Efficient Multi-Strategy Intra Prediction for Quality Scalable High Efficiency Video Coding

Dayong Wang, Ce Zhu, Yu Sun, Frederic Dufaux, Yuanyuan Huang

Efficient Multi-Strategy Intra Prediction for Quality Scalable High Efficiency Video Coding

Dayong Wang, Ce Zhu, *Fellow, IEEE,* Yu Sun, *Member, IEEE,* Frederic Dufaux, *Fellow, IEEE,* and Yuanyuan Huang,

***Abstract*—As an extension of High Efficiency Video Coding (HEVC), the Scalable High Efficiency Video Coding (SHVC) introduces multiple layers with inter-layer predictions, which greatly increases the complexity on top of the already complicated HEVC encoder. In Intra prediction for Quality SHVC, Coding Tree Unit (CTU) allows recursive splitting into four depth levels, which considers 35 Intra prediction modes and inter-layer reference (ILR) mode to determine the best possible mode at each depth level. This achieves the highest coding efficiency but incurs a substantially high computational complexity. In this paper, we propose a novel Intra prediction scheme to effectively speed up the enhancement layer Intra-coding in Quality SHVC. The new features of the proposed framework include: First, spatial correlation and its correlation degree are combined to predict most probable depth level candidates. Second, for a given depth candidate, based on the probabilities of ILR mode, we check the ILR mode by examining the residual distribution based on skewness and kurtosis to determine whether the residuals follow a Gaussian distribution. In that case, the Intra prediction comparisons, which require a high complexity, are skipped. Third, during Intra prediction selection from 35 Intra prediction modes, spatial and inter-layer correlations are combined with the local monotonicity of the Hadamard costs associated with the modes in a small neighborhood, to examine only a portion of Intra prediction modes. Finally, a hypothesis testing on the currently selected depth level is performed to examine whether the residuals present significant differences within their block to early terminate depth selection. The proposed multi-step multi-strategy scheme aims to minimize the number of depth selections while greatly reducing the mode decision complexity for a depth candidate in a hierarchical fashion. Our experimental results demonstrate that the proposed scheme can achieve a speedup gain of more than 75% in average on the test video sequences, while maintaining almost the same coding efficiency. .**

***Index Terms*—SHVC, mode decision, depth decision, Intra prediction, complexity.**

I. Introduction

WITH the extensive use of mobile devices, ever-increasing number of users are browsing and sharing video contents on social networks. Broadband networks, especially 3G/4G wireless networks, have allowed these video communications to become important parts of people's daily lives [1]. Notably, the number of diverse video applications, such as digital TV broadcasting, video conferencing, wireless video streaming, and smart phone communications, has consistently increased. In the above and many other applications, various resolution levels, different types of resource constraints, and network bandwidths are often required. Conventional video coding schemes cannot effectively meet these diverse needs [2]. Scalable Video Coding (SVC) provides an attractive alternative for these applications. SVC is the scalable extension of the H.264/AVC, whose stream consists of a base layer (BL) and one or more Enhancement layers (ELs). Through selecting an appropriate EL, SVC adapts to a wide variety of device capabilities, network conditions, and client applications [3]. In order to provide this adaptability, SVC needs to encode multiple layers and perform inter-layer prediction, which lead to a very complex and slow encoding process.

With the increasing popularity of high-resolution video applications and services, more recently, the next generation video coding standard, i.e., High Efficiency Video Coding (HEVC) has been developed. HEVC has evolved from previous video coding standards by adding more advanced features and higher-efficiency coding tools to improve compression performances. Notably, HEVC is capable of decreasing bit rates by approximatively 50% in comparison with H.264/AVC, while still maintaining the same high video quality level [4]. Since HEVC has very high coding efficiency, it receives extensive attention and has wide applications. Its high coding efficiency is obtained at the cost of high coding complexity, which is about two to four times that of H.264/AVC [5]. In order to accommodate different device capabilities, network conditions, and client applications, MPEG and ITU have introduced the scalable extension of HEVC, known as Scalable High Efficiency Video Coding (SHVC) [6].SHVC supports different scalability features, including temporal scalability (frame rate from low to high), spatial scalability (spatial resolution from low to high), quality scalability (quality from low to high), as well as bit-depth scalability (bit depth from low to high, e.g., 8 to 10 bit), and color gamut scalability (color gamut from narrow to wide, e.g., ITU-R Recommendation BT.709 to BT.2020). Since SHVC need to encode multiple layers and perform inter-layer prediction, and each layer has to perform HEVC encoding, the coding complexity of SHVC is even higher. Therefore, improving the coding speed is always highly desired, especially for wireless and real-time applications.

In this paper, we propose a multi-step multi-strategy scheme to accelerate the coding speed of Intra prediction for quality SHVC (QS). The novelties and the contributions of the proposed algorithm are summarized as follows: (1) spatial correlation and its correlation degree are used jointly in prediction; (2) based on the probabilities of the Inter-layer reference (ILR) mode, a statistical test is applied to test whether the residual coefficients of ILR obey a Gaussian distribution and to determine the necessity of Intra prediction checking; (3) spatial and inter-layer correlations are combined with the relationship between Intra modes (IMs) and their corresponding Hadamard

TABLE I
SUMMARY OF ACRONYMS

Acronym	Definition
SVC	Scalable Video Coding
BL	Base layer
EL	Enhancement layer
HEVC	High Efficiency Video Coding
SHVC	Scalable High Efficiency Video Coding
QS	Quality SHVC
ILR	Inter-layer reference
IM	Intra mode
HC	Hadamard cost
MB	Macro-block
RD	Rate-distortion
AZB	All-zero block
CU	Coding Unit
SAD	Sum of absolute difference
SMP	Symmetric motion partition
AMP	Asymmetric motion partition
QP	Quantization parameter
ME	Motion estimation
CTU	Coding Tree Unit
RMD	Rough Mode Decision
RDO	Rate Distortion Optimization
CBDP	Correlation-Based Depth Prediction
DB-IMD	Distribution-Based ILR Mode Decision
HCB-IMP	Hadamard-Cost Based Intra Mode Prediction
SDB-DET	Significant Difference Based Depth Early Termination
CSTC	Common SHM test conditions
PU	Prediction Unit
MPM	Most Probable Mode
LMP	Local Minimum Point
SS	Spatial SHVC

cost (HC) values to predict candidate IMs; (4) residual coefficients of current depth are tested for significant differences to determine early termination. Experimental results demonstrate that the proposed algorithm can significantly improve the coding speed with negligible losses in coding efficiency.

The remainder of this paper is organized as follows. Section II provides related work. Section III presents an overview of our proposed algorithm. Section IV proposes four fast decision methods to improve the coding speed. Experimental results and conclusions are presented in Section V and Section VI, respectively.

II. RELATED WORK

In this section, we first summarize fast algorithms for H.264/SVC, and then review fast algorithms for HEVC. Finally, we present fast algorithms for SHVC. For the reader's convenience, the acronyms used in the paper are listed in Table I.

