Hybrid Strategies for Efficient Intra Prediction in Spatial SHVC

Dayong Wang, Yu Sun, Weisheng Li, Lele Xie, Xin Lu, Frédéric Dufaux, Ce Zhu

Hybrid Strategies for Efficient Intra Prediction in Spatial SHVC

Dayong Wang, Yu Sun, Weisheng Li, *Member, IEEE,* Lele Xie, Xin Lu, Frederic Dufaux, *Fellow, IEEE,* Ce Zhu, *Fellow, IEEE*

***Abstract*—With multi-layer encoding and Inter-layer prediction, Spatial Scalable High Efficiency Video Coding (SSHVC) has extremely high coding complexity. It is very crucial to speed up its coding to promote widespread and cost-effective SSHVC applications. Specifically, we first reveal that the average RD cost of Inter-layer Reference (ILR) mode is different from that of Intra mode, but they both follow the Gaussian distribution. Based on this discovery, we apply the classic Gaussian Mixture Model and Expectation Maximization to determine whether ILR mode is the best mode thus skipping Intra mode. Second, when coding units (CUs) in enhancement layer use Intra mode, it indicates very simple texture is presented. We investigate their Directional Mode (DM) distribution, and divide all DMs into three classes, and then develop different methods with respect to classes to progressively predict the best DMs. Third, by jointly considering rate distortion costs, residual coefficients and neighboring CUs, we propose to employ the Conditional Random Fields model to early terminate depth selection. Experimental results demonstrate that the proposed algorithm can significantly improve coding speed with negligible coding efficiency losses.**

***Index Terms*—SHVC, Coding depth, ILR mode, Intra mode, Directional Mode.**

I. INTRODUCTION

This work was supported in part by the National Natural Science Foundation of China under Grants U19A2052, 62020106011, 61972060 and 62027827; in part by the National Key Research and Development Program of China under Grant 2019YFE0110800; in part by Natural Science Foundation of Chongqing under Grants cstc2020jcyj-msxmX0766, cstc2020jcyj-zdxmX0025 and cstc2019cxcyljrc-td0270; in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K202100604; in part by the Jiangxi Provincial Natural Science Foundation under Grant 20202BABL202006; and in part by Shangrao Basic Research Program under Grant 2021F003. *(Corresponding author: Weisheng Li.)*

D. Wang is with the Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the School of Mathematics and Computer Science, Shangrao Normal University, Shangrao 334001, 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).

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).

L. Xie is with School of Information and Communication Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: 1435439126@qq.com).

X. Lu is with the School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, United Kingdom (e-mail: xin.lu@dmu.ac.uk).

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).

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).

VIDEO applications, such as digital TV broadcasting, video conferencing, wireless video streaming, and smart phone communications, are deeply embedded in our daily life. Increasingly, more terminal devices with various spatial resolutions are constantly emerging. This poses the requirements for video streams to be adaptive to different display resolutions. As the latest scalable video coding standard, Spatial Scalable High Efficiency Video Coding (SSHVC) provides an efficient solution to keep up with the requirements, which encodes video sequences layer by layer at different spatial resolutions. At the decoding end, appropriate layers are extracted to adapt to various display resolutions on demand.

The bitstream generated by SSHVC consists of a base layer (BL) and one or more enhancement layers (ELs). BL frames are encoded using intra-layer prediction only, while the additional inter-layer prediction tools are employed to encode the EL frames. The intra-layer prediction process in SSHVC is identical to that in HEVC. As the picture contents of the BL and ELs are generated from the same video source at different resolutions, the coding unit (CU) in the BL can be up-sampled to predict the co-located CU in the ELs. This prediction mechanism is denoted as Inter-layer prediction, and the prediction mode is called Inter-layer Reference (ILR) mode. HEVC encoder is a single layer encoder, nevertheless, its computational complexity is quite high. As a multi-layer encoder, SSHVC encodes video sequences at several layers. The complexity of an SSHVC encoder is much higher than that of an HEVC encoder. This seriously restricts the wide application of SSHVC, especially in wireless and real-time application environments. Consequently, it is crucial to develop fast coding algorithms for SSHVC to speed up its encoding process.

For this purpose, we propose efficient hybrid strategies for Intra-prediction in SSHVC. The major novelties and contributions of the proposed algorithm include:

(1) We find that the rate distortion (RD) costs of both ILR modes and Intra modes follow the Gaussian distribution. However, there is a significant divergence between respective RD costs. Consequently, the classical Gaussian Mixture Model and Expectation Maximization (GMM-EM) in machine learning are employed in our proposed algorithm to early determine whether the ILR mode is the best mode.

(2) We reveal a special directional mode (DM) distribution when a CU uses Intra mode as the best mode in EL. Based on this distribution, we propose DM Distribution-Based Progressive DM Selection (DD-BPDS) to predict

the best DM. More specifically, we divide all DMs into three classes and propose different strategies for different classes to progressively predict the best DM: significant difference-based DM early termination, significant difference-based DM prediction-descent direction search, and variable stepsize check-binary search.

(3) As coding depths of neighbouring CUs, residual coefficients, and RD costs provides prior-knowledge for the coding depth selection, we propose a Conditional Random Fields-Based Depth Early Termination (CRF-BDET) algorithm. Specifically, the coding depths of neighbouring CUs, residual coefficients, and RD costs are considered with the CRF model in machine learning to determine the best coding depth, thus early terminating the depth selection.

To the best of our knowledge, the three contributions above have never been investigated. Furthermore, GMM-EM, DD-BPDS and CRF-BDET have never been suggested for fast video coding implementations.

The rest of this paper is organized as follows. Section II discusses related work. Section III provides the overview of the proposed algorithm. Section IV presents the proposed algorithm in detail, including three fast strategies to accelerate the encoding process. Section V discusses and analyses the experimental results. Finally, Section VI draws the conclusions of this research and plans for future work.

II. RELATED WORK

In SHVC [1], four coding depths from 0 to 3 are evaluated for each CTU. The CU sizes at coding depth 0, 1 and 2 correspond to 64×64, 32×32, 16×16, respectively. For coding depth 3, there are two CU sizes available, namely 8×8 and 4×4. For each CU, the encoder needs to evaluate both the ILR modes and Intra modes which also includes 35 DMs. In order to determine the best coding parameters, SHVC evaluates all of the depths and modes. The exhaustive evaluation enables SHVC to achieve high coding efficiency at the cost of significantly increased computational complexity. However, doing so can result in a very complicated coding process. In order to accelerate the encoding process, a number of algorithms have been developed targeting at the mode prediction, DM prediction and depth prediction in SHVC, which are reviewed and discussed below.

(1) Mode prediction

As the current CU and its relevant CUs contain similar picture contents, relevant CUs can be used to predict the most-likely coding modes of the current CU. Tohidypour et al. [2] predict the RD cost of the current CU by using those of it's relevant CUs to achieve an early termination. To reduce the computational complexity, the algorithm proposed in [3] adopts relevant CUs to predict the likely modes and skips unlikely modes in EL. Based on the depth and mode of the co-located CU in BL, the algorithm proposed in [4] first predictes the likely modes for the current CU in EL, and then further eliminates unlikely modes using Inter-layer and spatial correlations. Wang et al. [5-7] first checke the ILR mode and merge mode, and then compare the difference of their RD costs to early terminate the mode selection process. The correlations form the basis for the above algorithms.

Generally speaking, the residual coefficients are very small, and follow the Laplacian distribution [9] or Gaussian distribution [10] if the mode is accurately predicted. Wang et al. [7] first check ILR mode, and then calculate its part-zero block based on the distribution of its residual coefficients to early terminate mode selection. Wang et al. [8] first check ILR mode, and then decide whether its residual coefficients follow Gaussian distribution so as to early skip Intra mode. Wang et al. [11] combine ILR mode probabilities with its residual coefficients to skip Intra mode prediction. Pan et al. [12] combine depth correlation and all-zero block to early terminate mode selection.

(2) DM prediction

Intra mode prediction is a very time-consuming encoding process in SSHVC. The fast DM selection algorithms that skip the evaluations of unlikely DMs provide an effective solution to speed up the coding process. As textural features are closely related to DM selection, they are often used for fast DM prediction. Zhao et al. [13] use a Sobel operator to predict more likely DMs and skip unlikely ones to reduce encoding time. Zhang et al. [14] obtain the average gradients in both the horizontal (AGH) and vertical directions (AGV) and then calculate the ratio values of AGH/AGV to predict the likely DMs. Based on the improved pixels of CUs, neighbouring blocks and the Sum of Absolute Hadamard Transformed Difference (SATD) costs, the work in [15] predictes the best DM candidate and skip some DMs to speed up the coding process.

