

Fast Depth and Inter Mode Prediction for Quality Scalable High Efficiency Video Coding

Dayong Wang, Yu Sun, Ce Zhu, Weisheng Li, Frédéric Dufaux

▶ **To cite this version:**

Dayong Wang, Yu Sun, Ce Zhu, Weisheng Li, Frédéric Dufaux. Fast Depth and Inter Mode Prediction for Quality Scalable High Efficiency Video Coding. IEEE Transactions on Multimedia, 2020, 22 (4), pp.833-845. 10.1109/TMM.2019.2937240 . hal-02270536

HAL Id: hal-02270536
https://hal.science/hal-02270536

Submitted on 10 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fast Depth and Inter Mode Prediction for Quality Scalable High Efficiency Video Coding

Dayong Wang, Yu Sun, *Member, IEEE,* Ce Zhu, *Fellow, IEEE,* Weisheng Li, *Member, ACM,* and Frederic Dufaux, *Fellow, IEEE*

***Abstract*—The scalable high efficiency video coding (SHVC) is an extension of high efficiency video coding (HEVC), which introduces multiple layers and inter-layer prediction, thus significantly increases the coding complexity on top of the already complicated HEVC encoder. In inter prediction for quality SHVC, in order to determine the best possible mode at each depth level, a coding tree unit can be recursively split into four depth levels, including merge mode, inter2Nx2N, inter2NxN, interNx2N, interNxN, inter2NxnU, inter2NxnD, internLx2N and internRx2N, intra modes and inter-layer reference (ILR) mode. This can obtain the highest coding efficiency, but also result in very high coding complexity. Therefore, it is crucial to improve coding speed while maintaining coding efficiency. In this research, we have proposed a new depth level and inter mode prediction algorithm for quality SHVC. First, the depth level candidates are predicted based on inter-layer correlation, spatial correlation and its correlation degree. Second, for a given depth candidate, we divide mode prediction into square and non-square mode predictions respectively. Third, in the square mode prediction, ILR and merge modes are predicted according to depth correlation, and early terminated whether residual distribution follows a Gaussian distribution. Moreover, ILR mode, merge mode and inter2Nx2N are early terminated based on significant differences in Rate Distortion (RD) costs. Fourth, if the early termination condition cannot be satisfied, non-square modes are further predicted based on significant differences in expected values of residual coefficients. Finally, inter-layer and spatial correlations are combined with residual distribution to examine whether to early terminate depth selection. Experimental results have demonstrated that, on average, the proposed algorithm can achieve a time saving of 71.14%, with a bit rate increase of 1.27%.**

***Index Terms*—SHVC, coding depth, coding mode, mode prediction, early termination.**

This work was supported in part by the Natural Science Foundation of China under Grant 61401247, 61571102, U1713213, and 61972060, in part by Chongqing science and Technology Commission Project under Grant cstc2016jcyjA0543, cstc2017jcyjAX0142, cstc2018jcyjAX0525, cstc2018jcyjAX0225 and cstc2018jszx-cyzdX0124, in part by the Key Project of Sichuan Provincial Department of Science and Technology under Grant 2018JY0035, in part by Project of Sichuan Provincial Department of Science and Technology under Grant 2018RZ0072, in part by Open Project of Hubei University of Arts and Science under Grant XK2018013, in part by International Science and Cooperation Project of Hubei Province under Grant 2019AHB059, in part by faculty sabbatical leave fund from Uni. of Central Arkansas. *(Corresponding author: Ce Zhu)*

D. Wang is with Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wangdayong@cqupt.edu.cn).

Y. Sun is with the Department of Computer Science, University of Central Arkansas, Conway, AR 72035 USA (e-mail: yusun@uca.edu).

C. Zhu is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: eczhu@uestc.edu.cn).

W. Li is with Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: liws@cqupt.edu.cn).

F. Dufaux is with the Laboratoire des Signaux et Systèmes, CNRSCentraleSupelec- Université Paris-Sud, 91192 Gif-sur-Yvette, France (e-mail: frederic.dufaux@l2s.centralesupelec.fr.

Manuscript received Feb 1, 2019

I. Introduction

With the rapid development of science and technology, video applications have been widely used in our daily life. Various video applications, such as digital TV broadcasting, video conferencing, wireless video streaming, and smart phone communications, have become ubiquitous. These applications are also characterized by heterogeneous environments with a wide variety of devices. These devices may have different screen resolutions, processing capabilities, and network bandwidth requirements. In addition, different networks, such as broadband networks and wireless networks, may have different characteristics. Even within the same network, network bandwidths may also vary in different conditions. Consequently, video streaming should be able to adapt to different devices and network bandwidths. Scalable Video Coding (SVC), an extension of H.264/AVC, can efficiently solve these problems [1]. In order to provide scalability, SVC consists of a base layer (BL) and one or more enhancement layers (ELs). A SVC stream may support temporal, spatial and quality scalability, or a combination thereof. Scalability in temporal, spatial and quality refer to frame rate from low to high, spatial resolution from low to high and SNR from low to high, respectively. In recent years, the newest video coding standard, HEVC, has been developed. Since HEVC uses more advanced features and higher-efficiency coding tools, it has better coding efficiency than previous video coding standards. HEVC only uses about half the bit rate in comparison with H.264/AVC, while maintaining the same video quality level [2]. Since HEVC has very high coding efficiency, it attracts extensive attention and has wide applications. In order to accommodate heterogeneous device capabilities, network conditions and client applications, Scalable High Efficiency Video Coding (SHVC), an extension of HEVC, has been proposed. Similar to SVC, SHVC also supports scalability in temporal, spatial, and quality. In addition, it also provides bit-depth scalability, and color gamut scalability [3].

HEVC achieves very high coding efficiency, but at the cost of a very high encoding complexity, with an increase by about two to four times when compared to H.264/AVC [2]. In SHVC, multiple layers need to be encoded using inter-layer prediction, in addition to the same encoding process as that of HEVC for

each layer. Thus, SHVC is even more complex than HEVC. Therefore, it is highly desirable to improve the coding speed, especially for wireless and real-time applications.

For this purpose, a new algorithm to improve the coding speed of inter prediction for quality SHVC (QS) is proposed in this work. The contributions of the proposed algorithm are: (1) inter-layer correlation is combined with spatial correlation and its correlation degree in prediction; (2) inter-layer reference (ILR) mode and merge mode are predicted based on depth correlation and residual distribution; (3) Square partitions, including ILR mode, merge mode and inter2Nx2N, are early terminated based on their RD-cost differences; (4) Non-square partitions are predicted based on the difference in expected values of the residual coefficients of square partitions; (5) inter-layer and spatial correlations are combined with residual distribution to predict depth early termination.

The remainder of this paper is organized as follows. Section II discusses related work. Section III presents an overview of our proposed algorithm. Section IV, Section V and Section VI describe the proposed depth prediction, mode prediction and depth early termination in details. Experimental results are presented in Section VII and we draw conclusions in Section VIII.

II. RELATED WORK

Since both SHVC and H.264/SVC are related to scalable video coding, they share many similarities. Besides, as SHVC is the scalable extension of HEVC, they certainly have close relationship. Therefore, we first discuss fast algorithms for H.264/SVC, then for HEVC, and finally for SHVC.

To improve the coding speed for H.264/SVC, the authors of [4-6] use inter-layer correlation to predict candidate macroblock (MB) modes and skip unlikely modes in EL. Kim et al. [7] use both inter-layer and spatial correlations to predict the SKIP mode of the current MB in EL. The research in [8] statistically derive the Rate-Distortion (RD) cost expectations of each mode, and encode the modes according to this expectation from small to large. Then, the RD cost is compared with a preset threshold to determine whether to early terminate the coding process. Yeh et al. [9] use Bayesian theorem to predict the best mode of the current MB. If the best mode cannot be obtained by the Bayesian theorem, the Markov process is then used to further estimate and refine the best mode. Jung et al. [10] propose a fast mode decision algorithm for SVC which uses All-Zero Block (AZB) detection. Based on empirical analysis of inter-layer and spatial correlations of an AZB, a prediction is made of whether a MB in the EL would be an AZB. Then, the AZB detection algorithm is adopted to examine and terminate the predicted MBs. Zhao et al. [11] initialize the candidate mode list based on inter-layer and spatial correlations, and then early terminate the coding process by a constrained model with optimal stopping. Based on motion activities, picture detail levels, inter-layer and spatial correlations, Lu et al. [12] predict candidate modes and early terminate the encoding process. According to inter-layer and spatial correlations, the work in [13] estimates likely modes, skips unlikely modes, and then early terminates the coding process based on the RD cost and residual coefficients. Generally speaking, these algorithms use correlations to either predict candidate modes or use early termination, or jointly exploit mode prediction and early termination to improve the coding speed.