To improve the coding speed for H.264/SVC, Li et al. [7, 8] and Lin et al. [9] predict candidate macro-block (MB) modes in EL based on the co-located MB modes in BL. Since these algorithms use inter-layer correlation to remove unlikely modes, the coding speed is improved. However, as these algorithms only exploit inter-layer correlation in prediction, the speed-up is limited. Kim et al. [10] use the modes of the co-located MB and its neighboring MBs in BL to predict candidate modes of the current MB in EL. Since the algorithm uses both inter-layer and spatial correlations in MB modes to predict candidate modes and exclude unlikely modes, the coding speed is obviously improved. Generally, early terminations are also efficient ways to improve the coding speed. Using a statistical approach, Park et al. [11] derive the expectation of the Rate-distortion (RD) cost of each mode first, and then encode the modes according to this expectation. When the RD cost is smaller than a set threshold, the encoding procedure is terminated. Yeh et al. [12] firstly propose to analyze and predict modes in EL statistically. The Bayesian theorem is then used to detect whether the prediction mode of the current MB is the best. Finally the method further predicts and refines the aforementioned mode when it is detected not to be optimal by the Markov process. Jung et al. [13] predict MBs in EL to be All-zero blocks (AZBs) based on empirical analysis of the inter-layer and spatial correlation of the AZB. Then, only predicted MBs are examined and terminated by using AZB detection algorithm. Zhao et al. [14] develop a constrained model with optimal termination based on inter-layer and spatial correlations, and then use the model to initialize the candidate mode list and early terminate the coding process. These algorithms use early terminations to improve coding speed. However, if only early terminations are used, the speed-up is limited. Lu et al. [15] exploit inter-layer and neighboring correlations, as well as examine the level of picture details and motion activities to predict candidate modes and early terminate the encoding process. Wang et al. [16] use inter-layer and spatial correlations to estimate candidate modes and exclude low likelihood modes, and make use of RD cost and residual coefficients to end the coding process early. In this way, the coding speed is effectively improved.

In order to improve the coding speed, several fast algorithms have been proposed for HEVC [17]-[26]. Zhang et al. [17] selectively check the candidate IMs through the Hadamard cost-based progressive rough mode search, and early terminate Coding Unit (CU) split according to the RD cost. Min et

al. [18] calculate both global and local edge complexities in horizontal, vertical, 45 diagonal, and 135 diagonal directions and use them to decide the partitioning of a CU. Its four sub-CUs are then processed in the same way to early terminate CU split. Cho et al. [19] propose early CU split decision and early CU pruning decision at each CU depth level to improve coding speed. Shen et al. [20] skip some specific depth levels based on spatial correlations, and then skip some prediction modes based on RD cost and prediction mode correlations among different depth levels or spatial correlations.

The above algorithms are targeting fast Intra prediction for HEVC. Actually, inter prediction is even more complex, hence many fast inter prediction algorithms have also been developed for HEVC. Shen et al. [21] use three adaptive inter mode decision strategies to improve coding speed based on correlations of prediction modes, motion vectors and RD costs among different depth levels and among spatially temporally adjacent CUs. Lee et al. [22] define the upper-bound of sum of absolute difference (SAD). When the SAD is smaller than the upper-bound, a predefined threshold is compared to determine zero block. Vanne et al. [23] propose a conditional evaluation of symmetric motion partition (SMP) modes, range limitations primarily in the SMP sizes and secondarily in the asymmetric motion partition (AMP) sizes, and a selection of the SMP and AMP ranges as a function of the quantization parameter (QP), to optimize the decision of SMPs and AMPs. Zhao et al. [24] use the depth information of the collocated block from a previous frame to predict and check the size of the current block. The inter-prediction residuals are then analyzed to determine whether to terminate the mode decision process or to skip unnecessary modes and split the block into smaller sizes. Next, a fast discrete cross difference is adopted to detect the dominant IM. Finally, four early termination strategies are used to terminate coding process. The research proposed in [25] determine CU depth range and skip some specific depth levels based on temporal and spatial correlations, and early terminate coding process based on motion homogeneity checking, RD cost checking and SKIP mode checking to skip motion estimation (ME) on unnecessary CU sizes. Pan et al. [26] propose an early MERGE mode decision for the root CUs based on the AZB and the ME information of the Inter2Nx2N mode. An early MERGE mode decision is considered for children CUs based on the mode selection correlation between the root CU and the children CUs. When the root CUs are encoded in the non-MERGE modes, the AZB and the ME information are also used for early termination of children CUs.

Since HEVC uses advanced features and higher-efficiency coding tools compared to H.264/AVC, existing complexity reduction schemes proposed for H.264/SVC cannot directly be applied to SHVC [27]. Different from HEVC, SHVC exploits inter-layer correlation in prediction. Although fast algorithms developed for HEVC can be applied for SHVC, the coding speed cannot always be significantly improved. Therefore, it is crucial to improve the coding speed of SHVC. In this paper, we mainly focus on coding speed improvement for QS.

Bailleul et al. [28] propose to only encode the depth of the co-located CU in BL, and disallow Intra prediction and orthogonal block modes of the co-located CU to improve the coding speed. Since too many depths and modes are skipped, the coding speed is significantly improved but the coding efficiency is severely degraded. Ge et al. [29] skip the depths in EL that are larger than those of the co-located CU in BL. Since the process is very simple, the performance is not optimal. The method proposed in [30] improves the coding speed by investigating the early termination of motion prediction mode search based on inter-layer correlations. Since it only uses inter-layer correlations for early mode search termination, the speed-up remains limited. Tohidypour et al. [31] predict candidate modes and skip the remaining modes based on the depth and scalable layer for the current CU in the EL, when the best mode for at least one of the parent CUs in previous depth layers of the same quad-tree structure is the merge mode. Moreover, the correlations between the mode information of EL's CUs and BL's CUs are also exploited to further eliminate candidate modes. In this manner, the coding speed can be improved.

Although these above-reviewed algorithms can somewhat improve the coding speed, they are only applicable for inter prediction of QS. Conversely, only a few researches have focused on improving the coding speed of Intra prediction for quality SHVC so far. Wang et al. [32] propose to skip low likelihood depths based on inter-layer correlations, then combine IMs and their corresponding HC values to skip low likelihood IMs, so as to improve the coding speed of Intra prediction for QS. Although unlikely depths and IMs are skipped, the depth prediction is not extensively investigated and ILR mode is not studied. Tohidypour et al. [33] use the coding tree unit (CTU) partitioning structure of the already encoded CTUs in the EL and BL to predict the coding unit sizes of the current CTUs in the EL. However, this algorithm does not consider ILR and IMs, therefore the coding speed is not significantly increased.

In summary, all of the above algorithms using different approaches to improve the coding speed. However, according to the best of our knowledge, some aspects have not been considered yet in the literature in order to speed up the coding process, including: (1) Correlations are usually used to predict depths, however, correlation degrees have not been considered, which might affect the accuracy of prediction; (2) Since the content of a CU in EL and the co-located CU in BL are exactly the same, inter-layer correlation is very strong. Therefore, many CUs may select ILR mode as the best mode. Developing approaches targeting ILR are highly desirable; (3) Spatial and inter-layer correlations are often used to predict IMs. In addition, HC values are also used to predict IMs. However, the spatial and inter-layer correlations and HC values are not fully exploited, leading to insignificant improvement of coding speed; (4) AZBs or thresholds obtained by experiment are often used to early terminate depth selection. However, the number of AZBs is often limited. Moreover, a threshold obtained by experiments lacks theoretical basis and may not be optimal in all conditions.

Based on the above considerations, we propose in this paper a new and fast Intra prediction algorithm for QS. In order to improve the coding speed, first, the spatial correlation in

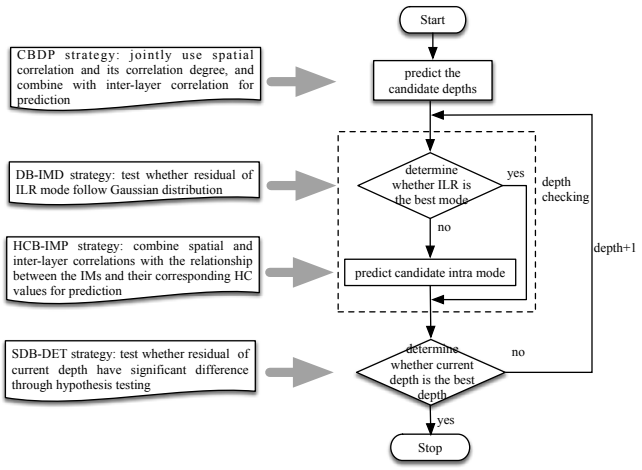


Fig. 1. Flowchart of the overall algorithm.

BL is studied and used as the spatial correlation in EL. It is combined with inter-layer correlations to predict candidate depths. Then, the ILR mode is checked and the distribution of residual coefficients is tested to determine the necessity of Intra prediction checking, where rough mode decision (RMD) and rate distortion optimization (RDO) procedures are included. In the RMD procedure, only a portion of IMs is checked by combining spatial and inter-layer correlations with the relationship between the IMs and their corresponding HC values. Finally, residual coefficients of the current depth are tested for significant differences in variances to determine early termination.

III. Overview of the Proposed Multi-Strategies Framework

The objective of our proposed multi-strategies framework for fast Intra prediction is to improve the coding speed and maintain the coding efficiency for QS. To achieve this objective, we propose four strategies: Correlation-Based Depth Prediction (CBDP); Distribution-Based ILR Mode Decision (DB-IMD); Hadamard-Cost Based Intra Mode Prediction (HCB-IMP); and Significant Difference Based Depth Early Termination (SDB-DET). The overview of the algorithm is summarized in Fig.1. First, the depth candidates are predicted through CBDP. For the selected current depth candidates, DB-IMD determines whether the ILR mode is the best mode. In the affirmative, Intra prediction do not need to be checked; otherwise the IM candidates are predicted through HCB-IMP. After the depth has been checked, the residuals of the depth are examined by SDB-DET to determine whether it is the best depth for early termination through SDB-DET.

The proposed four strategies are shown in the left part of Fig. 1.

IV. Proposed Fast Intra Prediction Algorithm

In this section, we present in more details the different components of the proposed fast Intra prediction algorithm. In order to develop the algorithm, we have conducted extensive experiments to investigate the relative methods and relationship for fast decision. To meet the different resolution requirements, two sequences in each format B, C, D and E are selected in our experiments. More specifically, the following test sequences are used: Sunflower and Tractor in B format; Flowervase and PartyScene in C format; BlowingBubbles and RaceHorses in D format; and Parkruner and Town in E format. Test conditions are listed as follows: each test sequence is encoded using all I-frame structure, the maximum CU size is 64 and maximum partition depth is 4. As suggested in common SHM test conditions (CSTC)[34], the QPs used for the BL are set as (26, 30, 34, 38), and the corresponding QPs used for the EL are set as (22, 26, 30, 34) and (20, 24, 28, 32), respectively. Experimental results show that these two settings of QPs produce similar performance. Therefore, we only present the experimental results obtained by QPs with (22, 26, 30, 34) for brevity. Based on these experiments, we propose our efficient fast decision methods that will be described below.

A. Correlation-based depth prediction (CBDP)

Similar to HEVC, SHVC usually allows the maximum size of CU to be 64, and the depth levels range from 0 to 3. Since every depth contains the whole Intra prediction and ILR prediction process, the induced coding complexity is very high. Therefore, skipping depths with low likelihood is very important in improving the coding speed.

1) Spatial correlation prediction: Since strong spatial correlations exist in natural video content, neighboring CUs are highly similar in motion and textural features, and their coding depths are also highly correlated. Therefore, the coding depth of the current CU can be predicted from its neighboring CUs. However, it is not optimal to directly and simply predict the current CU depth from its neighboring CU depths. The degree of correlation between the current CU and its neighboring CUs should be estimated. Afterwards, the candidate depths should be obtained and sequentially checked from the one with the highest probability to the one with the lowest probability. Thus, the approach to obtain the correlation degrees of the CUs in the EL becomes a crucial issue. Since CUs in the BL and the co-located CUs in the EL corresponds to the same content, the spatial correlations of the CUs in the BL can be equivalently utilized by the co-located CUs in the EL. Fig. 2 shows the CUs for the prediction of the current CU depth. C is the current CU in the EL, L is the left CU, U is the upper CU, UL is the upper-left CU, and UR is the upper-right CU. Accordingly, BC, BL, BU, BUL and BUR are the collocated CUs of C, L, U, UL and UR in the BL, respectively.

Obviously, the more similar the depths of two neighboring CUs in the BL are, the more likely the depths of the co-located CUs in the EL will be the same. Therefore, if the depths of neighboring CUs in the BL are very similar, the spatial correlation degree of the co-located CUs in the EL should be very strong, and vice versa. Obviously, the spatial correlation degrees are inversely proportional to the absolute depth difference between neighboring CUs in the BL. Note also that the maximum absolute difference of neighboring CUs in depth is 3. The spatial correlation degree is represented as

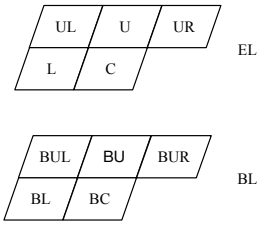


Fig. 2. Relative CUs for the prediction of the current CU's depth.

spatial depth weights. In order to represent different spatial depth weights, a power of two is used. Of course, other powers can also be used as long as they can distinguish different spatial correlation degrees. From the above analysis, the spatial weights of all depths are set as follows:

$$w_h^s = 2^{3-|h_{bn}-h_{bc}|} \tag{1}$$

.

where h refers to the h-th depth and s represents spatial correlation, w_h^s denotes the spatial weight of the h-th depth, and h_{bc} and h_{bn} represent the depths of the co-located CUs of the current CU and its neighboring CU in BL, respectively. For a depth in neighboring CUs, its spatial weight can be obtained according to (1). For neighboring CUs with the same depth, the spatial weight should be the sum of all these CUs' spatial weights. If neighboring CUs of current CU do not contain a depth, its spatial weight is set to be 0. Through the above process, spatial weights of all depths can be obtained.

2) Inter-layer correlation prediction: Since a CU's content in EL and the co-located CU in BL are exactly the same, inter-layer correlation is strong. Obviously, when a CU in BL uses a depth, the co-located CU in EL is more likely to use similar or even the same depths. Similar to the spatial correlation prediction above, more similar depths induce larger inter-layer weights and vice versa. The inter-layer weight of a depth is set by

$$w_h^l = 2^{3-|h-h_{bc}|} \tag{2}$$

.

where w_h^l represents the inter-layer weight of the $h-th$ depth and l represents inter-layer correlation. Inter-layer weights of all depths can be obtained according to Eq. (2).

3) Combine spatial and inter-layer correlations to predict candidate depths: The spatial and inter-layer weights of all depths can be obtained by applying the above procedures. Since both spatial and inter-layer correlations are strongly correlated with depth decision, the weight w_h of the h-th depth can be derived as

$$w_h = w_h^s + w_h^l \tag{3}$$

.

The weights of all depths are obtained in this way and sorted in a descending order. If the first two depths are smaller than or equal to 1, only depth 0 and 1 will be selected for checking. In this condition, the corresponding coding efficiency losses are listed in Table II, in which BDBR [35] measures the bitrate difference at equal PSNR in the EL.

TABLE II
THE CODING EFFICIENCY LOSSES FOR CBDP

Format	Sequences	BDBR
B	Sunflower	0.7%
B	Tractor	0.2%
C	Flowervase	0.4%
C	PartyScene	-0.2%
D	BlowingBubbles	0.0%
D	RaceHorses	0.2%
E	Park	-0.1%
E	Town	-0.1%
Average		0.14%

In Table II, we observe that the average coding efficiency loss is 0.14%, whereas the maximum loss is 0.7%, hence the coding efficiency loss is negligible. In terms of complexity, if two depths out of 4 are selected, the coding speed can be improved by more than 50%. As a good trade-off between coding efficiency and speed, the depth selection can be summarized as follows.

(1) The default depth is set to 3.