Hadamard Costs (HCs) are also strongly related to DM selection and are usually used to predict candidate DMs. HC-based progressive rough mode search is developed in [16] to check likely DMs, thus skipping unlikely ones to speed up the coding process. The work described in [7] integrates textural features with the relationship between DMs and their corresponding HC values to predict candidate DMs and skip unlikely ones. Wang et al. [8] combine relevant CUs with DMs and their corresponding HC values to predict candidate DMs and skip unlikely DMs. Wang et al. [11] adopte the difference of Hadamard Costs (HCs) of typical DMs to predict candidate DMs, and then combine percentages of gradient amplitudes with HCs to early terminate DM selection. Jamali et al. [17] develop an RDO cost statistical model to predict likely DMs.

(3) Depth prediction

In order to speed up the coding process, texture features, correlations, residual coefficients and RD costs are usually exploited to avoid evaluations on unnecessary coding depths.

Generally speaking, CUs containing simple textures usually use small coding depths, such as depth 0 and depth 1. Conversely, CUs comprising complex textures usually use large coding depths, such as depth 2 and depth 3. As textural features are related to depth selection, using textual features to predict candidate depths and skip unlikely coding depths is a common practice. Wang et al. [11] integrate depth probabilities with textural complexity to early skip unlikely depths and early terminated depth selection. The work in [18] employs a tunable decision model to predict candidate CU size partition in Versatile video coding. Lei et al. [19] use a content property analysis to predict candidate depths, thus filtering out unlikely depths. Based on textural features, Xu et al. [20] use deep learning to decide whether to check the current depth for coding speedup. These algorithms are developed based on textural features only.

As the current CU and its relevant CUs usually contains very similar contents, relevant CUs can be used to predict the candidate depths and to skip unlikely depths for the current CU. In [21], a Naive Bayesian Classifier exploites the spatial and Inter-layer correlations to predict candidate sub-CTU partitions for SHVC. In [22], a Support Vector Machine uses temporal and spatial features to predict likely depths for the current CU in EL. Wang et al. [5-8] exploite relevant CUs and their correlation degrees to predict likely depths. The work in [23] uses interlayer, spatio-temporal correlations and inter-level correlations to predict candidate modes. The Inter-layer and spatiotemporal correlations are adopted in [24] to build two feedforwards neural network-based learning models thus predicting candidate depths and modes. The above algorithms are developed based on correlations only.

Both textural features and correlations are related to depth selection. Therefore, it is a common approach to jointly exploit them in the coding depth prediction. Lu et al. [25][26] simultaneously use texture complexity and spatial-temporal correlation to predict the candidate depths, and then integrate Inter-layer correlation with temporal correlation to exclude unlikely DMs.

Residual coefficients reflect the predictive performance and are strongly related to depth selection. Therefore, they can be used to early terminate the depth selection process. Tan et al. [27] use prediction residuals statistics to prune the coding quad-trees. Commonly-used Significance Testing methods, such as t-test [7] and F-test [8], are used to test whether either expected values (t-test) or variances (F-test) of residual coefficients are very similar so as to early terminate the depth selection.

In addition, Yuan et al. [28] first checked whether a CU is a motion-homogeneous block, ane then proposed the corresponding early termination strategies. Yuan et al. [29] proposed an efficient Intra prediction method for H.264/AVC to improve coding speed.

(4) SSHVC Intra-Coding

Most of the existing fast coding algorithms are developed for quality SHVC. The research work on Intra prediction for SSHVC is limited. It is caused by the fact that the Inter-layer correlation is strong, as the resolutions of BL and EL in quality SHVC are the same. It is relatively easy to develop fast prediction algorithms for quality SHVC. However, the resolutions of BL and EL in SSHVC are different, i.e., 2x or 1.5x. Consequently, the Inter-layer correlation is weaker, which is more challenging to develop efficient fast coding algorithms for SSHVC.

Although the above algorithms can improve the coding speed, there are some issues to be addressed for SSHVC:

(1) Residual coefficients are often used to early terminate the mode selection process. Generally speaking, we can early terminate the mode selection process if residual coefficients are small enough. However, residual coefficients are not only related to the mode selection but also closely related to the complexity of texture in the CU. Therefore, it is unlikely to achieve the optimal performance by using the residual coefficients only to early terminate the mode selection.

(2) When Intra mode is selected as the best mode, its DM distribution has never been investigated yet. Predicting candidate DMs without considering their distributions might limit the potential improvement of coding speedup.

(3) Textural features and neighboring CUs are usually used to predict candidate depths. Actually, residual coefficients and RD costs can directly reflect predictive performance and are more strongly related to depth selection. However, they are often independently used in depth prediction. If they can be jointly exploited, the corresponding depth prediction performance can be further improved.

In order to address the above issues, in this research, we develop efficient hybrid strategies for Intra prediction of SSHVC as follows:

(1) We investigate the relationship between RD costs and mode selection, and then use the RD costs of ILR mode to early terminate mode selection. We discover that RD costs of ILR mode and Intra mode have significant difference, and the RD costs of these two modes follow Gaussian distribution. Based on this discovery, we propose to apply the classic GMM-EM in machine learning to decide whether ILR mode is the best mode for the current CU.

(2) Instead of directly predicting the best DM, when Intra mode is used as the best mode, we first investigate DM distribution and then develop corresponding methods to more effectively predict the best DM.

(3) Neighboring CUs, residual coefficients and RD costs are all strongly related to depth selection. We exploit the Conditional Random Field (CRF) Model in machine learning to effectively combine neighboring CUs, residual coefficients and RD costs to early terminate the depth selection of the current CU.

III. Overview of the Proposed Algorithm

To improve Intra coding speed and maintain coding efficiency for SSHVC, in this research work, we propose three strategies: RD Cost of ILR Mode-Based Mode Selection (RCIM-BMS), DM Distribution-Based Progressive DM Selection (DD-BPDS), and Conditional Random Fields-Based Depth Early Termination (CRF-BDET). The overview of the proposed algorithm is illustrated in Fig. 1. First, we use RCIM-BMS to determine whether ILR mode is the best mode. In the affirmative, we can directly skip Intra prediction. Otherwise, we use DD-BPDS to predict the best DM for Intra prediction. After the depth has been checked, we use CRF-BDET to determine whether the current depth can be early terminated. In Fig. 1, the left side shows the three strategies, and the right side presents the procedure of the proposed algorithm.

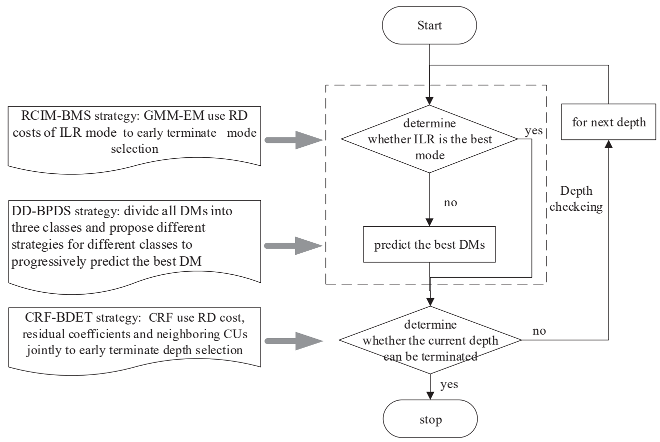


Fig. 1: Overview of the Proposed Algorithm.

We have carried out extensive experiments to analyse the encoding characteristics of Intra coding in SSHVC. Various video sequences that contain different degrees of motion and texture are employed, and they include Blue_sky, Ducks, Park_Joy, Pedestrian, Tractor, Town and Station2. According to the common SHM test conditions (CSTC) [30], two scalability ratios, namely 2x and 1.5x are tested. In cases of 2x and 1.5x, the ratios of the height and width of a frame in EL to those of the corresponding frame in BL are 2 and 1.5, respectively. For each scalability ratio, we utilise one QP set for BL: (22, 26, 30, 34), and two QP sets for EL: (22, 26, 30, 34) and (24, 28, 32, 36). For the aforementioned two scalability ratios and two QPs settings, their combinations in EL can be classified into four cases: case 1 refers to the scalability ratio of 1.5x and the QP set of (22, 26, 30, 34); case 2 indicates scalability ratio of 1.5x and the QP set of (24, 28, 32, 36); case 3 is the scalability ratio of 2x and the QP set of (22, 26, 30, 34); and case 4 denotes the scalability ratio of 2x and the QP set of (24, 28, 32, 36). Among these 4 cases, both the scalability ratio and the QP difference between EL and BL achieve their greatest values in case 4, therefore, the Inter-layer correlation should be the weakest. Consequently, if good encoding performance is achieved in case 4, the performances in the other 3 cases can obtain even better performances. Therefore, the parameter configuration for case 4 is used in our evaluation. Based on the extensive experiments, we propose the three fast Intra prediction strategies as shown below.