To reduce the coding complexity of HEVC, several methods have been proposed. Pan et al. [14] uses the mode of depth correlation, AZB, and motion estimation information to early detect merge mode. The authors of [15] divide modes into symmetric motion partition (SMP) modes and asymmetric motion partition (AMP) modes, and then propose three optimization techniques to predict SMP modes and AMP modes. The research in [16] predicts candidate depths based on the previous frame and neighboring CUs, and then early terminate unnecessary CUs according to motion homogeneity checking, RD cost checking and SKIP mode checking. Zhang et al. [17] combine coding features with a weighted support vector machine (SVM), and jointly optimize the complexity allocation at CU level with given rate-distortion (RD) cost constraints, to predict candidate depths. Chen et al. [18] propose an adaptive early termination decision for CU and PU, and also develop an adaptive search range determination according to the motion vector (MV), to reduce computational demand.

Some research focus on improving the coding speed of SHVC. Tohidypour et al. use the RD cost of the neighboring blocks of current block in EL, RD costs of the corresponding block and its four neighboring blocks in BL, to predict the RD cost of the to-be-encoded CU in EL [19-20]. They use the neighboring blocks' modes of current block in EL and the co-located block's mode of current CU in BL to predict current block's candidate modes in EL for quality SHVC [21]. The research in [22] uses spatial and inter-layer correlations to propose a probabilistic approach to predict the quad-tree structure of coding tree units (CTUs) for quality and spatial scalable SHVC. In [23], when at least one of the parent CUs in previous depth layers of the current CU select the merge mode, the depth and scalable layer of the current CU in the EL are jointly used to predict candidate modes. Moreover, inter-layer and spatial correlations are also exploited to further eliminate candidate modes. In this manner, the coding speed can be improved. Tohidypour et al. use coding information of relative CUs including neighboring CU in EL and the co-located CU in BL to build a probabilistic model, and then uses this model to predict the candidate modes of the current CU in EL to improve the coding speed [24]. Wang et al. [25] predict candidate depths. In each candidate depth, if ILR mode early termination condition is satisfied, intra mode prediction can be skipped; otherwise intra mode prediction needs to be checked. Afterwards, if depth early termination is satisfied, other depth levels can be directly skipped to improve the coding speed. Yeh et al. [26] propose a coding unit complexity (CUC)-based prediction method to improve the coding speed of HEVC-to-SHVC transcoding for quality scalability. Yeh et al. [27] use encoded Coding Unit (CU) sizes, prediction modes, motion vectors and Rate-Distortion Costs (RD-Costs) in BL and encoded CU sizes in EL to accelerate the coding speed of quality SHVC. Shen et al. [28] use conditional probability of a SKIP/Merge mode, motion activity and mode complexity

to describe the video content to predict the EL mode decision to improve the coding speed.

Although SHVC and SVC are very similar, their specific coding features are different. Algorithms developed for SVC may not suitable for SHVC or cannot significantly improve coding speed of SHVC. Since both SHVC and HEVC use the same advanced coding tools, the complexity reduction algorithms developed for HEVC can also be utilized for SHVC. As HEVC has only one coding layer, these algorithms developed for HEVC do not consider and exploit inter-layer correlations, and thus they cannot effectively improve SHVC's coding speed. In addition, the fast inter coding algorithms developed for SHVC only predict either CUs part or modes part. Since only one part is predicted, the coding speed improvement is limited. In addition, mode prediction is always predicted in the same approach without considering different features for different modes. It is obvious that using the same prediction method without any adaptation to different mode features is not always efficient, and thus, the improvement of coding speed is limited to some degree.

Based on the above considerations, in this paper, we propose a new fast depth and mode inter prediction for quality SHVC. In order to improve the coding speed, we divide the problem into the depth part and mode part. In the depth part, we propose depth prediction and depth early termination. In the mode part, we divide modes into three categories and separately develop their own methods for predictions. Experimental results demonstrate that our algorithm can significantly improve the coding speed with negligible losses in coding efficiency.

III. OVERVIEW OF THE PROPOSED MULTI-STRATEGIES FRAMEWORK

In order to improve coding speed and maintain coding efficiency for inter prediction in QS, we propose five strategies: Correlation-Based Depth Prediction (CBDP); Distribution-Based ILR and Merge Mode Decision (DB-IMMD); RD Cost Based Square Mode Early Termination (RCB-SMET); Expected Values Based Non-Square Mode Prediction (EVB-NSMP); and Correlation and Distribution Based Depth Early Termination (CDB-DET). The overview of the proposed algorithm is summarized in Fig.1. The left side of Fig.1 shows the five proposed strategies. First, depth candidates are predicted through CBDP. For the selected current depth candidates, DB-IMMD determines whether the best mode is in ILR and Merge Modes. In the affirmative, the other modes do not need to be checked; otherwise inter2Nx2N mode is checked and RCB-SMET is used to determine whether the best mode is in square modes. In the affirmative, the non-square modes do not need to be checked; otherwise EVB-NSMP is used to predict candidate non-square modes. After the depth has been checked, the residuals of the likelihood depth predicted by correlations are examined by CDB-DET to determine whether it is the best depth for early termination.

IV. CORRELATION-BASED DEPTH PREDICTION (CBDP)

Similar to HEVC, the maximum size of CU in SHVC is 64, and the range of depth levels is from 0 to 3. Since every depth contains the whole inter and intra mode prediction process, and the inter prediction includes eight partitions, the induced coding complexity is very high. In quality SHVC, a CU's content in EL and the co-located CU in BL are exactly the same, thus inter-layer correlation is very strong. In addition, since strong spatial correlations exist in natural video content, thus spatial correlation is also very strong. Therefore, we can use inter-layer and spatial correlations to predict candidate depths and skip depths with low likelihood to improve the coding speed. The relative CUs of the current CU, which are used to predict candidate depths of the current CU, are shown in Fig.2. C, L, U, UL and UR are the current, left, upper, upper-left and upper-right CUs in the EL respectively. Accordingly, BC, BL, BU, BUL and BUR are the collocated CUs of C, L, U, UL and UR in the BL, respectively.

In order to develop the fast inter prediction algorithm for quality SHVC, extensive experiments have been conducted to investigate related features and test corresponding methods. To meet the requirement of different resolutions, in our experiments, we use two sequences from class B (Tractor and Sunflower), class C (PartyScene and Flowervase), class D (RaceHorses and BlowingBubbles) and class E (Town and Parkruner). According to common SHM test conditions (CSTC) [29], we encode each test sequence under the Random Access (RA) main configuration. The QPs used for the BL are (26, 30, 34, 38), and the corresponding QPs used for the EL are (22, 26, 30, 34). Based on these experiments, we propose our efficient fast decision methods below.

A. Inter-Layer Correlation Prediction

As mentioned above, inter-layer correlation is very strong and can be used to predict candidate depths of current CU. Using the above experimental conditions, inter-layer correlation in depth between BL and EL is obtained by calculating proportions of the four depths used by C in EL given the depth used by BC in BL. The results are reported in Table I, where higher values indicate that the corresponding depths have higher possibilities to be used, namely stronger depth correlations.

In Table I, BC refers to a depth used by CU BC in BL, C refers to depths used by CU C in EL. From Table I, we can observe the following features:

(1)When BC in BL uses depth 0, C in EL has high probability to select depths 0 and 1, and has low probability to select depths 2 and 3.

(2)When BC in BL uses depth 1, C in EL has high probability to select depths 0, 1 and 2, and has low probability to select depth 3.

(3)When BC in BL uses depth 2, C in EL has high probability to select depths 2 and 3, and has low probability to select depths 0 and 1.

(4)When BC in BL uses depth 3, C in EL has high probability to select depths 2 and 3, and has low probability to select depths 0 and 1.

According to these observations, we can conclude that C in EL has high probability to select depth of BC and its neighboring depths, and other depths have low probability to