(2) If the first two depths are smaller than or equal to 1, only depth 0 and 1 are required to be checked.

(3) If the first three depths are smaller than or equal to 2, only depth 0, 1 and 2 are required to be checked.

(4) If the first two depths are 2 or 3, depth 0 is not likely to be selected and will be skipped.

B. Distribution-based ILR mode decision (DB-IMD)

1) Distribution of ILR mode: In QS, since frame resolutions between the BL and EL are the same, the inter-layer correlation is very high. Therefore, a CU in the EL searches for the best matching CU in the reconstructed pixels, by using the ILR mode. It should be well predicted and many CUs may select it as the best mode. In order to improve the coding speed, we only select ILR to test under the aforementioned test conditions. The corresponding coding efficiency losses are listed in Table III.

From Table III, it can be seen that the average coding efficiency loss is -0.19%, which is negligible. However, in sequence "Flowervase", the coding efficiency loss is significantly larger than the other sequences.

2) Distribution of ILR mode: According to the coding efficiency loss, if the Intra prediction were directly skipped, the coding efficiency would be obviously degraded in some sequences. On the contrary, if the Intra prediction were always checked, a large amount of unnecessary coding time would be wasted in most sequences. In order to improve the coding speed and maintain the coding efficiency, ILR mode is first

TABLE III
THE CODING EFFICIENCY LOSSES WITH ONLY ILR MODE

Format	Sequences	BDBR
B	Sunflower	0.0%
B	Tractor	-0.1%
C	Flowervase	-0.6%
C	PartyScene	-0.2%
D	BlowingBubbles	-0.2%
D	RaceHorses	-0.2%
E	Park	0.0%
E	Town	-0.2%
Average		-0.19%

checked and then it is determined whether it is the best mode. In the affirmative, the Intra prediction including RMD and RDO procedure will be skipped to improve the coding speed. Otherwise, the Intra prediction needs to be checked to maintain the coding efficiency. The key problem is how to determine if the ILR mode is the best mode. In general, if a mode has been predicted very well, the residual coefficients will obey a certain distribution [36]. Therefore, it can be determined whether the ILR mode is the best mode by studying the distribution of residual coefficients. Residual coefficients are typically modeled using a Gaussian distribution [37,38] or a Laplacian distribution [39]. The Gaussian distribution is selected due to its superior performance in our experiments. In order to determine whether residual coefficients obey a Gaussian distribution, a test based on skewness and kurtosis detection is proposed, as described below.

Suppose $x_1, x_2, \cdots, x_n$ are residual coefficients, the moment estimators G_1 and G_2 of skewness and kurtosis are

$$G_1 = \frac{B_3}{B_2^{\frac{3}{2}}}, G_2 = \frac{B_4}{B_2^2}, \tag{4}$$

where B_k(k=2,3,4) is the sample central moment of order k and is given by the expression

$$B_k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^k}{n}, \tag{5}$$

when n is relatively large (generally more than 100), G_1 and G_2 can be approximated by

$$G_1 \sim N(\mu_1, \sigma_1^2) = N(0, \frac{6(n-2)}{(n+1)(n+3)}), \tag{6}$$

$$G_2 \sim N(\mu_2, \sigma_2^2) = N(3 - \frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}), \tag{7}$$

where μ_1 , μ_2 are the expected values of G_1 and G_2 respectively, σ_1 and σ_2 are the variances of G_1 and G_2 respectively. Then, it can be determined whether the residual coefficients

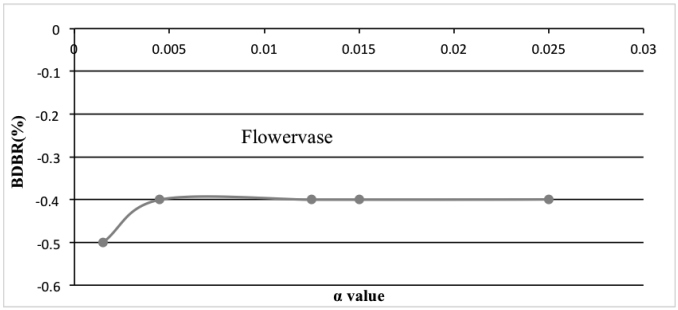


Fig. 3. α and the corresponding coding efficiency for DB-IMD.

obey the Gaussian distribution with expected values and variances. The standardized expected values are given by

$$u_1 = \frac{G_1}{\sigma_1}, u_2 = \frac{G_2 - \mu_2}{\sigma_2}. \tag{8}$$

.

According to statistical hypothesis testing, a significance level α refers to the probability of wrongly rejecting the null hypothesis that the distribution is Gaussian. Its corresponding test critical value z_α can be obtained by checking Gaussian distribution table. The decision of Gaussian distribution is then expressed as

$$|u_1| <= z_\alpha, |u_2| <= z_\alpha. \tag{9}$$

.

If the two conditions defined in Eq. (9) are satisfied, residual coefficients can be assumed to obey a Gaussian distribution. In that case, the ILR mode can be assumed to be the best mode, and then the Intra prediction in the CU will be skipped.

In order to significantly improve the coding speed and maintain the coding efficiency, obtaining an optimal value of α and the corresponding z_α is key. From Table III, we can find that sequence "Flowervase" has the largest coding efficiency loss; so skipping Intra prediction has the greatest effect on this sequence. If "Flowervase" can achieve very high coding efficiency, the other sequences definitely get higher performances. Therefore, we only need to test this sequence to select the optimal α. Toward this end, some commonly used α values are selected for testing, such as 0.0005, 0.0015, 0.0025, 0.0045, 0.0125, 0.015 and 0.025. The corresponding coding efficiency represented with BDBR is shown in Fig.3.

From Fig.3, we observe that there is a turning point when α is equal to 0.0045. If α is smaller than 0.0045, the corresponding coding efficiency changes dramatically. When α is greater than or equal to 0.0045, the coding efficiency stays approximately constant. Therefore, α should be greater than or equal to 0.0045. As we known, the smaller α is, the larger the corresponding z_α is and the encoding speed increases. Based on the above analysis α is set to 0.0045 and the corresponding z_α is 2.61.

C. Hadamard-cost based Intra mode prediction (HCB-IMP)

1) Relationship between IM and HC value: To enhance the coding efficiency of video frames, HEVC includes 35 IMs

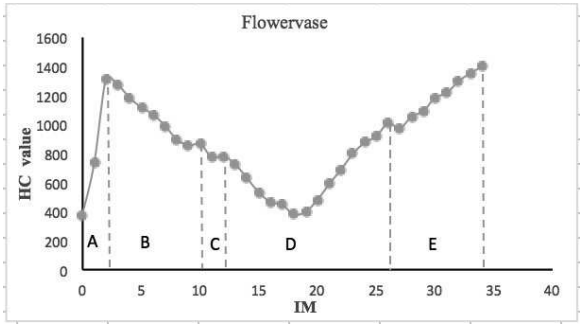


Fig. 4. The relationship between IMs and HC values.

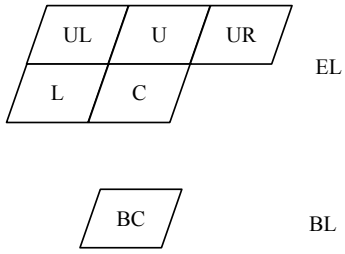


Fig. 5. Relative CUs for the prediction of the current CU's initial points.

for the prediction unit (PU) with 4-Level splitting of block partitions ranging from 64×64 to 8×8. In the RMD procedure, the first best N candidates are selected from all 35 modes. Subsequently, the Most Probable Modes (MPMs) obtained from the left and upper PUs in EL and the co-located PU in BL, are adopted as candidate modes. Then, those candidates enter the RDO process to select the optimal mode. To facilitate more accurate prediction results, a complex texture unit tends to select a smaller size of PU and 8 IMs as candidates, whereas a simple texture unit tends to select a larger size of PU and 3 IMs as candidates.

Since there are 35 IMs in every CU, the coding complexity is very high. To improve the coding speed while maintaining the coding efficiency, only IMs with high likelihood are chosen for checking and IMs with low likelihood are excluded. For this purpose, the relationships between the IMs and their corresponding HC values are firstly investigated. The candidate IMs are subsequently obtained based on their relationships. In Fig. 4, the relationships between the IMs and their corresponding HC values are presented with "Flowervase" sequence in C format. A local minimum point (LMP) with its neighborhood is partitioned as a zone, separated by the nearest local maximum points.

In Fig. 4, five zones A, B, C, D and E are presented, where each zone contains a LMP. Obviously, the point with the smallest HC should be identified within these LMPs. Each point is associated with its IM in the horizontal axis and its HC value in the vertical axis, respectively. As we known, the best IM should have the smallest RD cost in the 35 IMs. Since the RD cost has a very strong correlation with the HC value, the best IM is most likely to correspond to the identified point with the smallest HC, and the LMPs should be searched. In each zone, it can be observed that all the left and right points are monotonically decreasing when approaching their LMPs. Extensive experiments have demonstrated that the 35 IMs may be divided into a varying number of intervals for different sequences. Within each interval, both the left and right neighbors are monotonically decreasing while approaching their LMPs.

2) IM prediction based on the relationship between IM and HC value: According to the relationship between IM and HC value, the LMP within a zone can be obtained by following descent direction of the HC values. In other words, if the HC value of a point is smaller than those of its left and right neighbors, the point is declared to be the LMP of its zone. In order to improve the coding speed and preserve coding efficiency, the selection of the initial points is crucial in identifying the LMPs. Clearly, relative CUs, including neighboring CUs in the EL and the co-located CU in the BL, have IMs similar to the current CU. Therefore, the IMs of these five relative CUs are selected as initial points as shown in Fig. 5.

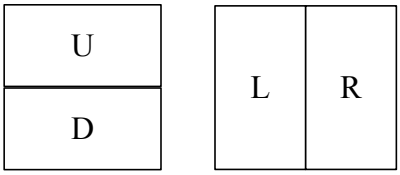
In Fig. 5, C is the current CU in the EL, L is the left CU, U is the upper CU, UL is the upper-left CU, UR is the upper-right CU, and BC is the co-located CU of C in BL. Initial points are obtained through the process as described above. Since duplicate points may exist, redundant points should be properly discarded to avoid unnecessary search. As discussed earlier, the LMPs are obtained by following the descent direction of the HC values from the initial points. Moreover, modes 0 and 1 which refer to the planar and DC modes have not been considered yet, thus they will also be selected for checking. It should be noted that all LMPs are arranged in ascending order of HC values. Obviously, if all initial points obtained from the five relative CUs are 0 or 1, the number of LMPs is 2. If all initial points obtained from the five relative CUs are different and not 0 or 1, the largest number of LMPs is 7.

After the RMD process, the best IM can be obtained from the LMPs through the RDO process. Obviously, the number of LMPs ranges from 2 to 7. When the size of PU is larger than 8×8, 3+MPM modes need to be checked. If there are only 2 LMPs, 1+MPM modes can be skipped during checking. When the size of PU is smaller than or equal to 8×8, 8+MPM modes need to be checked, so 1+MPM to 6+MPM modes can be skipped. In this way, the coding speed is further improved during the RDO process. In order to test the effectiveness of this approach, the coding efficiency is shown in Table IV.

From Table IV, we can observe that the maximum coding efficiency loss is -0.4% and the average coding efficiency loss is -0.11%. The coding efficiency loss is therefore negligible. Notably, since we only select 0,1 and the IMs of five relative CUs as the initial points to search, many IMs with low probabilities are initially skipped and some IMs are further skipped in the RDO procedure. In this way, the coding speed is significantly improved.

TABLE IV
THE CODING EFFICIENCY LOSSES FOR HCB-IMP

Format	Sequences	BDBR
B	Sunflower	0.0%
B	Tractor	-0.1%
C	Flowervase	-0.4%
C	PartyScene	-0.1%
D	BlowingBubbles	-0.1%
D	RaceHorses	-0.1%
E	Park	0.0%
E	Town	-0.1%
Average		-0.11%



(a) Top and bottom (b) Left and right

Fig. 6. The division of the CU

D. Significant difference based depth early termination (SDB-DET)

In the above-described process, both inter-layer and spatial correlations are used to predict the candidate depths and skip depths with low probabilities to improve the coding speed and maintain the coding efficiency. As it is known, if a depth is predicted accurately, the residual coefficients will obey a certain probability distribution. Therefore, the residual coefficients are firstly checked to determine whether they follow a Gaussian distribution, which however, is only the necessary condition. More concretely, even though residual coefficients obey a Gaussian distribution, the corresponding depth may not be the best depth. Hence, further check is needed. Toward this end, a CU for residual coefficients is divided into top and bottom parts as shown in Fig. 6(a) and a statistical test is carried out to determine whether significant differences exist for these two parts. A similar statistical test is then conducted for left and right parts as shown in Fig. 6(b). If there is no significant difference in both tests, it is not necessary to further check the next depth. In contrast, further checking is carried out when any of the two tests show significant differences.

Since residuals have already been checked to obey a Gaussian distribution, hypothesis testing is used to determine whether the two parts for each division present significant differences. For this purpose, F-test is used, which is the test statistics to investigate the significance of the difference between two sampled variances. The variances of the CU residual coefficients can be calculated by

$$s^2 = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(r_{ij} - \bar{r})}{M \times N}, \tag{10}$$

where $M \times N$ is the size of the partitioned CU,r_{ij} is the residual coefficient value in (i,j) and $\bar{r}$ is the average value of the corresponding residual coefficients in the partitioned CUs. CUs can take three sizes, 64×64, 32×32 and 16×16. Straightforwardly, in the case of top-bottom partitioning (Fig. 6(a)), respectively left-right partitioning (Fig. 6(b)), the corresponding sizes are 64×32, 32×16 and 16×8, respectively 32×64, 16×32 and 8×16. As mentioned above, the residual coefficients obey a Gaussian distribution. Suppose that the residual coefficients of two parts in each division can be respectively modeled as[40]

$$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \tag{11}$$

.where X_1 is a Gaussian distribution of one part, X_2 is a Gaussian distribution of the other part in a division. For example, in the case of up-down partitioning (Fig. 6(a)), if X_1 is a Gaussian distribution of up part, X_2 is a Gaussian distribution of down part.

Assuming σ_1=σ_2, the test statistics F can be computed as,

$$F = \frac{S_1^2}{S_2^2}, \tag{12}$$

where S_1^2 and S_2^2 are respectively the variances of the two parts and can be calculated by Eq. (10). In order to ensure σ_1=σ_2, the condition can be expressed as

$$\frac{1}{F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)} \leqslant F \leqslant F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1), \tag{13}$$

where n_1 and n_2 are the number of residual coefficients of the two parts in each subdivision, and β is the significance level value. Given β , the corresponding threshold $F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)$ is determined by F distribution table. The key is to select the best β and the corresponding $F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)$.

When the depth is 0 and the corresponding CU size is 64×64, we can firstly adopt commonly used values for β, such as 0.10, 0.05, 0.025, 0.01 and 0.005. Since the sample number is very large, all their corresponding $F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)$ is 1. If the $F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)$ is set to 1, the coding speed cannot be improved. In order to improve the coding speed while maintaining coding efficiency, values close to 1 are tested and the corresponding coding efficiencies are shown in Fig.7.

In Fig. 7, the horizontal axis represents the value of $F_{\frac{\beta}{2}}(n_1 - 1, n_2 - 1)$ abbreviated as F, and the vertical axis represents the coding efficiency loss denoted by $BDBR$. From Fig.7, we can observe that F equals 1 and 1.01 result in identical coding efficiency, however the speed-up is very negligible. When F is 1.04 and 1.05, the $BDBR$ is 0 for all test sequences, except for the sequence "town" which shows a very small loss of -0.1%. Additionally, experiments also show that there is no significant difference in terms of coding speed

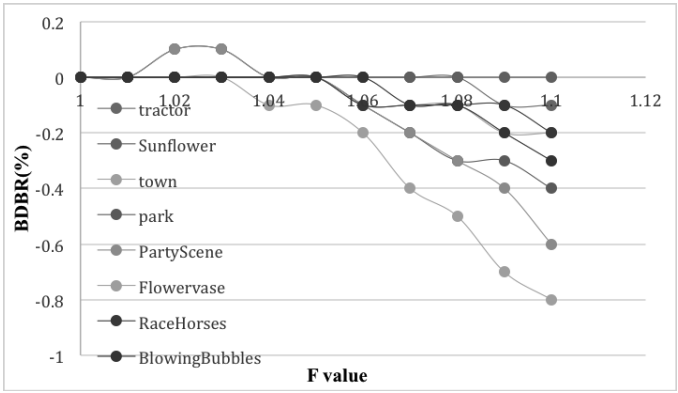


Fig. 7. F and the corresponding coding efficiency for SDB- DET.

between F = 1.04 and F =1.05. Therefore, F =1.04 is used hereafter. Adopting the same methodology, we can also obtain the threshold of depth 1 (CU size is 32×32) and depth 2 (CU size is 16×16) as follows

$$F = \begin{cases} 1.04 & 64 \times 64 \\ 2.0 & 32 \times 32 \\ 3.2 & 16 \times 16 \end{cases} \quad (14)$$

The CU sizes are 64×64, 32×32 and 16×16 and their corresponding depths are 0, 1 and 2, respectively. If a condition is satisfied, the corresponding depth can be terminated and the next depths do not need to be further checked.

V. Experimental results

The proposed fast Intra prediction algorithm for QS has been implemented in the SHVC reference software (SHM 11.0) on a server with Intel (R) 2.0 GHz CPU and 30 GB memory. The experimental parameters are set according to the CSTC[34]. As we are targeting Intra coding, both GOPSize and IntraPeriod are set to be 1. The QPs used for the BL are set to be (26, 30, 34, 38), and the corresponding QPs used for the EL are set to be (22, 26, 30, 34) and (20, 24, 28, 32), respectively.

Note that the eight training sequences, which were previously used to verify the effectiveness of the proposed algorithm, are no longer used as test sequences hereafter. This demonstrates that our proposed algorithm is generic and effective for different types of content.

The proposed algorithm includes four strategies, namely ”CBDP”, ”DB-IMD”, “HCB-IMP”, and “SDB-DET”. Since both “CB-DP” and “SDB-DET” are related to depth, these two methods are combined during testing and denoted as ”Depth”. Since the experimental results show that the EL with (22, 26, 30, 34) and (20, 24, 28, 32) provide similar performances, only the results for the EL with (22, 26, 30, 34) are provided in this section. Coding efficiency is evaluated by BDBR. A negative BDBR represents an increase in coding efficiency compared with the reference software, or more specifically the percentage of bitrate saving for a given quality. Computational complexity is measured by running coding time in the EL, and “TS” represents the percentage of coding time savings in the EL. The coding efficiency losses and the coding speed improvements of the different strategies are shown in Table V.

TABLE V
Performance Comparison among the different strategies

Sequences	HCB-IMP		DB-IMD		Depth	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.2%	29.80%	-0.2%	48.10%	-0.1%	51.35%
PeopleOnStreet	-0.1%	30.82%	-0.1%	45.34%	0.0%	50.90%
BasketballDrive	-0.2%	26.54%	-0.3%	55.28%	0.0%	60.96%
BQTerrace	-0.1%	27.72%	-0.2%	52.35%	0.1%	59.81%
Cactus	-0.2%	29.39%	-0.2%	52.75%	0.0%	59.76%
Kimono	-0.1%	35.59%	-0.2%	61.64%	0.4%	66.72%
ParkScene	-0.1%	33.52%	-0.2%	47.96%	0.0%	51.54%
Average	-0.14%	30.48%	-0.2%	51.92%	0.06%	57.29%

From Table V, we find that the average coding speed improvements in “HCB-IMP”, “DB-IMD” and “Depth” are 30.48%, 51.92%, 57.29%, respectively. The average coding efficiency losses in “HCB-IMP”, “DB-IMD” and “Depth” are -0.14%, -0.2%, 0.06%, correspondingly. Obviously, all the strategies can significantly improve the coding speed with negligible coding efficiency losses. Since every depth contains the whole Intra prediction and ILR prediction process, “Depth” leads to the most significant speed-up among the different strategies. Nevertheless, “DB-IMD” also achieves remarkable computational complexity gains. Since the IM is relative simpler than the two above methods in the whole coding process, “HCB-IMP” leads to the least significant time savings.

To further demonstrate the performance of our proposed algorithm, the overall performance when combining the four proposed strategies, i.e. “CBDP”, “DB-IMD”, “HCB-IMP”, and “SDB-DET”, is evaluated and compared with Hamid’s algorithm [33]. To the best of our knowledge, Hamid’s algorithm is the best performing method in the existing literature in term of improving coding speed of Intra prediction in QS. For fair comparisons, our algorithm and Hamid’s algorithm are implemented on the same computing platform. Two settings of QPs: Q1=(22, 26, 30, 34) and Q2=(20, 24, 28, 32) are used. The overall performance of our algorithm and Hamid’s algorithm are listed in Table VI and Table VII, with Q1 and Q2 respectively.

From Table VI and Table VII, it can be observed that the proposed algorithm can reduce coding time by an average of 75.33% and 75.07% for Q1 and Q2, respectively. The average BDBR increase by -0.07% and -0.17%, respectively. Therefore, it can be concluded from these experiments that our algorithm can significantly improve the coding speed with negligible increases in coding efficiency.

With Q1, the running time of the proposed algorithm is improved by 14.74% in EL and 8.37% in total when compared to Hamid’s algorithm, while a gain of 1.23% is achieved in terms of BDBR compared with Hamid’s algorithm. The gains are very similar with Q2. Therefore, we can draw the

TABLE VI
PERFORMANCE OF HAMID'S AND THE PROPOSED ALGORITHMS WITH Q1

Sequences	Hamid			Proposed		
	BDBR	TS		BDBR	TS	
		EL	Total		EL	Total
Traffic	1.28%	72.02%	40.85%	-0.2%	72.35%	41.04%
PeopleOnStreet	1.91%	60.36%	34.13%	-0.1%	70.80%	40.06%
BasketballDrive	1.33%	59.36%	33.44%	-0.2%	77.06%	43.43%
BQTerrace	0.65%	58.23%	33.62%	0.0%	75.87%	43.84%
Cactus	0.98%	58.42%	33.27%	-0.1%	76.48%	43.58%
Kimono	0.78%	51.96%	29.13%	0.2%	81.87%	45.88%
ParkScene	1.16%	63.78%	36.48%	-0.1%	72.88%	41.71%
Average	1.16%	60.59%	34.42%	-0.07%	75.33%	42.79%

TABLE VII
PERFORMANCE OF HAMID'S AND THE PROPOSED ALGORITHMS WITH Q2

Sequences	Hamid			Proposed		
	BDBR	TS		BDBR	TS	
		EL	Total		EL	Total
Traffic	1.30%	71.56%	42.08%	-0.3%	71.63%	42.14%
PeopleOnStreet	1.94%	60.78%	35.78%	-0.3%	71.08%	41.88%
BasketballDrive	1.32%	59.64%	35.79%	0.0%	77.14%	46.36%
BQTerrace	0.67%	59.13%	36.13%	0.0%	76.08%	46.52%
Cactus	0.96%	58.34%	35.33%	-0.2%	76.51%	46.37%
Kimono	0.80%	52.23%	30.73%	-0.2%	81.40%	47.89%
ParkScene	1.20%	63.87%	38.01%	-0.2%	71.63%	42.69%
Average	1.17%	60.79%	36.26%	-0.17%	75.07%	44.84%

conclusion that the coding speed of the proposed algorithm is much faster with significant higher coding efficiency savings compared with Hamid's algorithm.

The coding speed of the proposed algorithm is significantly improved due mainly to the following reasons: (1) The Intra prediction including the RMD and RDO procedure, and the ILR prediction in every depth, are generally very complicated. Therefore, skipping depths with low likelihood, based on the inter-layer, spatial correlation and early termination, can very significantly improve the coding speed; (2) as the ILR mode is almost always the best mode for CUs, time-consuming encoding process, including RMD and RDO, can be skipped through testing residual coefficients, and (3) instead of all the IMs, only a subset of the IMs is checked to search for the LMPs, based on the relationships between the IMs and their corresponding HC values.

Surprisingly, our scheme even exhibits slightly better RD performance with 0.07% and 0.17% BD rate saving in average, respectively, under CSTC. We take the sequence "BasketballDrive" as an example which shows 0.2% BDBR saving for Q1 with our scheme. The bitrate and Y PSNR with our scheme and the SHM reference software are listed in Table VIII, for QPs of 22, 26, 30, and 34 in EL, respectively.

TABLE VIII
COMPARISON BETWEEN BITRATE AND Y PSNR WITH THE PROPOSED SCHEME AND SHM REFERENCE SOFTWARE

QP in EL	SHM		Proposed	
	Bitrate (kbps)	Y psnr (dB)	Bitrate (kbps)	Y psnr (dB)
22	39449.80	41.15	39123.87	41.12
26	15000.90	38.93	14930.92	38.92
30	8326.92	37.36	8239.88	37.35
34	4698.75	35.83	4649.33	35.81

TABLE IX
RD COSTS IN HCB-IMP AND THE SHM REFERENCE SOFTWARE

SHM				HCB-IMP			
27178	26543	29854	26004	27182	26516	29997	25947
30327	30816	36089	32372	30398	30907	36146	32425
24810	24632	26825	42796	24888	24626	26744	43132

TABLE X
RD COSTS IN DB-IMD AND THE SHM REFERENCE SOFTWARE

SHM				DB-IMD			
32941	35538	35120	33293	33088	35540	35230	33325
29677	31610	33857	32732	29774	31745	33838	32787
27635	28519	33363	32299	27613	28460	33478	32114

From Table VIII, we observe that both bitrates and Y PSNR values with our scheme are smaller than those with SHM for each QP, although the differences are very small. More specifically, to further look into the reason behind, we show the RD costs for the blocks comprising 12 CTU (3 rows and 4 columns) with the left upper location at (64,128) and (256, 64) in the first frame in "HCB-IMP" and the SHM reference software in Table IX and "DB-IMD" and the SHM reference software in Table X, respectively.

From Table IX and Table X, we can find that there are some RD costs with our method are smaller than the corresponding RD costs with SHM. From these results, we can see that while some CTUs suffer a RD loss due to our fast scheme, a number of CTUs show RD cost reductions. Based on our analysis, the reason may be due to the fact that RDO is performed locally rather than globally, and a locally optimized CTU may not always lead to better coding efficiency for the following CTUs [41-43]. Generally, it is content dependent.

VI. Conclusion

In this paper, a novel and effective Intra prediction algorithm for QS is proposed. The encoding procedure includes four fast strategies: "CB-DP" allows skipping depths with low probabilities; "DB-IMD" permits skipping the unnecessary Intra prediction;"HCB-IMP" allows checking only a subset of the IMs instead of all 35 IMs; and "SDB-DET" allows early termination of depth selection. Therefore, the proposed algorithm can significantly improve the coding speed with negligible losses in coding efficiency. It is noted that some common schemes may be extended to inter QS and spatial SHVC (SS), such as CBDP and SDB-DET with some possible adaptations. In addition, DB-IMD and HCB-IMP can also be extended to Intra SS by taking into account the changed inter-layer correlation due to the differences between QS (same resolution but different QP among layers) and SS (same or similar QP but different resolution among layers). In other words, for inter QS and SS, the proposed schemes need to be adapted according to the statistics change of inter-layer correlation, as well as the partitions in the inter modes. This is our ongoing work.

Acknowledgment

This work is supported by National Natural Science Foundation of China under Grant 61401247, 61571102, 61501071 and 51307047, Nature Science Foundation Project of Chongqing under Grant cstc2016jcyjA0543 and cstc2017jcyjAX0142, the Fundamental Research Funds for the Central Universities under Grant ZYGX2014Z003, the National High Technology Research and Development Program of China under Grant 2015AA015903, the Applied Basic Research Program of Sichuan Province under Grant No.2014JY0168 and an open project of electromechanical-automobile discipline of Hubei Province under grant No.XKQ2016003.

References

[1] Z.B.Shi,X.Y.Sun,and F.Wu,"Spatially Scalable Video Coding For HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.12, pp.1813-1826, Dec. 2012.

[2] Z.B.Shi, X.Y.Sun, and J.Z.Xu,"CGS Quality Scalability for HEVC," *IEEE International Workshop on Multimedia Signal Processing(MMSP 2011)*, Hangzhou, China, October. 17-19, 2011.

[3] L.Q.Shen, and Z.Y.Zhang,"Content-Adaptive Motion Estimation Algorithm for Coarse-Grain SVC," *IEEE Transactions on Image Processing*, vol.21, no.5, pp.2582-2591, May. 2012.

[4] G.Correa, P.Assuncao, L.Agostini, and L.S.Cruz, "Performance and Computational Complexity Assessment of High-Efficiency Video Encoders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.12, pp.1899-1909, Dec. 2012.

[5] S.Q.Yan, L.Hong, W.F.He and Q.Wang, "Group-Based Fast Mode Decision Algorithm for Intra Prediction in HEVC," *IEEE International Conference on Signal Image Technology and Internet Based Systems*, Naples, Italy, November. 25-29, 2012.

[6] J.M.Boyce, Y.Ye, and J.l.Chen, and A.K.Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology, vol.26, no.1, pp.20-34, Jan. 2016.*

[7] He Li, Z.G. Li, C.Y Wen, and L.P.Chau, "Fast Mode Decision for Spatial Scalable Video Coding," *IEEE International Symposium on Circuits and Systems*, Kos, Greece, May. 21-24, 2006.

[8] He Li, Z. G. Li, and C.Y Wen, "Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.16, no.7, pp.889-895, Jul. 2006

[9] H.C.Lin, W.H.Peng, and H.M.Hang, "A fast mode decision algorithm with macroblock-adaptive rate-distortion estimation for intra-only scalable video coding," *IEEE International Conference on Multimedia and Expo*, Hannover, Germany, June. 23-26,2008.

[10] B.G.Kim, K.Reddy, Y.Y.Park, "Fast mode decision algorithm for inter-frame coding in H.264 extended Scalable Video Coding," *IEEE International Symposium on Circuits and Systems*, Taipei, Taiwan, May. 24-27, 2009.

[11] C.S.Park, B.K.Dan, H.Choi, and S.J.Ko, "A Statistical Approach for Fast Mode Decision in Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.19, no.12, pp.1915-1920, Dec. 2009.

[12] C.H.Yeh, K.J.Fan, and M.J.Chen, "Fast Mode Decision Algorithm for Scalable Video Coding Using Bayesian Theorem Detection and Markov Process," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.20, no.4, pp.563-574, Apr. 2010.

[13] S.W.Jung, S.J.Baek, C.S.Park, and S.J.Ko, "Fast Mode Decision Using All-Zero Block Detection for Fidelity and Spatial Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.20, no.2, pp.201-206, Feb. 2010.

[14] T.S.Zhao, S.Kwong, H.l.Wang, and C.C.Kuo, "H.264/SVC Mode Decision Based on Optimal Stopping Theory," *IEEE Transactions on Image Processing*, vol.21, no.5, pp.2607-2618, May. 2012.

[15] X.Lu, and G.R. Martin, "Fast Mode Decision Algorithm for the H.264/AVC Scalable Video Coding Extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.23, no.1, pp.846-855, Jan. 2013.

[16] D.Y Wang, C.Yuan, Y.Sun, J. Zhang, and X.Jin, "A fast mode decision algorithm applied to Coarse-Grain quality Scalable Video Coding," *Journal of Visual Communication and Image Representation*, vol.25, no.7, pp.1631-1639, Jul. 2014.

[17] H.Zhang, and Z.Ma, "Fast Intra Mode Decision for High-Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.24, no.4, pp.660-668, Apr. 2014.

[18] B.Min, and R.C.C. Cheung, "A Fast CU Size Decision Algorithm for HEVC Intra Encoder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.5, pp.892-896, May. 2015.

[19] S.Cho, and M.Kim, "Fast CU Splitting and Pruning for Suboptimal CU Partitioning in HEVC Intra Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.23, no.9, pp.1555-1564, Sep. 2013.

[20] L.Q.Shen, Z.Y.Zhang, and P.An, "Fast CU Size Decision and Mode Decision Algorithm for HEVC Intra Coding," *IEEE Transactions on Consumer Electronics*, vol.59, no.1, pp.207-213,Jan. 2013.

[21] L.Q.Shen, Z.Y.Zhang, and Z.Liu, "Effective CU Size Decision for HEVC Intra coding," *IEEE transaction on image processing*, vol.23, no.10, pp.4232-4241,Oct. 2014.

[22] K.Lee, H.J.Lee, J.Kim, and Y.Choi, "A Novel Algorithm for Zero Block Detection in High Efficiency Video Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol.7, no.6, pp.1124-1134,Dec.2013.

[23] J.Vanne, M.Viitanen,and T.D.Hmlinen,"Efficient Mode Decision Schemes for HEVC Inter Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.24, no.9, pp.1579-1593, Sep. 2014.

[24] W.J Zhao, T.Onoye, and T.Song, "Hierarchical Structure based Fast Mode Decision for H.265/HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.10, pp.1651-1664, Oct. 2015.

[25] L.Q.Shen, Z.Liu, X.P.Zhang, W.Q.Zhao, and Z.Y.Zhang, "An Effective CU Size Decision Method for HEVC Encoders," *IEEE Transaction on Multimedia*, vol.15, no.2, pp.465-470, Feb. 2013.

[26] Z.Q.Pan, S.Kwong, M.T.Sun, and J.J.Lei, "Early MERGE Mode Decision Based on Motion Estimation and Hierarchical Depth Correlation for HEVC," *IEEE transaction on Broadcasting*, vol.60, no.2, pp.405-412, Jun. 2014.

[27] H.R.Tohidypour, M.T. Pourazad, and P.Nasiopoulos, "Probabilistic Approach for Predicting the Size of Coding Units in the Quad-Tree Structure of the Quality and Spatial Scalable HEVC," *IEEE transaction on Multimedia*, vol.18, no.2, pp.182-195, Feb. 2016.

[28] R.Bailleul, J.Decock, and R.V.D.Walle., "Fast Mode Decision for SNR scalability in SHVC Digest of Technical Papers," *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2014, pp. 191-192.

[29] Ge. Q.Y, and Hu.D, "Fast encoding method using CU depth for quality scalable HEVC," *IEEE Workshop on Advanced Research and Technology in Industry Applications*, Ontario, Canada, September.1366-1370, 2014.

[30] H.R.Tohidypour, M.T. Pourazad, and P.Nasiopoulos, "Content adaptive complexity scheme for quality/fidelity scalable HEVC," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May. 26-31,2013.

[31] H.R.Tohidypour, M.T. Pourazad, and P.Nasiopoulos, "An Encoder Complexity Reduction Scheme for Quality/Fidelity Scalable HEVC," *IEEE transaction on Broadcasting*,vol.62, no.3, pp.664-674, Sep. 2016.

[32] D.Y.Wang, C.Yuan, Y.Sun, J.Zhang, and H.N.Zhou, "Fast Mode and Depth Decision Algorithm for Intra Prediction of Quality SHVC," *Intell. Comput. Theory*, ser. Lecture Notes in Comput. Sci., vol. 8588, pp. 693?699, 2014.

[33] H.R.Tohidypour, M.T. Pourazad, and P.Nasiopoulos, "Probabilistic Approach for Predicting the Size of Coding Units in the Quad-Tree Structure of the Quality and Spatial Scalable HEVC," *IEEE transaction on Multimedia*, vol.18, no.2, pp.182-195, Feb. 2016.

[34] Common SHM Test Conditions and Software Reference Configurations, *Doc. JCTVC-Q1009, ITU-T SG 16 WP 3 and ISO/IEC JTC1/SC 29/WG 11*, Mar. 2014.

[35] G.Bjontegaard. Calculation of average PSNR difference between RD-curves.*13th VCEG-M33 Meeting*, Austin, TX, Apr.2-4, 2001.

[36] H.L.Wang, and S.Kwong, "Hybrid Model to Detect Zero Quantized DCT Coeffcients in H.264," *IEEE Transactions on Multimedia*,vol.9, no.4, pp.728-735, June. 2007.

[37] S.Cho, and M.Kim, "Fast CU Splitting and Pruning for Suboptimal CU Partitioning in HEVC Intra Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.23, no.9, pp.1555-1564, Sep. 2013.

[38] N.Hu, and E.H.Yang, "Fast Motion Estimation Based on Confidence Interval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.24, no.8, pp.1310-1322, Aug. 2014.

[39] B.Lee, and M.Kim, "Modeling Rates and Distortions Based on a Mixture of Laplacian Distributions for Inter-Predicted Residues in Quadtree Coding of HEVC," *IEEE Signal Processing Letters*, vol.18, no.10, pp.571-574, Oct. 2011.

[40] B.Lee, M.Kim, S.Hahm, I.J.Cho, and C.Park, "A Low Complexity Encoding Scheme for Coarse Grain Scalable Video Coding," *IEEE International Conference on Visual Information Engineering*, Xian, China, July. 29-Aug. 1, 2008.

[41] S. Li, C. Zhu, Y.B. Gao, Y.M. Zhou, F. Dufaux, and M.T. Sun, "Lagrangian Multiplier Adaptation for Rate-Distortion Optimization with Inter-frame Dependency," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 117-129, Jan. 2016.

[42] Y. Gao, C. Zhu, S. Li, and T. Yang, "Layer-Based Temporal Dependent Rate-Distortion Optimization in Random-Access Hierarchical Video Coding," *The 18th International Workshop on Multimedia Signal Processing (MMSP 2016)*, Montreal, Canada, Sep. 21-23, 2016.

[43] T. Yang, C. Zhu, X. Fan, and Q. Peng, "Source Distortion Temporal Propagation Model for Motion Compensated Video Coding Optimization," *IEEE International Conference on Multimedia and Expo (ICME 2012)*, Melbourne, Australia, July 2012.