IV. RD Cost of ILR Mode-Based Mode Selection (RCIM-BMS)

The justification forming the basis of the proposed GMM-EM-based mode selection algorithm is first discussed, and mode implementation details are then described.

A. Justification of the proposed mode selection algorithm

Using the aforementioned condition in testing, we can obtain the RD Costs of ILR mode and Intra mode, which are listed in Table I. From Table I, we can observe that, from depth 0 to 2, the average values of the RD Costs of ILR mode are significantly larger than those of Intra mode. In contrast, for depth 3, the average value of the RD Costs of ILR mode is smaller than that of Intra mode, but not very significantly.

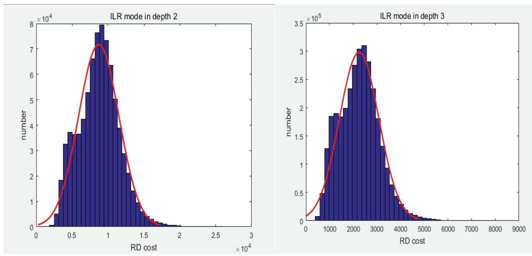
TABLE I: The RD Costs of ILR mode and Intra mode in All Depths

Sequence	Depth 0		Depth 1		Depth 2		Depth 3	
	ILR	Intra	ILR	Intra	ILR	Intra	ILR	Intra
Blue_sky	49801	11074	12172	3198	3039	1523	778	946
Ducks	140109	49873	34478	13686	8701	5732	2247	2444
Park_Joy	166607	21382	41191	5616	10308	3066	1888	2621
Pedestrian	37458	26252	9333	7063	2351	2189	609	763
Tractor	57344	22816	14445	5734	3719	2433	988	1189
Town	108752	63737	27495	16467	6875	4908	1721	1981
Station2	44750	12625	11201	4611	2839	2300	742	901
Average	86403	29680	21474	8054	5405	3164	1282	1549

This is because ILR and Intra mode use different ways in prediction. ILR mode directly upsamples the co-located pixels in horizontal and vertical directions in BL to predict the current CU in EL. Intra mode uses reference pixels from 35 DMs to predict the current CU. For large CUs, namely depth from 0 to 2, the distance between reference pixels and predicted pixels in the current CU is very large. If the texture is very complex, reference and predicted pixels may not be very similar, thus using Intra mode cannot predict the current CU very well. Hence, through upsampling the co-located pixels in BL, ILR mode can predict the current CU better than Intra mode. If the texture is very simple, reference and predicted pixels may be very similar, using 35 DMs, Intra mode can predict the current CU better than ILR mode. For small CUs with depth 3, the distance between reference and predicted pixels in the current CU is very small. Even if the texture is very complex, reference and the predicted pixels may still be very similar, Intra mode using 35 DMs can predict the current CU better than ILR mode. Thus, the average value of the RD Costs of Intra mode is larger than that of ILR mode. If the texture is very simple, both ILR mode and Intra mode can predict the current CU very well, the difference of their RD costs also should be very small. Therefore, the average difference of their RD costs is not very significant.

In addition to the RD cost relationships of ILR mode and Intra mode, we further investigate their RD cost distribution. We have conducted extensive experiments on the RD cost distribution of ILR mode and Intra mode. In Fig. 2, the horizontal axis represents RD cost, and the vertical axis represents the histograms, i.e., the corresponding number of CUs in each bin. Fig. 2. (a) and Fig. 2 (b) show that the RD cost distribution of ILR mode for sequence "Blue_sky" in depth 2 and depth 3, respectively.

From Fig. 2, we can observe that both the RD costs of ILR in depth 2 and depth 3 follow Gaussian distribution. We have performed additional testing by using Kolmogorov-Smirnov to further verify the goodness-of-fit of Gaussian distribution. Kolmogorov-Smirnov is a non-parametric distribution estimation method, which is robust and widely used in distribution estimation. Statistical Product and Service Solutions (SPSS) is commonly used statistical software. Here, we use

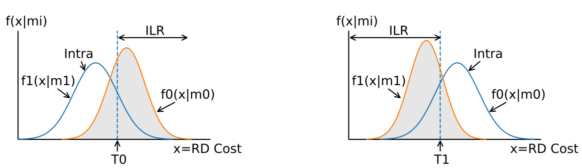


(a) Depth 2 (b) Depth 3

Fig. 2: The RD cost distribution of ILR mode for sequence “Blue_sky”: (a) depth 2, (b) depth 3.

Kolmogorov-Smirnov in SPSS to test the goodness-of-fit of Gaussian distributions. The asymptotic significances of the RD cost distribution of ILR mode for sequence “Blue_sky” in depth 2 and depth 3 are 0. 213 and 0.237 respectively. The commonly use significant level is 0.05. When the asymptotic significances of a data set are greater than 0.05, these data can be considered to follow Gaussian distributions. Therefore, the RD cost distribution of ILR mode for sequence “Blue_sky” in depth 2 and depth 3 follow Gaussian distribution. Our extensive experimental results have demonstrated that the RD cost distribution of both ILR mode and Intra mode in all sequences follow Gaussian distribution.

In light of the above analysis, it is concluded that the average RD cost of the ILR mode is different from that of the Intra mode, but they both follow the Gaussian distribution at all coding depths. To be specific, the average RD cost of the ILR mode is greater than that of the Intra mode at coding depths [0, 2], while less than that of the Intra mode at the coding depth of 3. The RD cost distributions for both the ILR mode and the Intra mode at coding depths [0, 2] and coding depth of 3 are shown in Fig. 3 (a) and Fig. 3 (b), respectively.



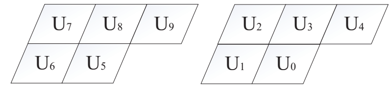
(a) depths from 0 and 2 (b) depth 3

Fig. 3: RD cost distributions of ILR and Intra modes at different depth levels.

B. GMM-EM-based Mode Selection

It is observed from Fig.3 that, despite the different average RD costs, the distribution for the ILR mode overlaps with that for the Intra mode. Therefore, it is impossible to directly decide which is the better choice. GMM-EM is very suitable for classifying the data that follow the Gaussian distribution but have significantly different average values.

GMM-EM is a widely used machine learning algorithm for clustering. It uses mixed Gaussian distribution as the parametric model and utilizes Expectation Maximization (EM) algorithm for training. Suppose we have different k parts and all data in each part follow Gaussian distributions, we add them into a mixed Gaussian distribution model. Based on the model, the probability of sample i belonging to each part can be derived based on the currently available mixture parameters. Then, the mixture parameters are refined. Repeating the process until convergence, then the probabilities of all samples belonging to each part can be obtained. Based on Gaussian distribution and EM algorithm, the theoretical basis of GMM-EM is very solid and can work very efficiently to cluster data. As mentioned above, the average value of RD costs of ILR mode and Intra mode are different from each other and both of them follow Gaussian distribution for all depths. Obviously, GMM-EM can use this feature to decide which mode is the best one based on its RD Cost. Therefore, in this research, we propose to apply GMM-EM in mode selection for SSHVC.



(a) previous frame (b) current frame

Fig. 4: The current CU and its relevant CUs

Since ILR mode is much more likely to be selected as the best mode [7-8], we encode the current CU using ILR mode first, and then use GMM-EM to determine whether ILR mode is the best mode based on its RD cost. In the affirmative, we can early terminate mode selection; otherwise, we need to further check Intra mode.

As shown in Fig. 4, U_0 represents the current CU, U_1, U_2, U_3 and U_4 are its neighboring CUs, U_5, U_6, U_7, U_8 and U_9 are respectively the co-located CUs of U_0, U_1, U_2, U_3 and U_4 in the previous frame. Obviously, these nine CUs are most similar to the current CU. Since there are only two parts, i.e., ILR mode and Intra mode, let part one represent ILR mode, and part two indicates Intra mode. The corresponding Gaussian Mixture Model can be written as

$$p\left(rd|\pi,\mu,\Sigma\right)=\pi_1 N\left(rd|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd|\mu_2,\Sigma_2\right), \tag{1}$$

where rd is the RD cost of a CU, π_1 is the probability of the CU using ILR mode, μ_1 and Σ_1 are respectively the expected value and variance of the RD costs of all these CUs selecting ILR mode as the best one; and π_2 is the probability of the CU using Intra mode, μ_2 and Σ_2 are respectively the expected value and variance of RD costs of all these CUs selecting Intra mode as the best one.

In order to obtain these six parameter values, based on Eq. (1), the corresponding maximum likelihood estimation function is

$$f=\prod_{i=1}^{M} p\left(rd|\pi,\mu,\Sigma\right)=\prod_{i=1}^{M}\left(\pi_1 N\left(rd|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd|\mu_2,\Sigma_2\right)\right), \tag{2}$$

where M is the number of the current CU and its relevant CUs. Since there are nine relevant CUs, from U_1 to U_9, counting the current CU U_0 itself, M is equal to 10.

The logarithm of the maximum likelihood function is

$$\log\left(f\right)=\sum_{i=1}^{M}\log\left(\pi_1 N\left(rd_i|\mu_1,\Sigma_1\right)+\pi_2 N\left(rd_i|\mu_2,\Sigma_2\right)\right), \tag{3}$$

where π_k, μ_k and Σ_k (k=1 or 2) can be calculated by

$$\frac{\partial \log(f)}{\partial \pi_k} = 0, \frac{\partial \log(f)}{\partial \mu_k} = 0, \frac{\partial \log(f)}{\partial \Sigma_k} = 0. \tag{4}$$

Then, we can derive

$$\mu_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k) rd_i, \Sigma_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)(rd_i-\mu_k)(rd_i-\mu_k), \tag{5}$$

where $N_k = \sum_{i=1}^{N}\gamma(i,k)$, then we have

$$\pi_k = \frac{N_k}{N}. \tag{6}$$

$\gamma(i,k)$ is the probability that the i-th CU, denoted as U_i, belongs to the k-th part, and it can be obtained by:

$$\gamma(i,k) = \frac{\pi_k N(x_i|\mu_k,\Sigma_k)}{\pi_1 N(x_i|\mu_1,\Sigma_1) + \pi_2 N(x_i|\mu_2,\Sigma_2)}. \tag{7}$$

By repeating (5), (6) and (7) until $\gamma(i,k)$ converges, the best $\gamma(i,k)$ can be obtained.

As mentioned above, the current CU is U_0 and ILR mode is part one, hence $\gamma(0,1)$ denotes the probability of the current CU selecting ILR mode as the best one. We denote the i-th iteration of $\gamma(0,1)$ as $\gamma_i(0,1)$. In order to avoid unnecessary iterations, if the absolute difference between $\gamma_{i-1}(0,1)$ and $\gamma_i(0,1)$ is small enough, we can terminate the iteration. We empirically select 0.01 as the threshold, then we obtain the early termination condition below

$$|\gamma_i(0,k) - \gamma_{i-1}(0,k)| \le 0.01. \tag{8}$$

If the condition (8) is met, we can terminate the iteration and obtain the probability of the current CU selecting ILR mode as the best mode. Since this probability is obtained based on RD cost, we define it as the RD-based ILR probability.

Obviously, the current CU and its relevant CUs are usually very similar. Therefore, if more relevant CUs use ILR mode, the current CU is also more likely to use ILR mode, and vice versa. Therefore, the probability of the current CU using ILR mode is proportional to the number of its relevant CUs using ILR mode. As shown in Fig. 3, there are 9 relevant CUs, so we simply set the possibility of the current CU using ILR mode as $\frac{k}{9}$, where k is the number of the relevant CUs using ILR mode. Since this probability is obtained based on the number of the relevant CUs using ILR mode, we define it as number-based ILR probability.

Since both the RD-based ILR probability and the number-based ILR probability strongly relate to ILR mode selection, we combine them to further predict the probability of the current CU using ILR mode. Let p(A) and p(B) denote the RD-based ILR probability and the number-based ILR probability, respectively. Since both of them are independent, we further derive the probability of the current CU selecting ILR mode as the best mode, p_r, by

$$p_r = p(A+B) = p(A) + p(B) - p(A)p(B). \tag{9}$$

Apparently, the larger the p_r is, the more likely the current CU will use ILR mode. Theoretically, if $p_r \ge 0.95$, the current CU is very likely to use ILR mode. Therefore, we select 0.95 as the threshold for p_r.

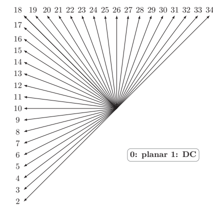


Fig. 5: 35 DMs in SHVC

V. DM Distribution-Based Progressive DM Selection (DD-BPDS)

In order to predict the likely DMs, we first investigate the distribution for DMs and categorise all DMs into three classes. The corresponding DM selection approaches proposed for each of the categories are introduced as shown below.

A. *Distribution-based DM Classification*

DMs in SHVC is illustrated in Fig. 5. There are 2 non-directional modes, i.e., DC (DM0) and planar (DM1), and 33 directional modes (DM2 $\cdots$ DM34). Generally speaking, DM0 and DM1 are very suitable for simple CUs. Similar to HEVC, SHVC first check 35 DMs in Rough Mode Decision to select N DMs with the smallest HCs, and then check these N DMs in Rate Distortion Optimization (RDO) process to select the DM with the smallest RDO value as the best DM. Using the above process, the best coding efficiency can be obtained. However, checking many unnecessary DMs cost much unnecessary computation time. Especially the Rough Mode Decision process always check 35 DMs, which is very time consuming. Therefore, we can skip unlikely DMs to improve coding speed.

As mentioned before, the average RD cost of the ILR mode is greater than that of the Intra mode at coding depths [0, 2]. In other words, if large CUs in EL use Intra mode, they are usually very simple. Similarly, for small CUs in EL, their texture cannot change very significantly due to small sizes, so they are also usually very simple. Obviously, simple CUs may have special DM features. Investigating these DM features in EL may help improve coding speedup. Different DMs may have different probabilities to be selected as the best DM among all CUs. We should first obtain their probabilities, and then investigate their distribution by grouping DMs with similar probabilities into a class. The probability of DM i is calculated by

$$p_i = \frac{n_i}{m}, \mathrm{i} = 0, \cdots, 34, \tag{10}$$

where n_i is the number of DM i is selected as the best DM among all CUs, m is the number of all CUs.

Through extensive experiments, we divide 35 DMs in EL into three classes according to the probability: Class 0, Class 1 and Class 2. Class 0 includes DM0 and DM1; Class 1 contains DM8, DM9, DM10, DM11, DM12, DM24, DM25, DM26, DM27 and DM28; Class 2 includes the remaining DMs. In addition, DMs in Class 1 is further divided into two subclasses,

TABLE II: DM distributions among classes for large CUs and small CUs in EL

Sequence	Large CU			Small CU		
	Class 0	Class1	Class 2	Class 0	Class 1	Class 2
Blue_sky	68.23%	14.15%	17.62%	76.41%	7.52%	16.07%
Ducks	20.28%	70.5%	9.22%	30.37%	55.02%	14.61%
Park_Joy	66.55%	17.29%	16.16%	67.70%	14.69%	17.61%
Pedestrian	57.92%	35.16%	6.92%	69.89%	24.73%	5.38%
Tractor	38.31%	34.69%	27.0%	52.58%	29.27%	18.15%
Town	72.85%	21.95%	5.2%	61.73%	30.86%	7.41%
Station2	41.27%	29.18%	29.55%	54.74%	19.58%	25.68%
Average	52.20%	31.85%	15.95%	59.06%	25.95%	14.99%

i.e., the horizontal subclass including DM8, DM9, DM10, DM11, DM12, and the vertical subclass including DM24, DM25, DM26, DM27 and DM28. For the convenience of the following descriptions, we define the DM distribution of a Class k, D_k, is the sum of the probabilities of all DMs in this Class. Specifically, D_k can be derived by

$$D_k = \sum_{i \in \text{class } k} p_i, \; k = 0, 1, 2. \tag{11}$$

As mentioned in the above section, in our experiments, we observe that the RD cost relationships between Intra mode and ILR mode in large CUs and small CUs are different, their DM distributions may also be different accordingly. Therefore, in our experiments, we separately investigate DM distributions among classes for large CUs and small CUs in EL. The corresponding results are presented in Table II.

From Table II, we can find that on average more than 50% CUs use DMs in Class 0, and about 85% CUs use DMs in Class 0 and Class 1 together. As mentioned above, DMs in Class 0 are very appropriate for simple CUs. In addition, DMs in Class 1, including horizontal subclass and vertical subclass, are also very suitable for simple CUs. Thus, DMs in both class 0 and class 1 are suitable for simple CUs. We define these DMs in Class 0 and Class 1 as simple DMs. It reflects that CUs in EL using Intra mode usually have very simple texture, which is completely consistent with the above-mentioned simple CUs using Intra mode in EL. According to the results in Table II, we found that around 15% CUs use DMs in Class 2. These CUs are generally complicated or irregular.

B. Class-based DM Selection

Even though there are two DMs only in Class 0, more than half of the CUs use these two DMs. Likewise, there are up to twelve DMs altogether in Class 0 and Class 1, but about 85% of CUs adopt the DMs in these two classes. The majority of DMs (23 DMs) are included in Class 2, and about 15% of the best DMs come from Class2. Therefore, the coding speed can be significantly improved if the best DM in Class 0 or Class 1 is determined in advance, thus terminating the DM selection. It is observed in Table II that the DM distributions of the three classes are quite distinct from each other. The number of DMs in each of these three classes are also significantly different. Targeting each of the classes, we develop different methods to obtain the best DM, thus effectively speeding up the coding process.

1) significant difference-based DM early termination: we develop a significant difference-based DM early termination approach for the DMs in Class 0. For simplicity, we denote the HC of DMi as HCi and min() is the smaller one of the two different HC values. DM0 and DM1 in Class 0 are non-directional modes. DM10 and DM26 are two typical ones in Class 1, where DM10 and DM26 represent the horizontal and vertical directions, respectively. If HC0 and HC1 are significantly less than HC10 and HC26, the best DM very likely comes from Class 0. HC0 and HC1 are greatly less than HC10 and HC26 if min (HC0, HC1) is significantly less than min (HC10, HC26). To decide if min (HC0, HC1) and min (HC10, HC26) are sufficiently different, their corresponding residual coefficients are investigated. Suppose $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ are the residuals matrixs of two DMs, the difference $\boldsymbol{R}$ between the two matrices is [7]

$$R = R_1 - R_2. \tag{12}$$

Applying Hadamard transformation to the Eq. (12), we have

$$HRH = HR_1H - HR_2H, \tag{13}$$

where $\boldsymbol{H}$ is a $m \times m$ Hadamard matrix and m is the size of the current CU. Suppose $\sum_{i=0}^{m}\sum_{j=0}^{m} HRH$ is the sum of the coefficients in the matrix $\boldsymbol{HRH}$, according to Cauchy-inequality, we can derive

$$\begin{aligned} \sum_{i=0}^{m}\sum_{j=0}^{m} HRH &\le \left|\sum_{i=0}^{m}\sum_{j=0}^{m}\left(HH^T\right)^2\right|^{\frac{1}{2}} \times \left|\sum_{i=0}^{m}\sum_{j=0}^{m} {r_{ij}}^2\right|^{\frac{1}{2}} \\ &\le \sqrt{m}\left|\sum_{i=0}^{m}\sum_{j=0}^{m} {r_{ij}}^2\right|^{\frac{1}{2}} \le \sqrt{m}\sum_{i=0}^{m}\sum_{j=0}^{m}\left|r_{ij}\right|. \end{aligned} \tag{14}$$

x_{ab} denotes the value at position (a, b) in $\boldsymbol{HRH}$, and it is

$$x_{ab} = \sum_{i=0}^{m}\sum_{j=0}^{m} h_{ai} r_{ij} h_{jb} \le \sum_{i=0}^{m}\sum_{j=0}^{m}\left|h_{ai}h_{jb}\right|\left|r_{ij}\right| \le \sum_{k=0}^{m}\sum_{p=0}^{m}\left|r_{ij}\right|. \tag{15}$$

If the quantised values in HRH are all less than k, $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ are not considered to be greatly different. The following condition is derived

$$\sum_{k=0}^{m}\sum_{p=0}^{m}\left|r_{kp}\right| < kQ_{step}. \tag{16}$$

Combining Eq. (13), (14), and (16), it gives

$$\left|\sum_{i=0}^{m}\sum_{j=0}^{m} HR_1H - \sum_{i=0}^{m}\sum_{j=0}^{m} HR_2H\right| < k\sqrt{m}Q_{step}. \tag{17}$$

Obviously, the following inequality holds

$$\left|\sum_{i=0}^{m}\sum_{j=0}^{m}|HR_1H| - \sum_{i=0}^{m}\sum_{j=0}^{m}|HR_2H|\right| \le \left|\sum_{i=0}^{m}\sum_{j=0}^{m} HR_1H - \sum_{i=0}^{m}\sum_{j=0}^{m} HR_2H\right|. \tag{18}$$

TABLE III: k and the corresponding BDBRs

Sequence	1	2	3	4	5	6
blue_sky	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%
ducks	-0.01%	-0.01%	0.00%	-0.01%	0.00%	0.00%
park_joy	-0.01%	-0.03%	0.00%	0.00%	0.01%	0.01%
pedestrian	-0.02%	0.00%	0.00%	0.00%	0.00%	0.01%
town	-0.16%	-0.11%	-0.10%	-0.07%	-0.03%	-0.02%
station2	0.16%	0.14%	0.13%	0.11%	0.08%	0.06%
tractor	0.01%	0.00%	0.01%	0.00%	-0.01%	-0.01%

Combining Eq. (17) and (18), we have

$$\left|\sum_{i=0}^{m}\sum_{j=0}^{m}|HR_1H| - \sum_{i=0}^{m}\sum_{j=0}^{m}|HR_2H|\right| < k\sqrt{m}Q_{step}, \tag{19}$$

Eq. (19) can be rewritten as

$$|HC_1 - HC_2| < k\sqrt{m}Q_{step}, \tag{20}$$

where HC_1 and HC_2 refer to sum of absolute values of Hadamard transform values of two DMs. If Eq. (20) is satisfied, they are not considered to be significantly different. Similarly, if the following condition holds, HC_1 and HC_2 are regarded to be significantly different.

$$|HC_1 - HC_2| > k\sqrt{m}Q_{step}. \tag{21}$$

We denote min (HC0, HC1) and min (HC10, HC26) as MC_1 and MC_2, respectively. To ensure MC_1 is significantly less than MC_2, Eq. (20) needs to be rewritten as

$$MC_1 - MC_2 > k\sqrt{m}Q_{step}. \tag{22}$$

If Eq. (22) holds, MC_1 is considered to be significantly less than MC_2. According to the test condition, we test a series of values, such as 1, 2, 3, etc. to obtain the best value of k. The corresponding BDBRs[31], which measures the bitrate difference at equal PSNR in the EL, are listed in Table III.

Table III shows a turning point when k equals 5. If k is greater than or equal to 5, the corresponding BDBRs for all the sequences are less than 0.1%. It means that k of 5 gives very good performance. If we choose a larger k, the corresponding increase in coding speed will be less significant. Therefore, k is set to 5.

If Eq. (22) holds, min (HC0, HC1) is significantly less than min (HC10, HC26). The best DM comes from Class 0 and we early terminate the DM selection process. Otherwise, we need to search for the best DM in Class 1.

2) DM prediction-descent-based direction search: As the DMs in Class 0 and Class 1 are very likely to be the best DM, if a DM and its two direct neighbours have been checked and its HC is the least in all checked DMs, this DM is very Likely the Best DM (LBD). To obtain an LBD as soon as possible, we follow the descent direction of the HCs to search for it. If an LBD is obtained, we early terminate the DM selection process. Otherwise, we need to search for the best DM in Class 2.

3) variable stepsize-binary search method: Table II shows about 15% of the CUs use DMs in Class 2 regardless of the size of the CU. Therefore, the probability of CUs using DMs in Class 2 is very small but cannot completely be ignored. If we always check all DMs in Class 2, it will cost a lot of unnecessary time. However, if we completely skip these DMs, the coding efficiency is significantly degraded. Therefore, it is crucial to obtain the best DM and skip unlikely DMs. For a DM in [DM2, DM18] or [DM18, DM34], if the distance between this DM and DM10 or DM26 is farther, the possibility of using this DM is smaller and the stepsize should be larger, and vice versa. Therefore, we use a variable stepsize to search for the best DM in Class 2. If there is a DM with the smallest HC among all checked DMs in Class 2, we then use a binary search to find the best DM. To be specific, we first check the midpoint between the DM with the least HC of all checked DMs and its nearest left checked neighbouring DM, then check the midpoint between the DM and its nearest right checked neighbouring DM, and finally, we select a DM with the least HC in all checked DMs. We repeat the process until a DM is an LBD. Since the above process includes variable stepsize check and binary search, it is called variable stepsize check-binary search.

VI. Conditional Random Fields-Based Depth Early Termination (CRF-BDET)

As we known, if the RD costs of a CU is very small, the CU is very likely to be early terminated. The residual coefficients of a CU are also strongly related to depth early termination. Due to the similarity of neighboring CUs, a CU and its neighboring CUs are usually very similar. Apparently, the three features of a CU, RD costs, residual coefficients and neighboring CUs, have strong relationships with its depth early termination. Based on these three features, we propose to apply the CRF Model in the field of machine learning to predict whether the current CU can be early terminated.

The CRF model is a discriminant probabilistic undirected graph model with strong probabilistic reasoning ability [32]. CRFs have been widely used for many areas, such as natural language processing and image area tagging. CRFs are excellent at modeling the structure of sequential data and can flexibly construct dependence between inputs and labels. This makes CRFs suitable to predict depth early termination, which can both exploit its characteristics on modeling sequential data structure and fuse the neighboring information from data. Particularly, CRFs have the advantage of modeling data with structured labels and simulating any features of observed sequences. In addition, the CRF needs not make independent assumptions on observations and can simulate arbitrary features of observed sequences.

Therefore, we propose to use the CRF model to efficiently simulate RD costs, residual coefficients and neighbouring CUs to achieve the early termination of coding depth evaluation. We first introduce the fundamentals of CRF and its application in video coding and then propose the CRF-based depth early termination approach as shown below.

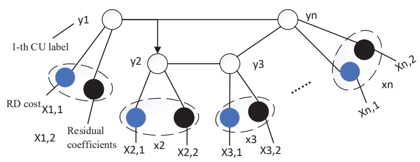
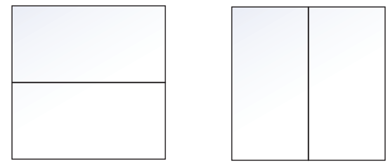


Fig. 6: A CRF of depth early termination



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

Fig. 7: The division of a CU

A. Fundamentals of CRF and Its Application in Video Coding

In the following, we first introduce the fundamentals of CRF, and then discuss how to apply CRF in depth early termination for video coding. Let X and Y represent an observed data sequence and its corresponding labeled sequence, they form an undirected graph. If Y is conditioned on X and it obeys the Markov property, then (Y, X) forms a CRF, which means

$$P\left(y_i|x_i, y_{w-i}\right) = P\left(y_i|x_i, y_{N_i}\right), \tag{23}$$

where x_i and y_i are the i-th variable and individual label of X and Y respectively; y_{w-i} refers to the remaining labels apart from y_i, $P\left(y_i|x_i, y_{w-i}\right)$ represents the probabilities of y_i conditioned on the observed variable x_i and the remaining labels y_{w-i}; y_{N_i} represents the neighbors of y_i, $P\left(y_i|x_i, y_{N_i}\right)$ represents the probabilities of y_i conditioned on the observed variable x_i and the neighbors y_{N_i}.

Using CRFs to derive the probability of y_i under the above condition, Eq (23) can be rewritten as

$$P\left(\mathrm{y}_i|\mathrm{x},\mathrm{y}_{N_i}\right) = \frac{1}{Z\left(x\right)} \exp\left(\sum_{m\in\varepsilon} \alpha_m U_m\left(y_i|x_{i,m}\right) + \sum_{n\in N_i} \beta_n I_n\left(y_i|y_{i,n}\right)\right), \tag{24}$$

where $Z(x)$ is a normalizing item for the conditional distribution, ε is the feature value set of y_i, U_m, $x_{i,m}$ and α_m are the m-th unary potential, the corresponding feature value and weighting parameters of y_i respectively; N_i is the neighbor set of y_i, I_n, $y_{i,n}$ and β_n are the n-th interaction potential, the corresponding neighbors and weighting parameters of y_i respectively. The 1^{st} part in Eq. (24) indicates the unary potential which refers to the local decision part without considering neighbors to express the local characteristics. The 2^{nd} part represents the interaction potential, which denotes the continuity between the labels and its neighbors.

In depth early termination for video coding, y_i is the label of the i-th CU. 0 and 1 for y_i indicate either early termination or split, respectively. Since unary potential express the local characteristics of a CU, both the RD cost and residual coefficients are actually unary potentials. As the interaction potential denotes the continuity between the labels of a CU and its neighbors, the neighboring CUs are obviously interaction potentials. Based on the above analysis, we can obtain the corresponding CRF for depth early termination in video coding, as shown in Fig.6. $x_{i,1}$ and $x_{i,2}$ are denoted as the RD cost and residual coefficients of the i-th CU, respectively; while y_i is the label of the i-th CU.

B. CRF-based Depth Early Termination

To effectively achieve an early termination of depth selection, we first use RD cost, residual coefficients and neighboring CUs to obtain their respective probabilities of depth early termination. The gradient ascent-based CRF solution is then subsequently developed to obtain the joint probability of depth early termination. Finally, we propose the corresponding optimization method. The details are described below.

1) The depth early termination probabilities for RD cost, residual coefficients and neighboring CUs:

(a) RD cost-based depth early termination probability

As mentioned above, a CU with a smaller RD cost is more likely to be split, and vice versa. Similar to ILR mode prediction in RCIM-BMS, both the RD costs of non-split and split CUs follow Gaussian distributions, and their expected values and variances are significant different [27]. Therefore, we can also use GMM to obtain the probability of depth early termination, based on its RD cost.

(b) Residual coefficients-based depth early termination probability

After the residual coefficients of a CU have been obtained, we first divide the CU into the top part and the bottom part (Fig.7 (a)), whose corresponding expected values are denoted as μ_1 and μ_2. Then, we divide the CU into the left part and the right part (Fig. 7 (b)), whose corresponding expected values are denoted as μ_3 and μ_4. If the expected values of the residual coefficients of the two parts in either of the above divisions, namely μ_1 and μ_2, or μ_3 and μ_4, are significantly different, the current CU is very likely to be further split, and vice versa.

We denote the maximum value between $|\mu_1{-}\mu_2|$ and $|\mu_3{-}\mu_4|$ as μ. The smaller the value of μ is, the current depth is more likely to be early terminated, and vice versa. Since the probability of depth early termination, denoted as $f(\mu)$, ranges from 0 to 1, we adopt an exponential function with base e to fit the relationship between μ and $f(\mu)$ as shown below.

$$f(\mu) = \exp(a_0 + a_1\mu + a_2\mu^2 + a_3\mu^3 + \cdots + a_b\mu^b), \tag{25}$$

where a_i is the value of i-th coefficient and b is the maximum exponent. The key is how to obtain the best a_i and b.

Using the aforementioned test condition in testing, we can obtain μ and the corresponding termination status for all CUs in all the above sequences. We fit the relationship using b of 1, 2, 3, etc. to obtain its best value through MATLAB. Theoretically, the larger the value of b is, the better the model can fit the relationship. However, during our experiments, we observe that if b is greater than or equal to 4, the differences in the fitting results are very small. Therefore, we set b to 4 and the corresponding model is presented below.

$$\begin{aligned} f(\mu) = \exp(&-3.2\times10^{-4}\mu^4 + 9.4\times10^{-4}\mu^3 - 0.085\mu^2 \\ &-0.059\mu - 0.004), \end{aligned} \tag{26}$$

We use Eq (26) to derive the corresponding $f(\mu)$ for any μ.

(c) Neighboring CUs-based depth early termination probability

Generally, a CU and its neighbouring CUs are very similar. Therefore, we can use its neighbouring CUs which can be early terminated, denoted as early terminated neighbouring CUs (ETNCUs), to predict the current CU for early termination. However, different ETNCUs may have various degrees of correlation with the current CU. To achieve a more accurate prediction, the degree of correlation should be taken into consideration when predicting the probability of depth early termination. We propose to use Pearson Correlation Coefficient (PCC) [22] to obtain the degree of correlation between the current CU and its ETNCUs. PCC is a commonly used method to calculate the degree of correlation between two sets of data. Adopting the pixel values of two neighboring CUs, we can calculate their corresponding PCC denoteds as r by

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}, \tag{27}$$

where x_i and y_i are the i-th pixel values in the current CU and one of its neighboring CUs, respectively. Since different CUs may have various numbers of ETNCUs, for a CU with more than one ETNCUs, we only select the one with the largest PCC. For a CU without ETNCUs, its PCC is set to 0.

PCC is strongly correlated with the probability of depth early termination. The larger the PCC is, the greater the probability of depth early termination is, and vice versa. Therefore, we can directly use PCC as the probability of depth early termination.

2) Gradient Ascent based-CRF Solution: Through the above process, we can obtain the RD cost-based depth early termination probability denoted as v_1, the residual coefficients-based depth early termination probability denoted as v_2 and the neighboring CUs-based depth early termination probability denoted as v_3. Their corresponding weighting parameters are denoted as w_1, w_2 and w_3, respectively. Accordingly, the probabilities of depth for further split in RD cost, residual coefficients and neighboring CUs are $1-w_1$, $1-w_2$ and $1-w_3$, respectively. Based on Eq. (24), the joint probability of depth early termination of a CU, p, can be obtained by

$$p = \frac{P\left(0|\mathrm{x},\mathrm{y}_{N_i}\right)}{P\left(0|\mathrm{x},\mathrm{y}_{N_i}\right) + P\left(1|\mathrm{x},\mathrm{y}_{N_i}\right)} = \frac{e^{\sum_{i=1}^{3} w_i v_i}}{e^{\sum_{i=1}^{3} w_i v_i} + e^{\sum_{i=1}^{3} w_i (1-v_i)}}. \tag{28}$$

Obviously, the larger the p is, the more likely the current CU will be terminated early. Therefore, we can determine whether a CU can be terminated early based on the maximum p. In order to obtain the maximum p, we must obtain the best w_i. Taking the logarithm of the above formula, we obtain

$$In(p) = \sum_{i=1}^{3} w_i v_i - In\left(e^{\sum_{i=1}^{3} w_i v_i} + e^{\sum_{i=1}^{3} w_i (1-v_i)}\right). \tag{29}$$

Taking the partial derivative of Eq. (29) with respect to w_i, we have

$$\frac{\partial In(p)}{\partial w_i} = \frac{(2v_i - 1)\, e^{\sum_{i=1}^{3} w_i (1-v_i)}}{e^{\sum_{i=1}^{3} w_i v_i} + e^{\sum_{i=1}^{3} w_i (1-v_i)}}. \tag{30}$$

Gradient ascent is a common method for calculating the maximum value. In order to obtain the maximum p, we use a gradient ascent in calculation of the best w_i below

$$w_i^{k+1} = w_i^k + \alpha \frac{\partial In(p)}{\partial w_i}, \tag{31}$$

where k is the number of repetitions, and w_i^k denotes the k-th iteration of w_i. Repeat Eq. (31) until the best w_i is obtained. To start iterations, we empirically set the initial weight w_i^0 to be 0 and the learning rate α to be 0.01. In order to avoid unnecessary iterations, if the absolute difference between w_i^{k-1} and w_i^k is small enough, we can terminate the iteration. We empirically select 0.001 as the threshold, then we can obtain the early termination condition below

$$\left|w_i^{k-1} - w_i^k\right| \leq 0.001. \tag{32}$$

Through the above process, we can obtain the best w_i, and then we can calculate the maximum probability of depth early termination, $p_{\max}$, through Eq (28). Theoretically, if $p_{\max} \geq 0.95$, the current CU is very likely to be early terminated and needs not to be further split. Therefore, we select 0.95 as its threshold value.

3) The optimization method: For the convenience of description, w_1, w_2 and w_3 are combined to form a vector $\mathbf{W}$, and v_1, v_2 and v_3 are combined to form a vector $\mathbf{V}$. Obviously, the best $\mathbf{W}$ depends on its corresponding $\mathbf{V}$. In other words, $\mathbf{V}$ and $\mathbf{W}$ have a one-to-one mapping relationship. To simplify the above iterative processing, we first obtain the corresponding best $\mathbf{W}$ through Eq. (31) and (32) for a given $\mathbf{V}$. If we obtain the same $\mathbf{V}$ afterwards, we can directly use the corresponding $\mathbf{W}$.

Therefore, we partition the range from 0 to 1 of each component in $\mathbf{V}$ into 10 equally spaced intervals. Theoretically, the average distance between the midpoint and other values in an interval are the shortest. Therefore, we select the midpoint as a representative value (**RV**) in each interval, and then use Eq (31) and (32) to calculate the corresponding best Representative **W**s (**RW**s) and store them in an **RW** Table. When encoding a new CU, we can obtain the **V** and its **RV** based on its RD cost, residual coefficients and neigbouring CUs, and then obtain the corresponding **RW** by checking the **RW** Table. We use the **RW** as the initial value of **W**, and then use Eq (31) and (32) to calculate its best **W**. Since **V** and its **RV** are very similar, their corresponding **RW** and the best **W** are also very similar. Therefore, we do not need to iterate many times and many unnecessary iterations can be skipped. Experiments show that if the gradient ascent is directly used to calculate the best **W**, we usually need to iterate hundreds of times for a CU. While using the proposed optimization method, we only need to iterate a few times. Therefore, this approach can significantly reduce the number of iterations and improve the coding speed effectively.

TABLE IV: Performance comparisons among different proposed strategies

Sequence	RCIM-BMS (%)		DD-BPDS (%)		CRF-BDET (%)	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.18	64.02	-0.02	35.32	0.0	34.30
PeopleOnStreet	-0.29	62.66	0.00	35.89	-0.2	34.41
Kimono	-0.30	66.29	-0.09	36.80	0.0	35.39
ParkScene	-0.15	65.04	-0.01	35.08	0.0	33.54
Cactus	0.41	60.33	0.31	34.84	0.0	34.08
BasketballDrive	0.96	54.42	0.39	34.29	0.1	33.39
BQTerrace	1.69	42.95	0.91	34.37	0.0	33.41
Average	0.31	59.39	0.21	35.23	-0.01	34.07

VII. Experimental Results

In order to evaluate the performance of the proposed efficient hybrid strategies for SSHVC, we use the reference software (SHM 11.0) and test the proposed algorithm on a server with Intel (R) 2.0 GHz CPU and 30 GB memory. The performances of the proposed algorithms are evaluated in terms of coding efficiency and coding speed. The coding efficiency is indicated by BDBR. A negative or positive BDBR represents an increase or decrease in coding efficiency compared with the reference software, respectively. Coding speed improvement is denoted by TS, which is the percentage of encoding run-time savings only in EL.

As mentioned above, we have proposed three strategies: RCIM-BMS, DD-BPDS and CRF-BDET. Since all the above processes use the parameter configuration in Case 4, we also use this parameter configuration for EL. The corresponding coding performance is provided in Table IV.

It is seen in Table IV that the average TS, i.e., coding speed improvements, in “RCIM-BMS”, “DD-BPDS” and “CRF-BDET” are 59.39%, 35.23%, 34.07%, respectively. Their corresponding average BDBRs, i.e., coding efficiency losses, are 0.31%, 0.21%, -0.01%, correspondingly. All three strategies can remarkably accelerate coding speed with negligible coding efficiency losses. The coding process of ILR mode is very simple, and most CUs use this mode only in mode prediction. For RCIM-BMS, most CUs use ILR mode only without having to check the Intra mode. Thus, the coding speed is most significantly improved, but its BDBR also increases correspondingly. As mentioned above, although Intra prediction includes RMD and RDO processes, we only skip some unnecessary DMs in RMD. Therefore, DD-BPDS achieves a relatively lower but still significant coding speed improvement. Due to the strict thresholds for early termination of depth selection, the coding speed improvement achieved by “CRF-BDET” is the smallest but its increase in BDBR is also negligible.

In order to demonstrate the overall performance of the proposed algorithm which incorporates all the proposed strategies, i.e., “RCIM-BMS”, “DD-BPDS” and “CRF-BDET”, we compare the performance of our algorithm with PAPS [21] and FIICA [26]. As FDMDIP [7] and EMIP[8] are developed for quality SHVC rather than spatial SHVC, it is pointless to compare with them. To the best of our knowledge, these two algorithms are the state-of-the-art fast Intra coding algorithms for SSHVC. To achieve fair comparisons, we test all algorithms on the same computing platform. As two scalability ratios and two QPs settings are used in the evaluation, we group the combinations into four cases. The corresponding overall performance comparisons are listed in Table V (case 1), Table VI (case 2), Table VII (case 3) and Table VIII (case 4), respectively. According to CSTC [30], there are five sequences only for case 1 and case 2, and seven sequences for case 3 and case 4.

In Table V (case 1), the average BDBRs of the proposed algorithm, PAPS and FIICA are -0.08%, 0.30% and 0.06%, respectively. While the average TS of the proposed algorithm, PAPS and FIICA are 74.69%, 67.03% and 47.15% correspondingly. In Table VI (case 2), the average BDBRs of the proposed algorithm, PAPS and FIICA are -0.22%, 0.38% and

TABLE V: Overall performance comparisons with case 1

Sequence	Proposed (%)		PAPS (%) [21]		FIICA(%) [26]	
	BDBR	TS	BDBR	TS	BDBR	TS
Kimono	-0.33	75.90	0.06	70.81	-0.21	62.35
ParkScene	-0.21	74.97	0.49	66.50	-0.12	38.17
Cactus	-0.13	75.17	0.10	65.37	-0.18	41.89
BasketballDrive	0.12	73.20	0.46	67.14	0.40	47.06
BQTerrace	0.16	74.20	0.39	65.32	0.41	46.27
Average	-0.08	74.69	0.30	67.03	0.06	47.15

TABLE VI: Overall performance comparisons with case 2

Sequence	Proposed (%)		PAPS (%) [21]		FIICA(%) [26]	
	BDBR	TS	BDBR	TS	BDBR	TS
Kimono	-0.21	77.04	0.17	68.76	0.81	61.67
ParkScene	-0.13	76.39	0.58	67.43	-1.22	37.46
Cactus	-0.31	76.67	0.27	63.21	0.31	40.13
BasketballDrive	-0.31	75.25	0.41	66.172	-0.80	44.36
BQTerrace	-0.13	75.40	0.48	63.68	0.00	45.15
Average	-0.22	76.15	0.38	65.85	-0.18	45.75

TABLE VII: Overall performance comparisons with case 3

Sequence	Proposed (%)		PAPS (%) [21]		FIICA(%) [26]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.03	74.61	0.24	76.40	0.41	36.37
PeopleOnStreet	0.28	75.14	0.22	62.80	0.10	39.43
Kimono	-0.23	75.03	0.21	73.12	-0.13	60.27
ParkScene	-0.16	73.39	0.42	64.51	0.22	36.49
Cactus	1.02	71.27	0.58	70.67	0.89	37.92
BasketballDrive	1.46	66.33	1.87	67.42	0.71	41.48
BQTerrace	2.32	70.23	0.83	63.21	0.50	43.56
Average	0.67	72.28	0.62	68.30	0.38	42.22

TABLE VIII: Overall performance comparisons with case 4

Sequence	Proposed (%)		PAPS (%) [21]		FIICA(%) [26]	
	BDBR	TS	BDBR	TS	BDBR	TS
Traffic	-0.17	75.89	0.27	74.56	-0.30	37.89
PeopleOnStreet	-0.29	76.54	0.29	61.23	-0.23	40.15
Kimono	-0.3	75.55	0.28	71.25	0.19	60.18
ParkScene	-0.15	74.44	0.46	61.46	0.11	38.13
Cactus	0.68	73.12	0.55	71.12	0.70	39.29
BasketballDrive	1.47	70.02	2.01	65.69	1.72	42.74
BQTerrace	2.15	72.47	0.91	61.37	0.61	44.37
Average	0.48	74.01	0.68	66.67	0.4	43.25

TABLE IX: Overall average performance comparison of all four cases

Case	Proposed (%)		PAPS (%) [21]		FIICA(%) [26]	
	BDBR	TS	BDBR	TS	BDBR	TS
Case 1	-0.08	74.69	0.30	67.03	0.06	47.15
Case 2	-0.22	76.15	0.38	65.85	-0.18	45.75
Case 3	0.67	72.28	0.62	68.30	0.38	42.22
Case 4	0.48	74.01	0.68	66.67	0.40	43.25
Average	0.21	74.28	0.50	66.96	0.17	44.59

-0.18%, respectively. While the average TS of the proposed algorithm, PAPS and FIICA are 76.15%, 65.85% and 45.75% correspondingly. In the above two cases, the BDBR of the proposed algorithm is less than those of PAPS and FIICA. The coding speed of the proposed algorithm is significantly faster than the other two algorithms. In Table VII (case 3), the average BDBRs of the proposed algorithm, PAPS and FIICA are 0.67%, 0.62% and 0.38%, respectively. While the average TS of the proposed algorithm, PAPS and FIICA are 72.28%, 68.30% and 42.22% correspondingly. Compared with the other two algorithms, the BDBRs of the proposed algorithm is greater than those of PAPS and FIICA, and its coding speed is also faster than the other two algorithms. In Table VIII (case 4), the average BDBRs of the proposed algorithm, PAPS and FIICA are 0.48%, 0.68% and 0.40%, respectively. While the average TS of the proposed algorithm, PAPS and FIICA are 74.01%, 66.67% and 43.25% correspondingly. The BDBR of the proposed algorithm is less than that of PAPS and slightly greater than that of FIICA, meanwhile, the coding speed of the proposed algorithm is significantly faster than the other two algorithms.

In order to further demonstrate the overall performance of the proposed algorithm, Table IX provides the overall average performance comparisons among these three algorithms for all four cases.

The overall average BDBRs of the proposed algorithm, PAPS and FIICA are 0.21%, 0.50% and 0.17%, respectively. While the overall average TS of the proposed algorithm, PAPS and FIICA are 74.28%, 66.96% and 44.59% correspondingly. Therefore, the overall average coding speed of the proposed algorithm is significantly faster than the other two algorithms. Meanwhile, the overall average BDBR of the proposed algorithm is less than that of PAPS algorithm and slightly greater than that of FIICA algorithms. From Table IX, we can find that the main BDBR loss comes from case 3 and case 4 (scalability ratio 2x). The main reason is that ILR prediction in scalability ratio 2x doesn't predict very accurately. ILR mode interpolates a CU in BL to predict the co-located CU in EL. In scalability ratio 2x, since the resolution difference between BL and EL is very large, it leads to poor interpolation and hence the prediction of ILR mode is not very accurately accordingly. Therefore, the corresponding BDBR loss increases significantly in scalability ratio 2x.

Here are the main reasons why the proposed algorithm can effectively improve coding speed. First, we observe that the RD cost of ILR mode is different from that of Intra mode in all depths, and their distributions all follow Gaussian distribution. These features are very suitable for adopting GMM-EM to determine whether ILR is the best mode so as to skip Intra prediction early. Second, we investigate DM distributions and discover a special distribution feature: Class 0 and Class 1 only have 12 DMs, but around 85% CUs use DMs in these 2 classes. While Class 2 has 23 DMs, but only around 15% CUs use DMs in this class. This finding is extremely effective in coding speedup since it indicates that we have very high probability, around 85%, to find the best DM in Class 0 and Class 1, and thus can early terminate DM selection. Third, neighboring CUs, residual coefficients and RD costs are strongly related to depth selection. The CRF model is a discriminant probabilistic model with strong probabilistic reasoning ability. Exploiting the CRF Model in the proposed algorithm can effectively integrate neighboring CUs, residual coefficients and RD costs to early terminate depth selection.

Generally speaking, the improvement in coding speed will lead to an increase in BDBR, namely a decrease in coding efficiency. However, in Tables V-IX, there are occasions that coding efficiency increases, i.e., BDBR savings, compared with the SHM reference software. The reason can be explained as follows. In the Intra prediction process, CUs are predicted by their reference pixels. For the current CU, different methods lead to different neighbouring CUs and different reference pixels, which in turn lead to different prediction performance. We have analyzed the reason in detail [7].

VIII. CONCLUSION

In this paper, we fully investigate the special features for SSHVC and propose a new mode distribution-based Intra prediction algorithm for SSHVC, which includes the following unique features and strategies: (1) we observe that the RD costs between ILR mode and Intra mode are significantly different, and both their RD costs of follow Gaussian distribution. Based on this observation, we apply GMM-EM in machine learning to determine whether ILR mode is the best mode so as to skip Intra prediction; (2) when CUs use Intra mode, they are normally very simple and usually use simple DMs. We investigate their DM distributions and divide all DMs into three classes, and then develop their corresponding approaches to progressively predict the best DM so as to skip many unlikely DMs; (3) neighboring CUs, residual coefficients and RD costs are strongly related to depth selection. We efficiently exploited simultaneously them to early terminate depth selection. Experimental results show that the proposed algorithm can improve the coding speed significantly with negligible coding efficiency losses. Deep learning has been a hot research topic recently, we will plan to explore it to further improve the coding speed in our future work [33].

REFERENCES

[1] J. M. Boyce, Y. Yan, J. 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.

[2] 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., Vancouver, BC, Canada, May 2013, pp. 1744-1748.

[3] Y. T. Kuo, P. Y. Chen, and H. C. Lin, “A spatiotemporal content-based CU size decision algorithm for HEVC,” *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 100–112, Mar. 2020.

[4] H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos, “An encoder complexity reduction scheme for Quality/Fidelity scalable HEVC,” *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 664–674, Sep. 2016.

[5] D. Wang, Y. Sun, W. Li, C. Zhu, and F. Dufaux, “Fast inter mode predictions for SHVC,” *in Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 1696–1701.

[6] D. Wang, Y. Sun, C. Zhu, W. