9], [22], [27], [28], [31], [40]. While the determination of the CU size at an early stage is also applicable to inter prediction, time reduction is also obtained by optimizing on the Motion Estimation (ME) process.

The complex quad-tree structure of the HEVC standard, which is the basis of the high encoding time, also implies that various optimization techniques could be implemented. Thus, most of the existing works have come up with heuristics to eliminate the processing of some of the possible configurations within the full range available so as to encode the input video with substantial time savings. The results of these techniques to accelerate the encoding process also come along with varied loss in quality and increase in the bit rate.

The high encoding time during inter prediction is largely contributed by the use of various Motion Estimation (ME) and motion compensation techniques to produce accurate prediction. In addition, the use of multiple reference frames in the ME techniques since the H.264/AVC standard, makes the motion estimation process even more complex. The common optimization method to reduce the encoder complexity therefore consists of decreasing the number of ME operations during the inter prediction process.

In this paper, the splitting of CUs in the quad tree is terminated based on the Mean Square (MS) of the residuals. In addition, CUs are also identified as split-CUs during the $2N \times 2N$ PU processing itself. In such cases, the processing for other PU modes are skipped. The different thresholds are identified after analyzing the splitting process of a sample of the sequences. The rest of this paper is thus organized as follows. Section II enumerates the techniques proposed by related works. Section III provides an overview of inter prediction in HEVC. The time consuming elements during inter-prediction are presented in Section IV. Section V describes the proposed approach consisting of the early termination schemes. The experiments conducted and discussions on the results obtained are subsequently presented in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORKS

Several algorithms have been proposed to reduce the HEVC complexity while ensuring negligible loss in quality and bit rate.

For example, significant time savings are obtained by avoiding the unnecessary splitting of some CUs. In [41], the quad-tree CU depth decision process is modeled as a three-level of hierarchical binary decision problem. In addition, a flexible CU depth decision structure is used to allow the performance of each CU depth decision be smoothly transferred between the coding complexity and RD performance along with binary classifiers to control the risk of false prediction. By making use of the mode information of the current CU and a depth range selection mechanism (DRSM), [19] produces an effective splitting decision process. The RD costs of the parent and current levels are used in [11] to terminate the quad-tree-based structure earlier, thus reducing the computational complexity of the encoder. In [38], the CU splitting decision

is constructed on a pyramid motion divergence (PMD) based CU selection along with a k nearest neighboring like method. Splitting and termination decision approaches are also explored in [15], [30]. In [30], a Bayesian decision rule is also used in the splitting decision. The splitting decision in [29] is based on skipping some specific depth levels rarely used in the previous frame and neighboring CUs in addition to using termination methods based on motion homogeneity checking, RD cost checking and SKIP mode checking. An early pruning method based on statistics of the prediction residuals is used in [33] to produce significant time gain. A fast CU decision approach is proposed in [36] based on an exponential model expressing the relationship between the motion compensation R-D cost and the SAD cost for the upper CU and its sub-CUs.

Fast prediction unit decision methods are proposed in [17], [18], [24], [34]. Motion Estimation is performed only on the selected inter prediction mode in [17] based on a priority level. Inter prediction information from previously predicted blocks and neighboring blocks enable the selection of only one inter prediction mode. Spatio-temporal analysis and depth correlations along with a classification of motion activity are explored in [18]. In [34], optimization techniques are proposed based on the rate-distortion-complexity characteristics of the HEVC inter prediction for the different block partitioning structures. In [24], the depth information of a CTU is already determined from information collected in the $2N \times 2N$ PU. It makes use of the high probability of $2N \times 2N$ being the best mode. Thus a fast scheme is proposed to make this decision at an early stage itself. In addition, a merge SKIP extraction method is developed and integrated with the CU depth decision algorithm to effectively decrease the encoding time.

A simple tree-pruning algorithm is proposed in [7] that exploits the observation where the sub-tree computations can be skipped if the coding mode of the current node is sufficient, i.e. a SKIP mode. An early detection of SKIP mode is also proposed in [39] to reduce an encoding complexity of HEVC. The proposed method is similar to the early skip detection scheme implemented the H.264/AVC, but slightly modified to address the different encoding scheme of HEVC. The SKIP mode and the number of transformed coefficients of a CU which are already computed values in HEVC are utilized in [10] and [16] to produce fast mode decision algorithm. A fast rate-distortion estimation algorithm for HEVC is proposed in [10] based on genuine zero blocks (GZBs) and pseudo zero blocks (PZBs) of the transformed coefficients. In [16], the early mode decision is modeled as a binary classification problem of SKIP/non-SKIP or split/unsplit along with the Neyman-Pearson-based rule to balance the rate-distortion (RD) performance loss and the complexity reduction.

In [25], a new global search pattern for finding the global minima and an adaptive early termination condition are proposed to speed up the Motion Estimation (ME) algorithm. By merging $N \times N$ PU partitions in order to compose larger ones, several Motion Estimation (ME) calls during the PU inter-prediction decision are avoided in [26] to reduce the overall encoding process.

Reference [13] proposes a system-level Adaptive Workload Management Scheme (AWMS). The AWMS collect feedbacks at a frame level and dynamically configures different parameters for the video-coding system such as the maximum CU

depth and the search range. A fast CU size decision based on Sobel operator is proposed in [14]. By using textual features of the video images, the algorithm identifies the coding depth of the final CU without having to traverse the coding layers and results in a reduced computational complexity.

III. OVERVIEW OF INTER PREDICTION IN HEVC

Each frame in HEVC is initially partitioned into blocks of 64×64 pixels prior to any prediction and encoding being made. Each of these blocks which is also known as the Largest Coding Block (LCU) is further split recursively into four children until the Smallest Coding Units (SCUs) of size 8×8 are reached. The quadtree structure formed is illustrated in Fig. 1. Intra prediction iterates among the 35 available modes to select the one that results in the cheapest Rate-Distortion (RD) cost. The reference data which are obtained from the already decoded top and left CUs are interpolated to form the prediction for the current CU regarding the angular modes. While intra prediction makes use of reference samples within the same frame, the references of inter prediction comes from the already decoded frames. As such, inter-prediction makes use of a list of reference frames in the case of uni-predictive frame (P-frame) and two lists (list 0 and list 1) for the bi-predictive frame (B-Frame).

Inter prediction exploits the motion data among frames. A block of pixels in a frame generally has a very close match with another block of the same size in a different frame. This match may be at the same location in the reference frame when there is no motion but at a different location when motion is present. HEVC represents this prediction as a set of motion vector which gives the translational movement of the block from the reference frame to the current one. The residuals which is the difference between the original block of pixels and the prediction block is transformed and together with the motion vector constitute the motion data.

During the inter-prediction process, a single motion data may not optimally represents the prediction of the CU in terms of RD cost as objects within this CU may have different motion vectors. CUs are therefore split into PUs to more accurately reflect the different motion vectors within the single block of pixels. In order to determine the best PU configuration, an intensive Motion Estimation (ME) process is undertaken. HEVC has to theoretically search through all the possible blocks in the search window which is commonly known as the full search algorithm so as to reach the best match in terms of minimum distortion and low number of bits to represent the encoded region. There are many fast ME algorithms which optimize this search. Efficient search techniques [20], [25] have been implemented as this process is the most time consuming element in the inter-picture process.

There are 8 possible PU configurations for each CU and they consist of one, two or four PUs. The different PU modes in inter prediction are illustrated in Fig. 2. The PUs are categorized as either symmetric PUs or asymmetric ones. For the symmetric PUs, the CU is either not split or split into 4 PUs or it may be split into two identical PUs (horizontal or vertical). The asymmetrical PUs are also known as Asymmetrical Motion Prediction (AMP) and consist of 4 PU configurations: the CU is split into $1/4$ and $3/4$ of the square region. This

is especially useful when only a small part within the region shows a different motion vector. There are fours AMP PU modes namely $2N \times nU$, $2N \times nD$, $nL \times 2N$ and $nR \times 2N$. The $N \times N$ PU configuration is only processed for the 8×8 CUs. It is to be noted that the smallest size CUs do not check for the AMP PU modes. They only adopt the $2N \times 2N$, $N \times 2N$, $2N \times N$ and the $N \times N$ PU configurations [4].
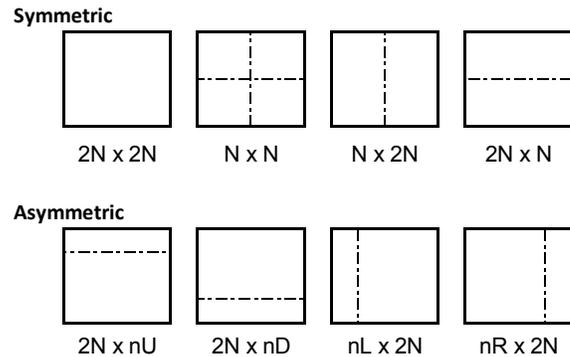

Fig. 2. HEVC inter-picture PU modes

Once the PU modes for each CU size has been identified in the quad tree, a bottom-up approach comparison is performed to determine whether the parent CU is more optimized compared to the combined RD-cost of the four children. In case the parent CU along with the appropriate PU mode is more optimized, it is retained. Otherwise, the 4 children, each with their independent PU configuration, are adopted.

IV. HEVC INTER PREDICTION COMPLEXITY

Although the new features introduced in the HEVC inter prediction contributes to the high performance, it is also the cause for the increase in encoding time compared to AVC. This complexity is distributed among many stages throughout the prediction process. The following sub-sections provide an insight into the time consuming elements during the inter prediction stage.

A. Inter Frames Encoding Time

Inter frame prediction is the most time consuming part in the HEVC encoder. The processing time of inter-predicted frames is in fact higher than their corresponding intra frames within the same sequence. A preliminary experiment is conducted by encoding 10 frames for each sequence and the result is illustrated in Fig. 3. It clearly shows that inter frames processing requires more than *thrice* the processing time of intra frames, indicating the complex nature of the inter prediction process. This higher computing time is associated with the exhaustive Motion Estimation (ME) processing which is performed for each possible PU configuration at the different depth of the quadtree structure.

B. Time spent at different depth of the LCU

The amount of time spent within each CU size or depth is further analyzed by collecting the data from the inter predicted frames for a number of sequences and the result is provided
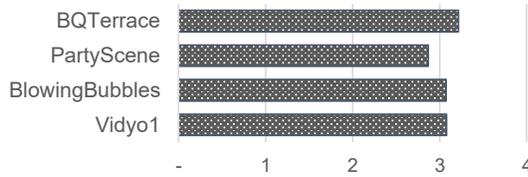
Fig. 3.   Inter frames processing time relative to corresponding intra frames

in Fig. 4. It is noticed that the 64×64 CU size shows the smallest percentage of time spent, i.e, 18%. This largest size CU seldom tests all the possible PU configurations and therefore the encoder spends a smaller proportion of time at this depth compared to others. Processing in the other CU depths averagely spend slightly more than 25% of the total inter prediction time spent for the whole LCU. Avoiding the processing of higher depths during inter prediction can therefore bring around 25% time savings for each depth at the level of the CU.
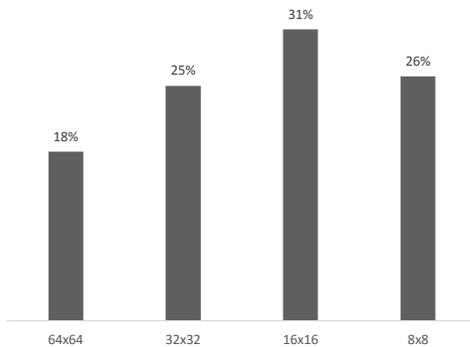

Fig. 4.   Time spent within each CU size during inter-prediction

As opposed to intra prediction where the complexity of the texture in the frame normally dictates the granularity of the CU structure, the movement of the objects in the frame relative to the background or to other objects determines the CU structure during inter prediction.

The ME operation precisely maps the region of the CU being processed to those within the searched window in order to determine the closest match in terms of the lowest distortion and bit rate. This intensive ME operation lends itself to a very time consuming process although some of these steps could be avoided by identifying the non-split CUs or specific PU modes at an early stage of the quad-tree traversal.

C. Frequency of PU modes

Data has also been collected from the inter predicted frames for a number of sequences and the frequency of occurrences of the different PU modes is provided in Fig. 5. Among the CUs which are not split, almost 90% of them adopt the $2N \times 2N$ PU mode.

By carefully identifying those terminating CUs that potentially will result in $2N \times 2N$ PU mode, the complexity of
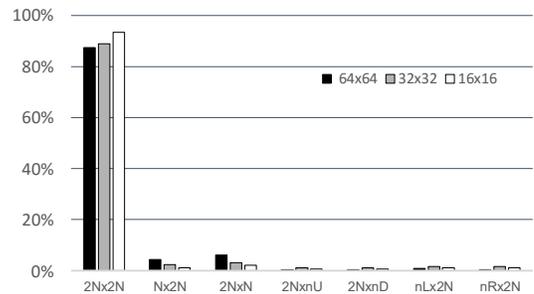

Fig. 5.   Occurrences of PU modes for different CU sizes

the inter prediction process may be further reduced. In the following section, a method of reducing these unnecessary computations is proposed by analyzing the residuals formed following the $2N \times 2N$ PU mode operation.

V.   PROPOSED APPROACH

Using the *lowdelay B main* profile for HEVC, there are three B-frames for every I-frame in a GOP. As illustrated in Fig. 3, inter-predicted frames encoding consumes more than *thrice* the amount of time compared to intra-predicted frames. The complexity of only the B-frames have therefore been reduced in this approach as they account for the larger portion of the encoding time. Moreover, since regions from the I-frames are the main references in the inter-prediction process and are used for assessing the quality of the predicted CUs, they are encoded using the conventional HEVC process.

The final structure of the LCU is determined after traversing the quadtree formed by HEVC with CUs of size 64×64 down to 8×8. The proposed approach of complexity reduction targets CUs of size 16×16 and larger only. By terminating CUs of size at least 16×16, only a reduced number of CUs of size 8×8 will be left. These remaining CUs follow the HEVC conventional processing.

By making an informed decision at the beginning of the CU processing for inter prediction, a number of unnecessary processing can be avoided leading a substantial decrease in the encoding time. For example, by correctly identifying that a CU will be split at the beginning itself can prevent the unnecessary checking of other modes. Next, determining that a CU will not be further divided can avoid the processing at higher depths. Within this very CU, the encoding time may be further reduced by recognizing that this terminating CU may be encoded as a single PU, thus avoiding the verification of other PU modes. These decisions, when applied collectively during the quad tree traversal, can lead to enormous time savings for the encoder.

The CUs under consideration are thus classified as

1)   Single PU mode ($2N \times 2N$)
2)   Other PU modes (two PUs), and
3)   Splitting of the CU into four children

In this paper, CUs are classified into the different categories based on the luma residuals. Once the best $2N \times 2N$ PU motion data is obtained following the merge and the SKIP operations,

the luma residuals which are also of the order of $2N \times 2N$ are further analyzed.

The residuals, which is the difference between the original block of pixels and the prediction block, obtained after performing the $2N \times 2N$ inter prediction processing are analyzed to determine the probable configuration of that CU. The residuals are grouped in blocks of 8×8 and the Mean Square (MS) of the residuals is computed for each one of these blocks as follows

$$MS = \frac{1}{h \times w} \sum_{h,w} [Org(i,j) - Pred(i,j)]^2 \qquad (1)$$

where h and w represent the *height* and *width* of the CU and i and j represent the coordinates of each pixel within the block
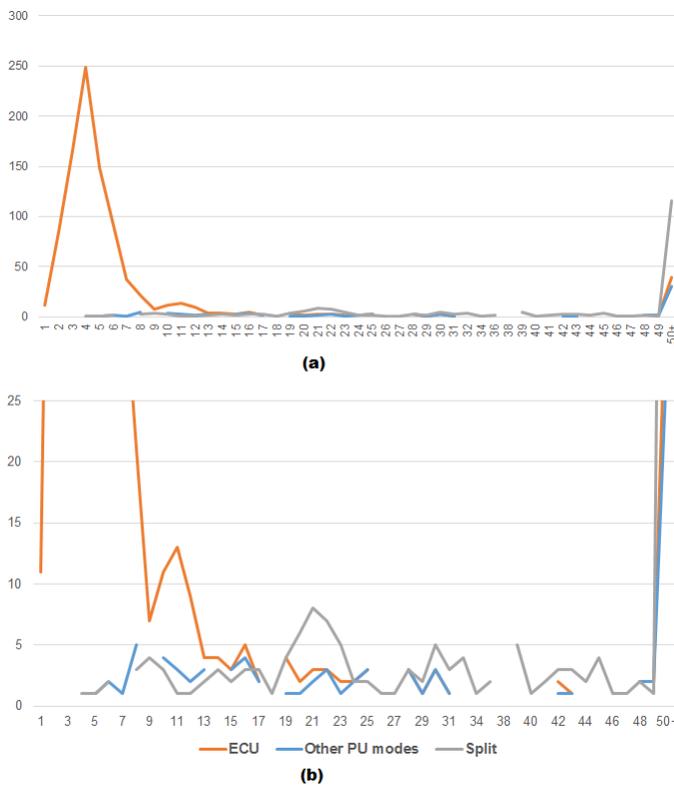

(a)

(b)

Fig. 6. a) Frequency of occurrence of ECU, Other PU modes and Split CUs as a relationship of MS_{Max} for the *BasketballPass* sequence with QP 22 b) Magnified version for MS_{Max} values for low frequencies of occurrences

The largest MS value, termed as MS_{Max} is used as the coefficient for the threshold values. The first 10 frames of a sample of sequences are studied to find the relationship between the early CU size determination and MS_{Max}. Fig. 6 illustrates the frequency of occurrence of the Early CUs (ECUs), other PU modes and Split CUs as a relationship of MS_{Max} for 16×16 CUs of the *BasketballPass* sequence encoded with a QP value of 22. The ECUs consist of the $2N \times 2N$ CUs (1 PU) as well as those CUs with 2 PUs ($N \times 2N, 2N \times N, 2N \times nU, 2N \times nD, nL \times 2N$ and $nR \times 2N$). It is noted that the majority of the non-split CUs (ECUs) are concentrated at the lower MS_{Max} values while the split CUs

are at the other end with relatively larger MS_{Max} values. In Fig. 6, it is observed that the ECUs are found mostly below 25. In addition, CUs adopt the $2N \times 2N$ PU configuration for values of MS_{Max} less than 10.
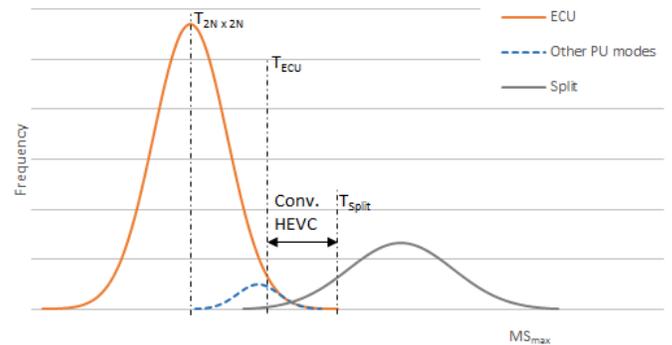

Fig. 7. Using Normal curves to approximate the occurrences of Early CU, Other PU modes and split CUs

The distribution of the three classifications can be approximated to three Normal curves as shown in Fig. 7. The early CU termination threshold is set to the MS_{Max} value for which 90% of the ECUs are captured. A preliminary analysis is performed on the first 10 frames of 4 sequences (one for each class) to determine this value. Table I shows that there is no apparent relationship between the threshold values and the QPs. The thresholds, however, is dependent on the sequences and varies accordingly. In this paper, the average value of 22 is adopted to terminate the CU (for QP = 22) and denoted by T_{ECU}. The threshold value to terminate the CU as $2N \times 2N$ PU is taken as $1/2\ T_{ECU}$ and denoted by $T_{2N \times 2N}$. As shown in Fig. 6, most of the CUs with values below $T_{2N \times 2N}$ (11) actually terminates as $2N \times 2N$ PUs. The high MS_{Max} in Fig. 7 values are the split CUs. The split threshold is therefore set above the ECU threshold, i.e. $1.25 \times T_{ECU}$. These CUs will not be allowed to perform the various modes checks but considered directly as split CUs.

TABLE I. 90% THRESHOLD VALUES FOR ECU BASED ON FIRST 10 FRAMES WITH QP=22

Sequence	CU Size		
	16	**32**	**64**
Cactus	31	33	29
BasketBallDrill	13	13	14
BlowingBubbles	35	30	-
Vidyo1	18	19	15

It is also observed that the MS of the residuals is proportional to the QP values. A large QP (lower quality) normally results in higher residual values. By observing the MS values obtained with same motion vectors and different QP values, the following relationship is formed

$$MS_{27} = 2.4 \times MS_{22} \qquad (2)$$

$$MS_{32} = 5.0 \times MS_{22} \qquad (3)$$

$$MS_{37} = 10.0 \times MS_{22} \qquad (4)$$

Combining the relationship among the different QP values with the threshold defined above for QP value of 22, the different threshold values are set as shown in Table II.

TABLE II.   MEAN SQUARE (MS) THRESHOLD VALUES FOR DIFFERENT QP VALUES

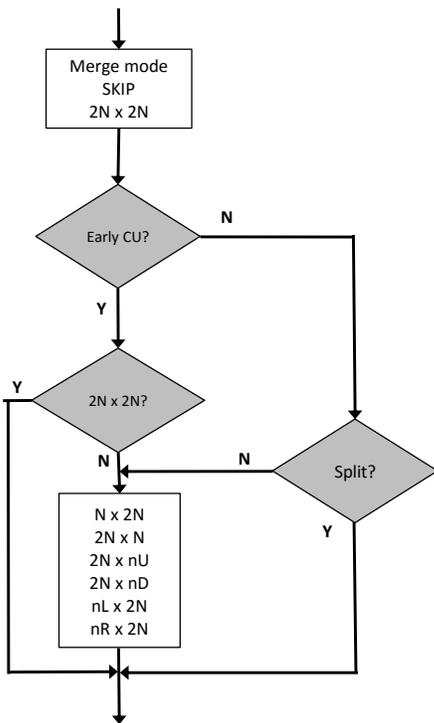| QP	T_{ECU}	$T_{2N \times 2N}$	T_{Split}
22	22	11	28
27	53	26	66
32	110	55	138
37	220	110	275

Fig. 8.   Illustrating early PU mode decisions based on residuals obtained following the $2N \times 2N$ inter-prediction operation

As illustrated in Fig. 7, the region between the ECU threshold and the split threshold is not considered in our complexity reduction approach. The inter-prediction modes of CUs falling in this area, where the curves intersect, are difficult to predict. The risk of an incorrect decision is quite high and may weight significantly on the quality. For this reason, these CUs have be left to HEVC normal processing.

The overall approach is summarized in Fig. 8 where the highlighted (gray) decision boxes indicate the modifications brought about to the conventional HEVC processing for inter-prediction.

VI.   EXPERIMENTS, RESULTS AND DISCUSSION

HEVC Test Model Reference software version 10 (HM10) has been used to implement the proposed early termination

schemes for complexity reduction in inter prediction using the QP values of 22, 27, 32 and 37. Thirteen standard sequences, ranging from class B to class E, as defined in [3] have been used so as to cover a broad range of resolutions. The experiments were performed on 50 frames for each sequence using the *lowdelay B main* encoding configuration with an IBBB Group of Picture (GOP) structure where optimization was carried out on the B frames only.

The performances of the proposed methods, compared to HM10, are reported in terms of the change in average bit rate, peak signal-to-noise ratio (PSNR), Total Encoding Time (TET) and Motion Estimated Time (MET) based on the following formula:

$$\Delta BitRate(\%) = \frac{BitRate(proposed) - BitRate(HM)}{BitRate(HM)} * 100 \quad (5)$$

$$\Delta PSNR(dB) = PSNR(proposed) - PSNR(HM) \quad (6)$$

$$\Delta TET(\%) = \frac{TET(proposed) - TET(HM)}{TET(HM)} * 100 \quad (7)$$

$$\Delta MET(\%) = \frac{MET(proposed) - MET(HM)}{MET(HM)} * 100 \quad (8)$$

TET is the processing time for both the I and the B frames while the MET represents only the encoding time of the B frames for the sequences under consideration. The Bjøntegaard metrics [2] rate distortion performance is used in the computation of BD-Rate and BD-PSNR.

A. Complexity Reduction using Early CU termination

The result of experiments conducted for the early CU termination is presented in Table III. By stopping the splitting process using the early CU termination threshold, T_{ECU}, an average complexity reduction of 41.9% (47.1% for ME) is achieved with a BD-Rate of 0.44%. This reduction is accompanied with a drop of 0.015 dB in terms of PSNR along with a *decrease* of 0.05 in bitrate. The time savings range from 23.3% for the *PartyScene* sequence to 63.0% for the *Vidyo1* sequence.

Class E sequences show the highest reduction in encoding time as they contain large stationary regions leading to large CUs being formed when encoded. Compared to the *PartyScene* sequence (lowest performance), the *Vidyo1* sequence has more CUs of larger sizes. For example, during the first 10 HEVC encoded frames (7 inter-frames), 54% of the 64×64 CUs are terminated in *Vidyo1* as opposed to only 30% for *PartyScene*. However, these high performance sequences from Class E also display high BD-Rates since the fixed thresholds used are slightly higher for these sequences. CUs containing small regions of motion are incorrectly terminated in the proposed approach leading to a relatively higher drop in quality.

It is also noted that the high complexity reduction of 42.7% achieved for the *BQTerrace* sequence yields a BD-Rate of only 0.09%.

TABLE III.   RESULTS OF COMPLEXITY REDUCTION USING EARLY CU ONLY

Class	Sequence	Δ PSNR [dB]	Δ TET [%]	Δ MET [%]	Δ Bit-Rate [%]	BD-PSNR [dB]	BD-Rate [%]
B	BQTerrace	-0.004	-42.7	-48.5	-0.06	-0.003	0.09
	Cactus	-0.010	-42.6	-47.9	-0.08	-0.010	0.44
	Kimono	-0.010	-42.9	-46.4	0.07	-0.013	0.37
	ParkScene	-0.022	-45.1	-50.3	-0.12	-0.030	0.86
C	BasketBallDrill	-0.015	-39.0	-43.7	-0.01	-0.015	0.34
	BQMall	-0.015	-33.1	-38.0	-0.03	-0.016	0.33
	PartyScene	-0.010	-23.3	-26.9	-0.01	-0.011	0.26
D	BasketBallPass	-0.011	-38.5	-44.0	0.09	-0.013	0.36
	BlowingBubbles	-0.012	-23.8	-26.6	-0.01	-0.013	0.38
	BQSquare	-0.015	-29.5	-34.4	-0.10	-0.012	0.18
E	Vidyo1	-0.029	-63.0	-71.0	-0.17	-0.030	0.83
	Vidyo3	-0.021	-58.0	-64.5	-0.09	-0.015	0.45
	Vidyo4	-0.020	-62.8	-69.9	-0.10	-0.023	0.77
	Average	**-0.015**	**-41.9**	**-47.1**	**-0.05**	**-0.016**	**0.44**

B. Complexity Reduction using Early CU and Early PU Termination

Table IV shows the result of applying the early CU and the early PU thresholds simultaneously. An average complexity reduction of 48.1% (54.4% for ME) is achieved with a BD-Rate of 0.55%. The drop of PSNR is only 0.023 dB while a *decrease* in the bit rate is also observed. In fact, almost half of the number of CUs terminated earlier are identified as $2N \times 2N$ PUs. For these CUs, the other modes processing are avoided leading to the additional time savings. A similar trend as in the early CU threshold is found for the performance of each sequence.

TABLE IV.   RESULTS OF COMPLEXITY REDUCTION USING EARLY CU AND EARLY PU

Class	Sequence	Δ PSNR [dB]	Δ TET [%]	Δ MET [%]	Δ Bit-Rate [%]	BD-PSNR [dB]	BD-Rate [%]
B	BQTerrace	-0.004	-48.0	-54.4	-0.07	-0.004	0.12
	Cactus	-0.014	-48.8	-55.0	-0.09	-0.010	0.48
	Kimono	-0.019	-51.1	-55.2	0.12	-0.022	0.72
	ParkScene	-0.034	-51.9	-58.1	-0.15	-0.045	1.24
C	BasketBallDrill	-0.026	-46.6	-52.2	-0.03	-0.023	0.58
	BQMall	-0.020	-39.9	-45.5	-0.02	-0.022	0.46
	PartyScene	-0.015	-27.0	-31.1	-0.04	-0.020	0.32
D	BasketBallPass	-0.023	-46.6	-53.8	0.00	-0.021	0.51
	BlowingBubbles	-0.016	-26.2	-30.0	-0.07	-0.023	0.46
	BQSquare	-0.020	-34.1	-39.4	-0.13	-0.015	0.23
E	Vidyo1	-0.045	-73.2	-82.4	-0.20	-0.041	1.15
	Vidyo3	-0.031	-67.1	-75.1	-0.09	-0.020	0.70
	Vidyo4	-0.036	-71.9	-80.0	-0.18	-0.034	1.03
	Average	**-0.023**	**-48.7**	**-54.8**	**-0.07**	**-0.023**	**0.62**

C. Complexity Reduction using Early CU and Early PU Termination along with Early Splitting

In addition to the early termination of CUs and PU modes, CUs with high MS values at the level of the $2N \times 2N$ mode processing are identified as split CUs. The processing associated with the other modes processing are therefore avoided and the splitting is performed directly. The results of combining all three techniques is provided in Table V. The reduction in TET ranges from 50.1% for the *BlowingBubbles* sequence to 76.0% for the *Vidyo1* sequence. An average overall complexity reduction of 62.2% (70.8% for ME) is thus achieved along with a BD-Rate of only 1.14%. The MET shows an average reduction of 70.8% in encoding time.

TABLE V.   RESULTS OF COMPLEXITY REDUCTION USING EARLY CU, EARLY PU AND EARLY SPLITTING

Class	Sequence	Δ PSNR [dB]	Δ TET [%]	Δ MET [%]	Δ Bit-Rate [%]	BD-PSNR [dB]	BD-Rate [%]
B	BQTerrace	-0.012	-60.2	-68.9	0.04	-0.017	0.46
	Cactus	-0.024	-61.5	-69.8	0.16	-0.031	1.13
	Kimono	-0.030	-64.5	-70.8	0.56	-0.051	1.56
	ParkScene	-0.043	-63.6	-71.9	0.00	-0.048	1.47
C	BasketBallDrill	-0.037	-62.3	-70.6	0.37	-0.055	1.28
	BQMall	-0.040	-58.7	-67.7	0.43	-0.066	1.44
	PartyScene	-0.032	-50.1	-58.5	0.18	-0.039	0.84
D	BasketBallPass	-0.033	-62.1	-72.5	0.43	-0.056	1.15
	BlowingBubbles	-0.038	-50.3	-58.2	0.19	-0.051	1.10
	BQSquare	-0.035	-51.5	-60.7	0.06	-0.042	0.64
E	Vidyo1	-0.045	-76.0	-85.8	-0.06	-0.045	1.30
	Vidyo3	-0.042	-72.7	-81.8	0.21	-0.059	1.23
	Vidyo4	-0.036	-75.1	-83.8	0.02	-0.036	1.28
	Average	**-0.034**	**-62.2**	**-70.8**	**0.20**	**-0.046**	**1.14**

The best and worst performances in terms of BD-Rate during the experiments are illustrated in Fig. 9 for the *BQTerrace* sequence and in Fig. 10 for the *Kimono* sequence. The *BQTerrace* sequence shows practically no deviation with the proposed approach from the standard HEVC encoder while the *Kimono* sequence (worst case) indicates only a very slight deviation. This comparison further confirms the high time savings produced by the proposed approach with negligible deterioration in quality.

D. Comparison with related works

A number of works on complexity reduction for HEVC inter prediction have already been published. Comparison with existing works have been limited to those based on the *low-delay main* profile of the HEVC. In addition, since different degree of reductions with varying BD-Rates are achieved, the performance indicator ratio, *BD-Rate/ΔTET*, proposed in [8] is used for comparison purposes in Table VI. The comparison is grouped into 3 categories with complexity reduction around 40%, 50% and 60%. The proposed approach for each category, outperforms the other works in terms of the ratio *BD-Rate/ΔTET*, confirming the effectiveness of the proposed approach.
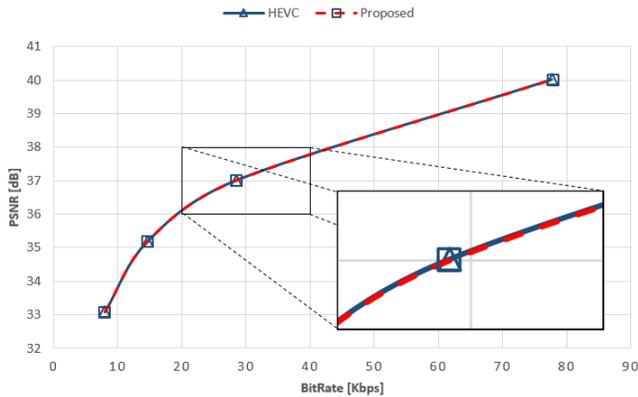
Fig. 9. Performance of *BQTerrace* sequence for complexity reduction using Early CU, Early PU and Early Splitting
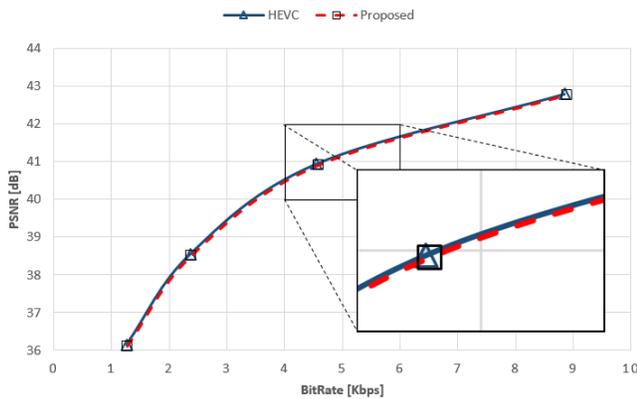

Fig. 10. Performance of *Kimono* sequence for complexity reduction using Early CU, Early PU and Early Splitting

TABLE VI.   COMPARISON WITH RELATED WORKS

Category	Related Works	Δ TET [%]	BD-Rate [%]	BD-Rate/ ΔTET [%]
40%	Goswami [12]	-39.53	0.50	1.26
	Shen [29]	-41.00	1.15	2.80
	Ahn [1]	-42.70	1.00	2.34
	Proposed (ECU)	**-41.90**	**0.44**	**1.05**
50%	He [14]	-45.34	1.23	2.71
	Lin [21]	-49.87	0.74	1.48
	Zhang [41]	-51.45	1.98	3.85
	Xiong [36]	-52.00	1.61	3.10
	Proposed (ECU+EPU)	**-48.70**	**0.62**	**1.27**
60%	Liu [23]	-56.71	1.05	1.85
	Xiong [37]	-59.67	2.19	3.67
	Proposed (ECU+EPU+Split)	**-62.20**	**1.14**	**1.83**

For the 40% category, the BD-Rate of 0.44% achieved in the proposed ECU termination approach is well below those of comparable works leading to a *BD-Rate/ΔTET* value of only 1.05. The proposed $ECU + EPU$ termination approach in the 50% category results in 48.7% time savings. Compared to other works, the BD-Rate is lower along with a *BD-Rate/ΔTET* ratio of only 1.27 . The proposed approach in the high reduction category (60%) also surpasses other works in terms of the *BD-Rate/ΔTET* ratio and with an overall complexity reduction of 62.2%.

VII.   CONCLUSION

In this paper, a new approach is proposed to enhance the complexity reduction in the HEVC inter prediction process for the *lowdelay Main* profile. The Mean-Square (MS) of each 8×8 block of the residuals following the $2N \times 2N$ inter prediction processing are computed. The highest MS value is used for comparison with the determined thresholds for early termination of the CU and PU processing along with the early identification of the splitting decision.

When this technique is applied to CUs of size 16×16, 32×32 and 64×64, an average overall complexity reduction of 62.2% (70.8% for ME) is achieved at a BD-Rate of only 1.14%.

The proposed approach achieves a higher encoding time reduction compared with the state-of-the-art algorithms while maintaining a good average bitrate performance. In the proposed work, a set of static thresholds are used. The good performances of the proposed work in predicting the termination of CUs and PU processing are observed by the low *BD-Rate/ΔTET* ratios. However, since the proposed approach is based purely on static thresholds, it produces relatively higher BD-Rates when the actual sequence thresholds should have been slightly higher or lower. Therefore, the proposed work can still be enhanced by including in the thresholds some characteristics intrinsic of each sequence.

REFERENCES

[1] S. Ahn, B. Lee, and M. Kim. A Novel Fast CU Encoding Scheme Based on Spatiotemporal Encoding Parameters for HEVC Inter Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):422–435, March 2015.

[2] G. Bjøntegaard. Calculation of Average PSNR Differences between RDCurves . Doc. VCEG-M33, VCEG, April 2001.

[3] F. Bossen. Common test conditions and software reference configurations . Doc. JCTVC-B300, JCT-VC, July 2010.

[4] B. Bross, P. Helle, H. Lakshman, and K. Ugur. Inter-picture prediction in HEVC. In *High Efficiency Video Coding (HEVC)*, pages 113–140. Springer, 2014.

[5] 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 *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 53–56, May 2013.

[6] G. Chen, Z. Pei, L. Sun, Z. Liu, and T. Ikenaga. Fast intra prediction for HEVC based on pixel gradient statistics and mode refinement. In *Signal and Information Processing (ChinaSIP), 2013 IEEE China Summit International Conference on*, pages 514–517, July 2013.

[7] K. Choi, S-H Park, and E.S. Jang. Coding tree pruning based CU early termination . Doc. JCTVC-F092, JCT-VC, July 2011.

[8] G. Correa, P. Assuncao, L. Agostini, and L.A. da Silva Cruz. Four-step algorithm for early termination in HEVC inter-frame prediction based on decision trees. In *Visual Communications and Image Processing Conference, 2014 IEEE*, pages 65–68, Dec 2014.

[9] T.L. da Silva, L.V. Agostini, and L.A. da Silva Cruz. Speeding up HEVC intra coding based on tree depth inter-levels correlation structure. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5, Sept 2013.

[10] H. Fan, R. Wang, L. Ding, X. Xie, H. Jia, and W. Gao. Hybrid Zero Block Detection for High Efficiency Video Coding. *IEEE Transactions on Multimedia*, 18(3):537–543, March 2016.

[11] K. Goswami, B-G Kim, D. Jun, S-H Jung, and J.S. Choi. Early Coding Unit-Splitting Termination Algorithm for High Efficiency Video Coding (HEVC). *ETRI Journal*, 36(3):407–417, June 2014.

[12] K. Goswami, J-H Lee, and B-G Kim. Fast algorithm for the high efficiency video coding (hevc) encoder using texture analysis. *Inf. Sci.*, 364(C):72–90, October 2016.

[13] M. Grellert, M. Shafique, M.U.K. Khan, L. Agostini, J.C.B. Mattos, and J. Henkel. An adaptive workload management scheme for HEVC encoding. In *2013 IEEE International Conference on Image Processing*, pages 1850–1854, Sept 2013.

[14] J. He, X. He, X. Li, and L. Qing. Fast Inter-Mode Decision Algorithm for High-Efficiency Video Coding Based on Textural Features. *Journal of Communications*, 9(5):441–447, May 2014.

[15] W.J. Hsu and H.M. Hang. Fast coding unit decision algorithm for HEVC. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–5, Oct 2013.

[16] Q. Hu, X. Zhang, Z. Shi, and Z. Gao. Neyman-Pearson-Based Early Mode Decision for HEVC Encoding. *IEEE Transactions on Multimedia*, 18(3):379–391, March 2016.

[17] A. Lee, D. Jun, J. Kim, J.S. Choi, and J. Kim. Efficient Inter Prediction Mode Decision Method for Fast Motion Estimation in High Efficiency Video Coding. *ETRI Journal*, 36(4):528–536, August 2014.

[18] J.H. Lee, B.G. Kim, D.S. Jun, S.H. Jung, and J.S. Choi. Complexity reduction algorithm for prediction unit decision process in high efficiency video coding. *IET Image Processing*, 10(1):53–60, 2016.

[19] J.H. Lee, C.S. Park, and B.G. Kim. Fast coding algorithm based on adaptive coding depth range selection for HEVC. In *Consumer Electronics - Berlin (ICCE-Berlin), 2012 IEEE International Conference on*, pages 31–33, Sept 2012.

[20] X. Li, R. Wang, W. Wang, Z. Wang, and S. Dong. Fast motion estimation methods for HEVC. In *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–4, June 2014.

[21] T-L Lin, C-C Chou, Z. Liu, and K-H Tung. HEVC early termination methods for optimal CU decision utilizing encoding residual information. *Journal of Real-Time Image Processing*, pages 1–17, 2016.

[22] Y. C. Lin and J. C. Lai. Edge Density Early Termination Algorithm for HEVC Coding Tree Block. In *Computer, Consumer and Control (IS3C), 2014 International Symposium on*, pages 39–42, June 2014.

[23] Z. Liu, T-L Lin, and C-C Chou. Efficient prediction of CU depth and PU mode for fast HEVC encoding using statistical analysis. *Journal of Visual Communication and Image Representation*, 38:474–486, 2016.

[24] C-S Park, G-S Hong, and B-G Kim. Novel Intermode Prediction Algorithm for High Efficiency Video Coding Encoder,. *Advances in Multimedia*, 2014, 2014.

[25] N. Purnachand, L.N. Alves, and A. Navarro. Fast Motion Estimation Algorithm for HEVC. In *Consumer Electronics - Berlin (ICCE-Berlin), 2012 IEEE International Conference on*, pages 34–37, Sept 2012.

[26] F. Sampaio, S. Bampi, M. Grellert, L. Agostini, and J. Mattos. Motion Vectors Merging: Low Complexity Prediction Unit Decision Heuristic for the Inter-prediction of HEVC Encoders. In *2012 IEEE International Conference on Multimedia and Expo*, pages 657–662, July 2012.

[27] K. Saurty, P. C. Catherine, and K. M. S. Soyjaudah. Terminating CU splitting in HEVC intra prediction using the Hadamard Absolute Difference (HAD) cost. In *SAI Intelligent Systems Conference (IntelliSys), 2015*, pages 836–841, Nov 2015.

[28] K. Saurty, P. C. Catherine, and K. M. S. Soyjaudah. *Fast Intra Mode Decision for HEVC*, pages 361–383. Springer International Publishing, Cham, 2016.

[29] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang. An Effective CU Size Decision Method for HEVC Encoders. *IEEE Transactions on Multimedia*, 15(2):465–470, Feb 2013.

[30] X. Shen, L. Yu, and J. Chen. Fast coding unit size selection for HEVC based on Bayesian decision rule. In *Picture Coding Symposium (PCS), 2012*, pages 453–456, May 2012.

[31] Y. Shi, O.C. Au, H. Zhang, X. Zhang, L. Jia, W. Dai, and W. Zhu. Local saliency detection based fast mode decision for HEVC intra coding. In *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on*, pages 429–433, Sept 2013.

[32] G.J. Sullivan, J. Ohm, W-J. Han, and T. Wiegand. Overview of the High Efficiency Video Coding (HEVC) Standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1649–1668, Dec 2012.

[33] H. L. Tan, C. C. Ko, and S. Rahardja. Fast Coding Quad-Tree Decisions Using Prediction Residuals Statistics for High Efficiency Video Coding (HEVC). *IEEE Transactions on Broadcasting*, 62(1):128–133, March 2016.

[34] J. Vanne, M. Viitanen, and T. D. Hämäläinen. Efficient Mode Decision Schemes for HEVC Inter Prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(9):1579–1593, Sept 2014.

[35] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, July 2003.

[36] J. Xiong, H. Li, F. Meng, Q. Wu, and K.N. Ngan. Fast HEVC Inter CU Decision Based on Latent SAD Estimation. *IEEE Transactions on Multimedia*, 17(12):2147–2159, Dec 2015.

[37] J. Xiong, H. Li, F. Meng, S. Zhu, Q. Wu, and B. Zeng. MRF-Based Fast HEVC Inter CU Decision With the Variance of Absolute Differences. *IEEE Transactions on Multimedia*, 16(8):2141–2153, Dec 2014.

[38] J. Xiong, H. Li, Q. Wu, and F. Meng. A Fast HEVC Inter CU Selection Method Based on Pyramid Motion Divergence. *IEEE Transactions on Multimedia*, 16(2):559–564, Feb 2014.

[39] J. Yang, J. Kim, K. Won, H. Lee, and B. Jeon. Early SKIP Detection for HEVC . Doc. JCTVC-G543, JCT-VC, November 2011.

[40] H. Zhang and Z. Ma. Early termination schemes for fast intra mode decision in High Efficiency Video Coding. In *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 45–48, May 2013.

[41] Y. Zhang, S. Kwong, X. Wang, H. Yuan, Z. Pan, and L. Xu. Machine Learning-Based Coding Unit Depth Decisions for Flexible Complexity Allocation in High Efficiency Video Coding. *IEEE Transactions on Image Processing*, 24(7):2225–2238, July 2015.