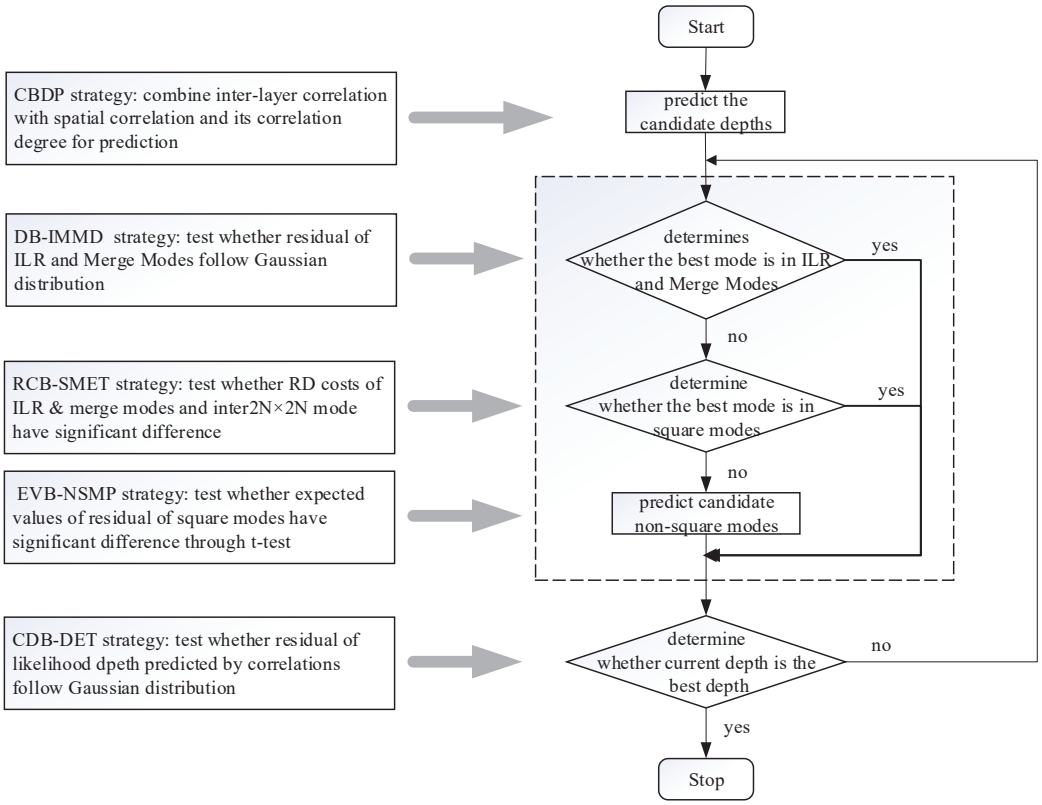


Fig. 1: Flowchart of the overall proposed algorithm.

TABLE I: Inter-layer depth distribution between BL and EL

BC	0				1				2				3			
C	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Sunflower	90	6	3	1	35	27	18	20	25	24	21	30	5	6	7	83
Tractor	52	24	15	10	27	28	27	18	11	18	30	42	4	8	15	73
Flowervase	81	10	5	3	9	74	10	6	8	13	25	53	3	3	14	81
PartyScene	66	12	9	12	20	47	29	4	1	6	21	71	1	1	7	91
BlowingBubble	48	22	16	14	26	27	38	9	1	2	52	46	1	1	7	91
RaceHorses	25	35	24	16	33	12	52	3	0	2	30	68	1	1	6	92
Parkrunner	55	12	16	17	27	16	43	15	1	1	60	38	0	1	14	85
Town	80	11	6	3	21	23	33	23	1	1	73	25	0	0	7	92
Average	62	16	12	10	25	32	31	12	6	8	39	47	2	3	10	85

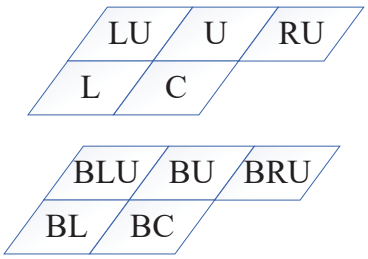


Fig. 2: Relative CUs of current CU.

be selected. Since a CU's content in EL and the co-located CU in BL are exactly the same, when a CU in BL uses a depth, the co-located CU in EL is more likely to use similar depths.

B. Spatial Correlation Prediction

Due to high similarity in motion and textural features for neighboring CUs, their spatial correlations are very strong. Therefore, we can predict the current CU's coding depth from its neighboring CUs'depths. Since the degrees of correlation between the current CU and its neighboring CUs may be vary, it is not appropriate to straightforwardly predict the current CU's depth from its neighboring CUs' depths. In this paper, we propose to first obtain spatial correlation degrees, then to combine neighboring CUs with their correlation degrees in prediction. The key point is how to obtain the correlation degrees between the current CU and its neighboring CUs in the EL. As the contents of CUs in the BL and the co-located CUs in the EL are the same, if two neighboring CUs' depths in BL are more similar, the co-located CUs' depths in EL will more likely be the same. To justify this, we tested the possibility of

TABLE II: Spatial Correlation under different correlation degrees

QPB, QPE	0	1	2	3
26, 22	82.0	61.0	48.0	40.0
30, 26	71.0	49.0	25.0	21.0
34, 30	64.0	35.0	25.0	19.0
38, 34	66.0	41.0	15.0	14.0
Average	70.75	46.5	28.25	23.5

the co-located CUs in EL having the same depth given by the absolute differences (0, 1, 2 and 3) between two neighboring CUs' depths in BL. The experimental results of the sequence "BlowingBubbles" are listed below.

In Table II, correlation degree refers to the absolute differences between two neighboring CUs' depths in BL, spatial correlation denotes as the possibility of the co-located CUs having the same depth in EL. QPB, QPE refer to QPs in BL and EL, respectively. 0, 1, 2 and 3 refer to the absolute differences between two neighboring CUs' depths in BL. From Table II, we can find that when the absolute differences between two neighboring CUs' depths in BL is 0, 1, 2 and 3 in BL, the average possibility of the co-located CUs having the same depth is 70.75%, 46.5%, 28.25% and 23.5% in EL, respectively. As a result, when neighboring CUs'spatial correlation degrees in the BL are very strong, then depths of the co-located CUs' depths in the EL should be very similar, and vice versa. Undoubtedly, spatial correlation degrees are inversely proportional to absolute depth differences among neighboring CUs in BL. We know that the maximum absolute difference of neighboring CUs in depth is 3. Therefore, the spatial correlation degrees d can be written as

$$d = 3 - |h_{bn} - h_{bc}| \tag{1}$$

where h_{bc} represents the depths of the co-located CUs of the current CU in BL, and h_{bn} represents the depth of one of its neighboring CUs in BL. The spatial correlation degree d is defined as spatial correlation weight. Through the above process, we can obtain each neighboring CU's spatial correlation weight. Then, we can calculate the weights of all depths by summing the weights with the same depths. Subsequently, we compute the probability of each depth by

$$p_h = \frac{w_h}{\sum_{h=0}^{3} w_h} \tag{2}$$

where h refers to the $h{-}th$ depth, w_h denotes the spatial weight of the $h-th$ depth, and p_h denotes the spatial probability of the $h-th$ depth. Obviously, the larger p_h is, the larger the probability of the $h-th$ depth to be selected.

C. Combine Spatial and Inter-Layer Correlations to Predict Candidate Depths

Through the above process, we can obtain the inter-layer depth distribution and the spatial probabilities of all depths. Since depth selection has strong relationship with both spatial and inter-layer correlations, we first predict depth likelihoods based on inter-layer correlation, and then predict depth likelihoods based on spatial correlation. If one or more depths are predicted to have a high probability to be selected based on inter-layer correlation, and they are also predicted to have a high probability to be selected based on spatial correlation, we can only check these depths and skip the other ones. The details are described below.

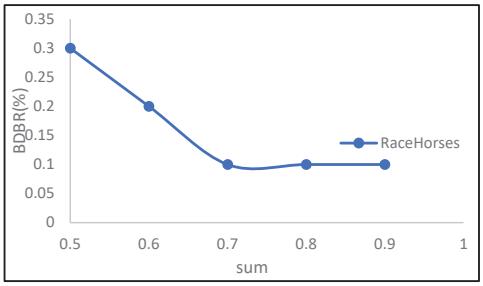


Fig. 3: Sum of probabilities of depth 0 and 1 and the corresponding coding efficiency loss.

As mentioned above, if BC use a depth, the co-located CU C in EL is very likely to select the depth and its neighboring depths. We can then combine their spatial probabilities to determine whether to only select these depths or to further check the other depths. For example, when BC uses depth 0, C in EL is very likely to use depths 0 and 1 based on inter-layer correlation. Then, we combine the spatial probabilities of depths 0 and 1 to decide whether further checking is needed. If depths 0 and 1 are very likely to be selected, the sum of their spatial probabilities should be at least equal to or greater than 0.5. Thus, we have selected 0.5, 0.6, 0.7, 0.8 and 0.9 under the above condition for experiments. From Table I, when BC in BL uses depth 0, C in EL has the largest sum of depths 2 and 3 in sequence "RaceHorses", so skipping depths 2 and 3 has the greatest impact for this sequence. In other words, if our method obtains very good coding efficiency for "RaceHorses", then we could definitely achieve better coding efficiency for the other sequences. Thus, through testing this sequence, we could obtain the optimal sum of the spatial probabilities for depth 0 and depth 1. The sum of their spatial probabilities and the corresponding coding efficiency losses are shown in Fig.3.

In Fig.3, the horizontal axis represents the sum of spatial probabilities of depth 0 and 1, and the vertical axis represents the coding efficiency loss denoted by BDBR, which represents the bitrate difference at an equal PSNR [30]. A negative BDBR represents an increase in coding efficiency when compared to the reference software. From Fig.3, we observe that there is a turning point when the sum is equal to 0.7. If the sum is smaller than 0.7, the corresponding coding efficiency changes dramatically. When sum is greater than or equal to 0.7, the coding efficiency stays approximately constant. Moreover, as we know, the smaller the sum is, the larger the encoding speed increases. Therefore, the sum is set to 0.7. The other depths are predicted in a similar way. Then, our depth prediction methods of C in EL are proposed as follows:

(1) If depth of BC is 0, go to (2); else if depth of BC is 1, go to (3); else if depth of BC is 2 and 3, go to (4).
(2) If two depths with the largest spatial probabilities are

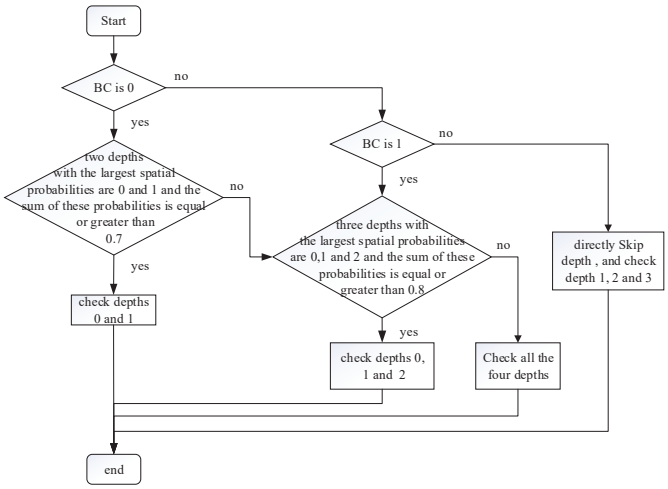


Fig. 4: Flowchart of the depth selection.

0 and 1 and the sum of these probabilities is equal or greater than 0.7, only these two depths are selected; else go to (3).

(3) If three depths with the largest spatial probabilities are 0, 1 and 2 and the sum of these probabilities is equal or greater than 0.8, only these three depths are selected; else all the four depths need to be checked.

(4) Since depth 0 is usually seldom used, when depth of BC is 2 and 3, the possibility of selecting depth 0 is further decreased. Depth 0 is directly skipped, and other three depths need to be checked.

The flowchart in Fig.4 illustrates the depth selection algorithm.

Since neighboring CUs may have different spatial correlation degrees, we combine neighboring CUs with spatial correlation degrees in prediction. In addition, we also integrate inter-layer correlation in prediction. Therefore, the proposed inter-prediction is very efficient.

V. PROPOSED SQUARE AND NON-SQUARE MODE DECISION

In every coding depth of SHVC, a CU is predicted by square and non-square modes. Square modes include ILR mode, merge mode, inter2Nx2N and intra2Nx2N mode. Non-square modes include 2NxN, Nx2N, 2NxnU, 2NxnD, nLx2N and nRx2N. As NxN is only used when depth is 3, we have not considered this mode for the sake of simplicity. More specifically, we have developed two methods to improve the coding speed for square and non-square modes, respectively.

A. Square partition mode prediction

For square partition, we have conducted experiments to obtain the proportions of the four modes (ILR mode, merge mode, inter2Nx2N mode, intra2Nx2N mode). Based on these experiments, the proportion of modes in square partition are listed in Table III.

TABLE III: The proportions of modes in square partition

Sequence	ILR and Merge	inter 2Nx2N	intra 2Nx2N
Sunflower	83.45	15.4	1.2
Tractor	86.1	12.9	1.0
Flowervase	79.0	19.9	1.1
PartyScene	84.5	14.2	1.3
BlowingBubble	82.9	16.5	0.6
RaceHorses	81.1	17.5	1.4
Parkrunner	87.5	12.1	0.4
Town	88.1	11.6	0.3
Average	84.1	15.0	0.9

TABLE IV: The depth correlation degree of ILR and merge modes

Sequence	Depth Correlation Degree
Sunflower	98.1
Tractor	95.4
Flowervase	97.8
PartyScene	94.7
BlowingBubble	95.4
RaceHorses	92.3
Parkrunner	96.0
Town	98.5
Average	96.0

From Table III, we can observe that: (1) ILR and merge modes have the highest probability; (2) inter2Nx2N has a smaller probability but remains non-negligible; (3) intra2Nx2N has a very small probability. Therefore, ILR and merge modes always need to be checked, and intra2Nx2N can be skipped directly. Regarding inter2Nx2N, if the mode is always checked, much unnecessary time would be spent. On the contrary, if the mode is skipped directly, the coding efficiency would be obviously degraded. Therefore, we first check ILR and merge modes, and decide whether they are the best modes. In the affirmative, we can only select these modes and skip inter2Nx2N mode. Otherwise, we still need to check inter2Nx2N mode and compare the difference of RD costs between ILR & merge modes and inter2Nx2N mode. If the difference is not significant, square modes provide a good prediction and non-square modes can be skipped directly. Otherwise, non-square modes need to be further checked. Base on the above analysis, in this paper, we develop ILR and merge modes prediction and square modes early termination. The details are presented below.

1) Distribution-Based ILR and Merge Modes Decision (DB-IMMD): Since their depths are strongly correlated, the parent CU and its children CUs may use similar modes[14]. In order to investigate the depth correlation degree between parent CU and its children CUs in ILR and merge modes, we have conducted experiments using the same conditions mentioned above. Based on these experiments, the depth correlation degrees of ILR and merge modes are listed in Table IV.

In Table IV, depth correlation degree refers to the probability that children CUs select ILR or merge modes given parent CU using ILR or merge modes. From Table IV, we can observe that the average depth correlation degree is 96%. In other words,

when a parent CU uses ILR or merge modes, its children CUs basically also use ILR or merge modes. However, depth correlation degrees are slightly lower in some cases, such as "RaceHorses" whose degree is only 92.3%. Consequently, when a parent CU uses ILR or merge modes, if its children CUs only check ILR and merge modes, the coding efficiency may degrade to some extent. In order to improve the coding speed and maintain coding efficiency, when a parent CU uses ILR or merge modes, we first encode ILR and merge modes in its children CUs, and then determine whether residual coefficients satisfy early termination condition. In the affirmative, we can skip the other modes. Otherwise, we have to further check other modes. As we know, residual coefficients normally follow a certain distribution if the prediction is accurate [31]. Thus, by analyzing the distribution of residual coefficients, we can decide if the early termination condition is met. Generally, a Gaussian distribution [32-33] or a Laplacian distribution [34] are normally used to model residual coefficients. Through our experiments, we found out that Gaussian distribution can better model residual coefficients, when compared to Laplacian distribution. Therefore, we adopt Gaussian distribution in this research. In [25], we proposed a test for intra prediction, which is based on skewness and kurtosis detection, to check whether residual coefficients follow the Gaussian distribution. Here, we apply this test to decide whether residual coefficients for inter prediction follow a Gaussian distribution.

Let x_1, x_2,...,x_n represent residual coefficients, the moment estimator G_1 for skewness and the moment estimator G_2 for kurtosis are obtained by

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

where B_k represents the sample central moment for order k (k=2,3,4). We can calculate B_k as

$$B_k = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^k}{n}, \tag{4}$$

where $\bar{x}$ denotes the average of x_i. G_1 and G_2 usually follow Gaussian distribution when n is relatively large (generally more than 100). Thus, we can approximated G_1 and G_2 by

$$\begin{aligned} G_1 &\sim N\left(\mu_1, \sigma_1^2\right) = N\left(0, \frac{6(n-2)}{(n+1)(n+3)}\right), \\ G_2 &\sim N\left(\mu_2, \sigma_2^2\right) = N\left(3 - \frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}\right), \end{aligned} \tag{5}$$

where σ_1 and σ_2 are the variances of G_1 and G_2 respectively, and μ_1 and μ_2 represent the expected values of G_1 and G_2. Then, we can decide if the residual coefficients follow the Gaussian distribution with expected values and variances according to

$$\mu_1 = \frac{G_1}{\sigma_1}, \mu_2 = \frac{G_2 - \mu_2}{\sigma_2}. \tag{6}$$

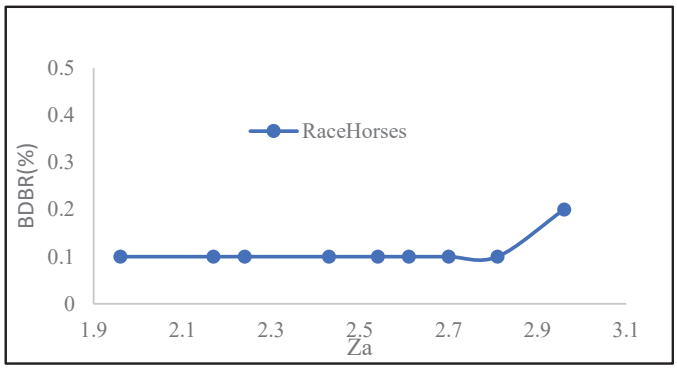


Fig. 5: z_α and the corresponding coding efficiency losses for ILR and merge modes.

Based on statistical hypothesis testing, a significance level α indicates the probability of falsely rejecting the null hypothesis, i.e. the distribution is Gaussian. Through examining the table of Gaussian distribution, we can obtain the corresponding critical test value z_α. The conditions of following Gaussian distribution are described by

$$|\mu_1| \leqslant z_\alpha, |\mu_2| \leqslant z_\alpha. \tag{7}$$

Residual coefficients are regarded to follow a Gaussian distribution if the two above conditions are met. In this case, our algorithm considers ILR or merge modes as the best modes, and then skips the other modes. Obtaining optimal α and the corresponding z_α is essential for effectively improving coding speed while maintaining coding efficiency. We observe in Table IV that "RaceHorses" has the lowest depth correlation degree. Therefore, if we terminate the prediction of ILR and merge modes early, we can obtain the greatest effect on "RaceHorses". As explained before, if we achieve good coding efficiency on "RaceHorses", we could obtain better coding performance on the other sequences. Therefore, we use "RaceHorses" in our experiments to find out the optimal values for optimal α and z_α. In our expriments, we select some commonly used values α, including 0.025, 0.015, 0.0125, 0.0075, 0.0055, 0.0045, 0.0035, 0.0025 and 0.0015. Correspondingly, the common values of z_α are 1.96, 2.17, 2.24, 2.43, 2.54, 2.61, 2.70, 2.81 and 2.96.

The resulting coding efficiency is shown in Fig. 5, where the horizontal axis represents the value of z_α, and the vertical axis indicates the coding efficiency loss, in terms of BDBR. It is obvious that there exists a turning point (z_α=2.81) in Fig. 5. We can observe that the corresponding coding efficiency losses are almost the same when z_α is less than or equal to 2.81. When z_α is greater than 2.81, the coding efficiency loss obviously increases. Besides, the greater z_α is, the larger the encoding speed increases. Based on the above analysis, z_α is set to 2.81 and α is set to 0.0025. Thus, when a parent CU uses ILR or merge mode, its children CUs are first encoded by ILR and merge modes. If their residual coefficients satisfy the early termination condition, the other modes can be skipped in these children CUs. Otherwise, the other modes need to be further checked.

2) RD Cost Based Square Modes Early Termination (RCB-SMET: When the condition of DB-IMMD is not satisfied or the parent CU of the current CU does not use ILR or merge modes, inter2Nx2N needs to be checked. As mentioned above, after all square modes have been checked, if their RD cost difference is not significant, it indicates that square modes provide a good prediction. Therefore, we only need to check square modes and skip non-square modes directly. In order to determine whether they have a significant difference, we can compare their quantized residual coefficients. If the differences of their quantized residual coefficients between ILR & merge modes, and inter2Nx2N mode are very small, then their RD costs are not significantly different [35]. Since every CU consists of multiple 4x4 blocks, we calculate quantized residual coefficients in a 4x4 block and derive the decision as follows. For a DCT coefficient w, its quantized value z is

$$z(i,j) = \frac{w(i,j)}{Q_{step}} + \frac{1}{6}, \tag{8}$$

where Q_{step} is the quantization step. As we know, if the condition (9) is met, no information needs to be coded,

$$\left| \frac{(w_1(i,j) - w_2(i,j))}{Q_{step}} \right| \leqslant 1, \tag{9}$$

where $w_1(i,j)$ is the DCT coefficients at the (i,j) position of ILR and merge modes, and $w_2(i,j)$ is the DCT coefficients at the (i,j) position of inter2Nx2N mode. Thus, we can consider that difference is not significant. As a result, we can derive

$$|w_1(i,j) - w_2(i,j)| \leqslant Q_{step}. \tag{10}$$

In SHVC, DCT coefficient $w(i,j)$ which is the DCT coefficients at the (i,j) position of a 4x4 block is computed as

$$w(i,j) = \frac{\sum_{u=0}^{3}\sum_{v=0}^{3} h(i,u)\, x(u,v)\, h(v,j)}{128 \times 128}. \tag{11}$$

where $h(i,u)$ refers to the value in the integer DCT transform matrix at the (i,u) position in SHVC, and $x(u,v)$ refers to the residual value in the residual matrix at the (u,v) position. According to (11), we can derive

$$\left|w_{1(x,j)} - w_{2(x,j)}\right| < \frac{\sum_{u=0}^{3}\sum_{v=0}^{3} |h(i,u)\, h(v,j)|\, |x_1(u,v) - x_2(u,v)|}{128 \times 128}, \tag{12}$$

In integer DCT transform of HEVC, $h(i,u)$ can be set to 64, 83 or 36, so the maximum value of $|h(i,u)|$ is 83. Combining (10) and (12), we obtain

$$\sum_{u=0}^{3}\sum_{v=0}^{3} |x_1(u,v) - x_2(u,v)| < 2.5 Q_{step}. \tag{13}$$

From (13), we can derive

$$\sum_{u=0}^{3}\sum_{v=0}^{3} |x_1(u,v)| - \sum_{u=0}^{3}\sum_{v=0}^{3} |x_2(u,v)| < 2.5 Q_{step}. \tag{14}$$

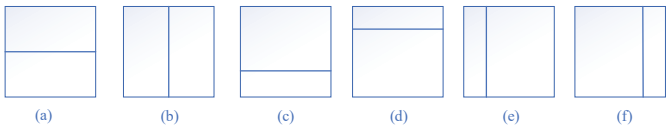


Fig. 6: Non-square partition

Sum of Absolute Difference (SAD) of two 4x4 blocks is obtained by (14) and can be written as

$$|SAD_1 - SAD_2| < 2.5 Q_{step}. \tag{15}$$

Since (15) is developed for 4x4 blocks, each 16x16 CU includes 16 4x4 blocks, we can derive

$$|SAD_1 - SAD_2| < 40 m Q_{step}. \tag{16}$$

Where m is the number of 16x16 CUs included in the current CU. Accordingly, each 32x32 CU includes 4 16x16 CUs, each 64x64 CU includes 16 16x16 CUs, thus m is set 16, 4, 1, corresponding to 64x64, 32x32, 16x16 CU, respectively. In order to maintain coding efficiency, we have not further tested 8x8 CUs.

Since SAD costs have strong relationship with RD costs, for convenience purposes, we have replaced SAD costs with RD costs, and (16) can be written as

$$|RD_1 - RD_2| < 40 m Q_{step}. \tag{17}$$

If the above condition is met, we can only select square modes and skip non-square modes directly. Otherwise, we still need to check non-square modes.

B. Expected Values Based Non-Square Mode Prediction (EVB-NSMP)

Non-square modes include inter2NxN, interNx2N, inter2NxnU, inter2NxnD, internLx2N, internRx2N, as shown in Fig. 6. In general, if the prediction is efficient, residual coefficients will obey a certain distribution. Therefore, we can also predict the best partition by studying the distribution of residual coefficients. As mentioned above, residual coefficients are modeled using a Gaussian distribution in this research. Suppose residual coefficients of a CU obey a Gaussian distribution, we can divide a CU into two parts and decide whether they have a significant difference to select the corresponding partition mode. More precisely, hypothesis testing is used to determine if the two parts for each division present a significant difference. For this purpose, t-test is used, which is the test statistics to investigate the significance of the difference between two sampled expected values. The expected values of the CU residual coefficients can be calculated by

$$\overline{X_1} = \frac{\sum_{u=0}^{3}\sum_{v=0}^{3} r_{ij}}{M \times N}, \tag{18}$$

where M and N are the width and height of the partitioned CU respectively, and r_{ij} is the residual value in the residual

matrix at the (i, j) position. For the sake of simplicity, we only consider CUs with sizes of 64x64, 32x32 and 16x16. According to the knowledge of mathematical statistics, if the residual coefficients of a CU, obeying a Gaussian distribution, are divided into two parts, these two parts will also obey a Gaussian distribution [36]. We can model the residual coefficients of two parts in each division as

$$X_1 \sim N\left(\mu_1, \sigma_1^2\right), X_2 \sim N\left(\mu_2, \sigma_2^2\right), \tag{19}$$

where X_1 and X_2 are two parts of a Gaussian distribution, μ and σ^2 are their respective expected value and variance.

We can compute the test statistics t by

$$t = \frac{\left(\overline{X_1} - \overline{X_2}\right) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{20}$$

where n_1 and n_2 are the number of coefficients in the first and second parts respectively, and S_w is calculated by the expression

$$S_w = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}, \tag{21}$$

where s_1^2 and s_2^2 are the variances of the first and second parts respectively. Assuming μ_1=μ_2, the condition can be expressed as

$$\left|\overline{X_1} - \overline{X_2}\right| \leqslant t_{\frac{\alpha}{2}}\, (n_1 + n_2 - 1)\, S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \tag{22}$$

where α is the significance level value. For any α, we can check T distribution table to obtain the corresponding threshold $t_{\frac{\alpha}{2}}\, (n_1 + n_2 - 1)$. For the sake of simplicity, we use a unified value for all six non-square modes. If the condition is satisfied, we can assume that there is no significant difference, and we do not need to check the corresponding modes. For example, in Fig. 6. (a), if top and bottom parts have a significant difference, then inter2NxN needs to be checked. Similarly, in Fig. 6. (b), if the left and right parts have a significant difference, then inter Nx2N need to be checked. The other modes are selected in the same way.

The optimal α and $t_{\frac{\alpha}{2}}\, (n_1 + n_2 - 1)$ can be obtained through experiments. We have performed experiments using the same conditions as above. Commonly used values for α are 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.02, 0.01 and 0.005, and the corresponding threshold values for $t_{\frac{\alpha}{2}}\, (n_1 + n_2 - 1)$ 0.674, 0.842, 1.036, 1.282, 1.645, 1.96, 2.326, 2.576 and 2.807, are tested.

In Fig.7, the horizontal axis represents the value of $t_{\frac{\alpha}{2}}\, (n_1 + n_2 - 1)$, abbreviated as t, and the vertical axis indicates the coding efficiency loss denoted by BDBR. From Fig. 7, we can observe that there are no obvious change rules in the loss of coding efficiency under different t, and all the losses of coding efficiencies are very small. Obviously, the higher the threshold is, the greater the coding speed improves. Therefore, the highest threshold value, 2.807, is selected.

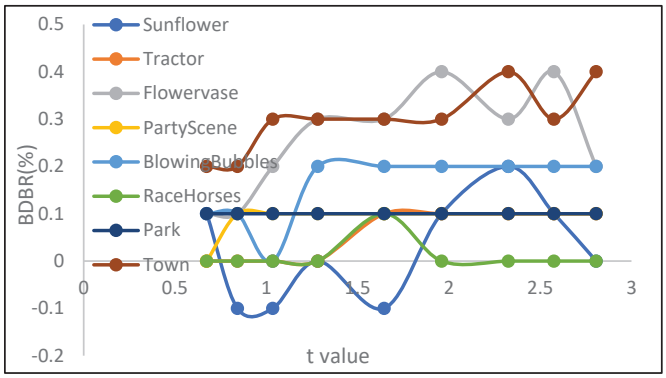


Fig. 7: t and the corresponding coding efficiency for Non-square mode prediction

VI. Correlation and Distribution Based Depth Early Termination (CDB-DET)

In the above described process, both inter-layer and spatial correlations are used to predict candidate depths and skip depths with low probabilities to improve the coding speed and maintain the coding efficiency. In each candidate depth, DB-IMMD, RCB-SMET and EVB-NSMP are used to further improve the coding speed. To this end, the residual coefficients of a candidate depth are used to test for depth early termination. As mentioned above, if the prediction is accurate, the residual coefficients will obey a Gaussian distribution. Since different depths may have different likelihoods to be selected, if residual coefficients are directly used in prediction, we may not obtain optimal thresholds that meet all conditions very well. Therefore, we combine both inter-layer and spatial correlations in prediction. When both inter-layer and spatial correlations predict a depth that has a high probability to be selected, we further use residual coefficients in prediction. Thus, we can improve the coding speed and maintain the coding efficiency. Based on the above analysis, we propose depth early termination as follows:

(1)If depth of BC is 0, go to (2); else if depth of BC is 1 go to (3).

(2)If two depths with the largest spatial probabilities are 0 and 1, these two depths have high probability to be selected. After depth 1 has already been checked, we use a test based on skewness and kurtosis detection in DB-IMMD for early termination. Using the same way as in DB-IMMD, z_α is set to 1.96. If the early termination condition with z_α (1.96) is satisfied, other depths can be skipped. Otherwise, go to (3).

(3)If three depths with the largest spatial probabilities are 0, 1 and 2, after depth 2 has already been checked, early termination with z_α (2.61) to test. If the early termination is satisfied, depth 3 can be skipped. Otherwise, depth 3 needs to be further checked.

The corresponding flowchart in Fig. 8 illustrates the process.

VII. Experimental Results

The proposed fast inter prediction algorithm for QS has been implemented in the SHVC reference software (SHM 11.0) on

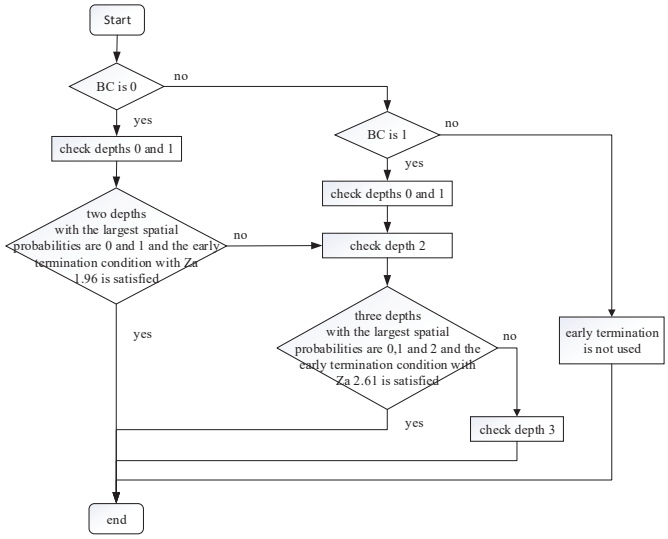


Fig. 8: Flowchart of the depth early termination

a server with Intel (R) 3.6 GHz CPU and 30 GB memory. The experimental parameters are set according to the CSTC [29]. The proposed algorithm is developed for RA coding structure. The QPs in the BL are set to (26, 30, 34, 38), and the corresponding QPs in the ELs are set to (22, 26, 30, 34) and (20, 24, 28, 32), respectively. We have used the eight training sequences (Sunflower, Tractor, Flowervase, PartyScene, BlowingBubbles, RaceHorses, Parkruner and Town) in the four proposed strategies: CBDP, DB-IMMD, EVBNSMP and CDB-DET. More precisely, these training sequences have been used to tune the proposed approach. This demonstrates that our algorithm performs well for various sequences with diverse video content.

The proposed algorithm includes five strategies: “CBDP”, “DB-IMMD”, “RCB-SMET”, “EVB-NSMP”, and "CDB-DET". It is very complicated to separately test them. Since the proposed algorithm is developed for depth and mode predictions, the five strategies are divided into depth and mode predictions. Since both “CBDP” and “CDB-DET” are related to the depth, we combine these two strategies, referred to as "Depth", in our performance evaluation. As “DB-IMMD”, “RCB-SMET", and “EVB-NSMP” are related to mode, we also integrate these three strategies in our experiments, and referred to it as "Mode". We have tested both "Depth" and "Mode" for the EL with (22, 26, 30, 34). We adopt BDBR to evaluate the coding efficiency loss. It measures the difference of coding efficiency between the proposed algorithm and the reference software. A negative BDBR indicates that our algorithm achieves a bitrate saving. We measure the coding time in EL, and "TS" represents the coding time savings (in percentage). Table V presents the performance of the different methods.

From Table V, we find that the average coding speed improvements in "Depth", "Mode" and "Non-Square mode" are 25.78% , 60.38% and 31.30%, respectively. The average coding efficiency losses in "Depth", "Mode" and "Non-Square Mode" are 0.14% ,1.04% and 0.83%, correspondingly.

TABLE V: Performance comparison of the different methods

Sequence	Depth		Mode		Non-Square Mode	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	0.1%	31.40%	1.0%	61.06%	0.9%	31.76%
PeopleOnStreet	0.1%	10.09%	0.9%	60.64%	0.6%	28.88%
Kimono	0.1%	32.23%	0.5%	60.43%	0.4%	25.60%
ParkScene	0.2%	26.77%	1.0%	61.02%	0.9%	33.37%
Cactus	0.2%	25.87%	0.9%	60.14%	0.6%	32.20%
BasketballDrive	0.2%	26.55%	1.2%	59.13%	0.8%	30.45%
BQTerrace	0.1%	27.57%	1.8%	60.26%	1.6%	36.86%
Average	0.14%	25.78%	1.04%	60.38%	0.83%	31.30%

TABLE VI: Performance comparison with Q1

Sequence	Proposed		Hamid [22]		Shen [28]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	1.2%	71.21%	1.3%	59.12%	1.5%	70.0%
PeopleOnStreet	1.0%	67.93%	0.97%	55.96%	1.9%	57.0%
Kimono	0.8%	72.34%	0.77%	62.23%	0.9%	64.0%
ParkScene	1.2%	72.98%	1.17%	65.79%	1.8%	69.0%
Cactus	1.2%	71.42%	0.96%	54.26%	1.3%	67.0%
BasketballDrive	1.5%	71.86%	1.35%	65.83%	2.1%	67.0%
BQTerrace	2.1%	70.26%	0.68%	61.16%	2.1%	71.0%
Average	1.27%	71.14%	1.13%	60.62%	1.67%	66.43%

TABLE VII: Performance comparison with Q2

Sequence	Proposed		Hamid [22]		Shen [28]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	2.5%	66.86%	2.2%	54.11%	3.50%	68.00%
PeopleOnStreet	2.2%	65.14%	2.4%	53.23%	2.40%	54.00%
Kimono	1.8%	67.92%	1.9%	56.17%	1.90%	60.00%
ParkScene	2.1%	68.87%	2.2%	60.74%	3.20%	67.00%
Cactus	2.1%	66.14%	2.0%	52.18%	2.20%	64.00%
BasketballDrive	2.3%	68.27%	2.6%	52.21%	3.30%	64.00%
BQTerrace	2.1%	68.86%	1.8%	59.63%	3.20%	68.00%
Average	2.21%	67.43%	2.18%	55.46%	2.81%	63.57%

After testing "Depth", "Mode" and "Non-Square Mode" methods, we further evaluate the overall performance of our proposed algorithm. We combine the five proposed strategies, i.e. CBDP, DB-IMMD, RCB-SMET, EVB-NSMP, and CDB-DET, and compare it with Hamid’s algorithm [22] as well as Shen’s algorithm [28]. For fair comparisons, all these three algorithms 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 71.14% and 67.43% for Q1 and Q2, respectively. The average BDBR increases by 1.27% and 2.21%, respectively. From these results, we can conclude that our algorithm can significantly improve the coding speed with small increase in coding efficiency loss. In addition, we can observe that the coding efficiency (BDBR) with Q2 is obviously degraded when compared to BDBR with Q1. This is because the QP difference between BL and EL is 6 in Table VII, whereas it is 4 in Table VI. The larger QP difference causes the weaker inter-

layer correlation. Therefore, coding efficiency is obviously degraded with Q2.

When compared with Hamid's algorithm, the running time of the proposed algorithm with Q1 is improved by 10.52% in EL, while the coding efficiency loss is increased by 0.14% in terms of BDBR. With Q2, the running time of the proposed algorithm is improved by 11.97% in EL, while the coding efficiency loss is increased by 0.03% in terms of BDBR, when compared with Hamid's algorithm. When compared with Shen's algorithm, the running time of the proposed algorithm with Q1 is improved by 4.71% in EL and the coding efficiency is increased by 0.4%; the running time of the proposed algorithm with Q2 is improved by 3.86% in EL and the coding efficiency is increased by 0.6%. Therefore, we can draw the following conclusion: (1) when compared with Hamid's algorithm, the proposed algorithm's coding speed is significant faster with negligible coding efficiency loss; (2) when compared with Shen's algorithm, the proposed algorithm's coding speed is faster with significant coding efficiency increase. All these results further demonstrate the robust performance of the proposed algorithm.

The BDBRs of "BQTerrace" and "Cactus" obtained by the proposed method are larger than those obtained by Hamid's method. They correspond to the two largest differences in BDBRs between both methods. Therefore, the RD curves of these two sequences for the proposed method, Hamid [22], and the original SHVC are presented in Fig. 9.

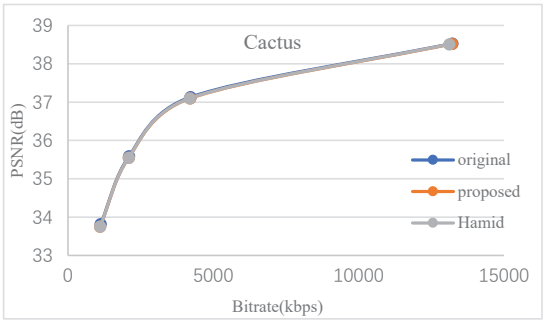
From Fig. 9, we observe that RD curves for "BQTerrace" and "Cactus" for the three methods are essentially overlapping one another. In other words, the proposed method does not introduce any noticeable coding efficiency loss when compared to Hamid and the original SHVC. This also holds for other sequences which have smaller BDBR differences among these three methods.

In order to demonstrate the subjective performance, the 15th decoded video frame of Cactus and BQTerrace are shown in Fig.10 and Fig.11, respectively.

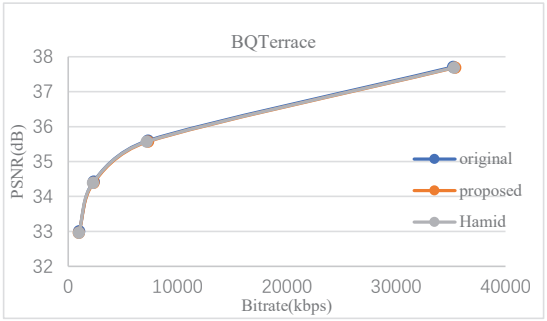
Fig. 10 demonstrates subjective performance of the 15th frame (Cactus) using the proposed method and Hamid's method. Since the BDBR between these two methods is very small, it is extremely difficult to notice a subjective difference. Similar observations are made in Fig. 11. Therefore, we can conclude that the decoded video frames obtained by the proposed method and Hamid method have no noticeable subjective differences.

The greatly reduced complexity of the proposed algorithm is mainly due to the following reasons: (1) skipping and early terminating depths with low likelihood, based on the inter-layer and spatial correlation and residual distribution; (2) combining depth correlation and residual distribution to skip inter2Nx2N and non-square modes; (3) skipping non-square modes through comparing the difference of square modes' RD costs, and (4) skipping non-square modes with low likelihoods through testing expected values of residual coefficients.

Experiments show that both coding efficiency and coding speed under RA are significantly better than those under low-delay (LD) coding structure. Indeed, due to the worse prediction of LD structure, its residual coefficients are larger than those of RA, which leads to more encoding times under



(a) RD curve for "Cactus"



(b) RD curve for "BQTerrace"

Fig. 9: RD curves for the proposed method, Hamid [22] and the original SHVC



(a) Proposed method



(b) Hamid's method

Fig. 10: Cactus - the 15th decoded frame



(a) Proposed method



(b) Hamid's method

Fig. 11: BQTerrace - the 15th decoded frame

LD. As we focus on high coding efficiency and low encoding complexity in this paper, we will not further pursue the fast schemes suitable for LD structure.

VIII. Conclusion

In this paper, we have proposed a novel and effective inter prediction algorithm for Quality SHVC. The algorithm mainly includes depth and mode predictions. In depth prediction, the inter-layer and spatial correlations are combined to predict candidate depths and skip depths with low likelihoods. Then depths are further early terminated through combing correlations with residual distribution. In mode prediction, we first encode ILR and merge modes, and then combine depth correlation and residual distribution to skip inter2Nx2N and non-square modes. Second, we encode inter2Nx2N, and then compare RD costs of square modes to skip non-square modes. Third, we test residual coefficients to skip non-square modes with low probabilities. Since many strategies have been proposed, the proposed algorithm can significantly improve the coding speed with small encoding efficiency losses

In addition, the coding speed of depth prediction is not very high, but its coding efficiency loss is also very small. In order to improve the coding speed, we could select larger thresholds, but then its coding efficiency loss would also increase. Indeed, there is a tradeoff between coding efficiency and coding speed, and we can select the most suitable thresholds according to application requirements. The proposed algorithm can be further improved by combining with other methods [37-39]. Machine learning, especially deep learning, is attracting significant research interest, we will use these technologies to further improve the coding speed of SHVC in the future.

References

[1] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] G.Correa, P.Assuncao, L.Agostini and L.S.Cruz, "Performance and Computational Complexity Assessment of High Efficiency Video Encoders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1899-1909, Dec. 2012.

[3] 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 Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20-34, Jan. 2016.

[4] H. Li, Z. G. Li, C. Y. Wen and L. P. Chau, "Fast mode decision for spatial scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20-34, Jan. 2016.

[5] H. Li, Z. G. Li and C. Wen, "Fast mode decision algorithm for interframe coding in fully scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 889–895, Jul. 2006.

[6] H. C. Lin, W. H. Peng and H. M. Hang, "Fast mode decision algorithm for interframe coding in fully scalable video coding," *in Proc. IEEE Int. Conf. Multimedia Expo.*, Hannover, Germany, Jun. 2008, pp. 765–768.

[7] L. Shen, Y. Sun, Z. Liu and Z. Zhang, "Efficient skip mode detection for coarse grain quality scalable video coding," *IEEE Signal Process. Lett.*, vol. 17, no. 10, pp.887–890, Oct. 2010.

[8] C. S. Park, B. K. Dan, H. Choi and S. J. Ko, "A statistical approach or fast mode decision in scalable video coding," IEEE Trans. Circuits Syst. Video Technol.," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1915–1920, Dec. 2009.

[9] C. H. Yeh, K. J. Fan, M. J. Chen and G.L. Li, "Fast mode decision algorithm for scalable video coding using Bayesian theorem detection and Markov process," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 563–574, Apr. 2010.

[10] 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 Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 201-206, Feb. 2010.

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

[12] X. Lu and G. R. Martin, "Fast mode decision algorithm for the H.264/AVC scalable video coding extension," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 23, no. 1, pp. 846-855, Jan. 2013.

[13] 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," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1631–1639, Jul. 2014.

[14] Z. Pan, S. Kwong, M. T. Sun and J. Lei, "Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 405–412, Jun. 2014.

[15] J. Vanne, M. Viitanen and T. D. Hamalainen, "Efficient mode decision schemes for HEVC inter prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1579-1593, Sep. 2014.

[16] L. Shen, Z. Liu, X. Zhang, W. Zhao and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia.*, vol. 15, no. 2, pp. 465-470, Feb. 2013.

[17] Y. Zhang, S. K wong, 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 Trans. Image Process.*, vol. 24, no. 7, pp. 2225-2238, Jul. 2015.

[18] M. J. Chen, Y. D. Wu, C. H. Yeh, K. M. Lin and S. D. Lin, "Efficient CU and PU Decision Based on Motion Information for inter prediction of HEVC," *IEEE Trans. Industrial Informatics.*, vol.14, no. 11, pp. 4735-4745, Nov. 2018.

[19] H. R. Tohidypour, M. T. Pourazad and P. Nasiopoulos, "Content adaptive complexity reduction scheme for quality/fidelity scalable HEVC," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP).*, Vancouver, BC, Canada, May 2013, pp. 1744-1748.

[20] H. R. Tohidypour, M. T. Pourazad and P. Nasiopoulos, "Content Adaptive Complexity Reduction Scheme for Quality/Fidelity Scalable HEVC," *document JCTVC-L0042, ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11.*, Geneva, Switzerland, Jan. 2013.

[21] H. R. Tohidypour, H. Bashashati, M. T. Pourazad and P. Nasiopoulos, "Fast mode assignment for quality scalable extension of the High Efficiency Video Coding (HEVC) standard: A Bayesian approach," *in Proc. 6th Balkan Conf. Inf. (BCI).*, Thessaloniki, Greece, Sep. 2013, pp. 61-65.
[22] 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 Trans. Multimedia.*, vol. 18, no. 2, pp. 182–195, Feb. 2016.
[23] H. R. Tohidypour, M. T. Pourazad and P. Nasiopoulos, "An encoder complexity reduction scheme for qual-ity/fidelity scalable HEVC," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 664–674, Sep. 2016.
[24] H. R. Tohidypour, H. Bashashati, M. T. Pourazad and P. Nasiopoulos, "Online-Learning-Based Mode Prediction Method for Quality Scalable Extension of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2204-2215, Oct. 2017.
[25] D. Wang, C. Zhu, Y. Sun, F. Dufaux and Y. Huang, "Efficient Multi-Strategy intra Prediction for Quality Scalable High Efficiency Video Coding," *IEEE Trans. Image Process.*, vol. 28, no.4, pp. 2063- 2074, Apr. 2019.
[26] C. Yeh, W. Tseng, L. Kang, C. Lee, K. Muchtard, M. Chen, "Coding unit complexity-based predictions of coding unit depth and prediction unit mode for efficient HEVC-to-SHVC transcoding with quality scalability," *J Vis Commun Image.*, vol. 55, pp. 342-351, Aug, 2018.
[27] C. Yeh, J. Lin, M. Chen, C. Yeh, C. Lee, K. Tai, "Fast prediction for quality scalability of High Efficiency Video Coding Scalable Extension," *J Vis Commun Image.*, vol. 58, pp. 462-476, Jan, 2019.
[28] L. Shen, G. Feng, "Content-based adaptive SHVC mode decision algorithm," *IEEE Trans. Multimedia.*, DOI:10.1109/TMM.2019.2909859.
[29] 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.
[30] G.Bjontegaard, "Calculation of average PSNR difference between RD-curves," *13th VCEG-M33 Meeting.*, Austin, TX, Apr.2-4, 2001.
[31] H. L. Wang and S. Kwong, "Hybrid model to detect zero quantized DCT coeffcients in H.264," *IEEE Trans. Multimedia.*, vol. 9, no. 4, pp. 728–735, Jun. 2007.
[32] S. Cho and M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1555–1564, Sep. 2013.
[33] N. Hu and E.-H. Yang, "Fast motion estimation based on confidence interval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1310–1322, Aug. 2014.
[34] 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 Process. Lett.*, vol. 18, no. 10, pp. 571-574, Oct. 2011.
[35] J. Ren, N. Kehtarnavaz and M. Budagavi, "Computationally efficient mode selection in H.264/AVC video coding," *IEEE Trans. Consum. Electr.*, vol. 54, no. 2, pp. 877-886, May. 2008.
[36] B. Lee, M. Kim, S. Hahm, I. J. Cho, and C. Park, "A low complexity encoding scheme for coarse grain scalable video coding," *in Proc. IEEE Int. Conf. Vis. Inf. Eng.*, vol. 1. Xian, China, Jul./Aug. 2008, pp. 753–758.
[37] Ce Zhu, Xiao Lin, L. Chau and Lai-Man Po, "Enhanced hexagonal search for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp.1210-1214, Oct. 2004.
[38] Bing Xiong and Ce Zhu,"A New Multiplication-Free Block Matching Criterion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no.10, pp. 1441-1446, Oct. 2008.
[39] Ce Zhu and Bing Xiong,"Transform-Exempted Calculation of Sum of Absolute Hadamard Transformed Differences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1183-1188, Aug. 2009.



Dayong Wang Dayong Wang received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China in 2010. He was a Lecturer with the Hubei University of Arts and Science from 2010 to 2012. He held a postdoctoral research position with the Graduate School, Shenzhen, Tsinghua University, from 2012 to 2015. He is currently an Associate Professor with the Chongqing University of Posts and Telecommunications, and also a Post-Doctoral Fellow with the University of Electronic Science and Technology of China. His main research interest is video coding.



Yu Sun Yu Sun received the B.S. and M.S. degrees in computer science from the University of Electronic Science and Technology of China, and the Ph.D. degree in computer science and engineering from the University of Texas at Arlington in 2004. She is currently a Professor with the Department of Computer Science, the University of Central Arkansas, USA. Her main research interests include video compression, multimedia communication, and image processing. Her research work in these areas has been published in over 70 technical papers in refereed journals and conferences. She is a member of the IEEE Multimedia Communications Technical Committee and was a Co-Editor of the IEEE Distributed Systems Online. She has served as a session chair, a member of organizing committees and technical program committees for over 55 international conferences.



Ce Zhu Ce Zhu (M'03–SM'04–F'17) received the B.S. degree from Sichuan University, Chengdu, China, in 1989, and the M.Eng and Ph.D. degrees from Southeast University, Nanjing, China, in 1992 and 1994, respectively, all in electronic and information engineering. He held a post-doctoral research position with the Chinese University of Hong Kong in 1995, the City University of Hong Kong, and the University of Melbourne, Australia, from 1996 to 1998. He was with Nanyang Technological University, Singapore, for 14 years from 1998 to 2012, where he was a Research Fellow, a Program Manager, an Assistant Professor, and then promoted to an Associate Professor in 2005. He has been with University of Electronic Science and Technology of China, Chengdu, China, as a professor since 2012. His research interests include video coding and communications, video analysis and processing, 3D video, visual perception and applications. He has served on the editorial boards of a few journals, including as an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON BROADCASTING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SIGNAL PROCESSING LETTERS, an Editor of IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and an Area Editor of SIGNAL PROCESSING: IMAGE COMMUNICATION. He has also served as a Guest Editor of four special issues in international journals, including as a Guest Editor for an ongoing special issue of "Perception-driven 360-degree Video Processing" in the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is an IEEE Distinguished Lecturer of Circuits and Systems Society (2019-2020) and of Broadcast Technology Society (2012-).



Weisheng Li Weisheng Li is currently a professor with the Chongqing Key Laboratory of Image Cognition, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China. He received his PhD degree in computer application technology from Xidian University in 2004. His research focuses on intelligent information processing and pattern recognition.



Frederic Dufaux (S'93-M'95-SM'09-F'16) is a CNRS Research Director at Laboratoire des Signaux et Systèmes (L2S, UMR 8506), CNRS-CentraleSupelec-Université Paris-Sud, where he is head of the Telecom and Networking division. He is also Editor-in-Chief of Signal Processing: Image Communication.

Frederic received his M.Sc. in physics and Ph.D. in electrical engineering from EPFL in 1990 and 1994 respectively. He has over 20 years of experience in research, previously holding positions at EPFL, Emitall Surveillance, Genimedia, Compaq, Digital Equipment, and MIT.

Frederic is a Fellow of IEEE. He is Chair of the IEEE SPS Multimedia Signal Processing (MMSP) Technical Committee and Chair of the EURASIP Special Area Team on Visual Information Processing. He has been involved in the standardization of digital video and imaging technologies, participating both in the MPEG and JPEG committees. He is the recipient of two ISO awards for his contributions.

His research interests include image and video coding, 3D video, high dynamic range imaging, visual quality assessment, and video transmission over wireless network. He is author or co-author of 3 books ("High Dynamic Range Video", "Digital Holographic Data Representation and Compression", "Emerging Technologies for 3D Video"), more than 200 research publications and 20 patents issued or pending.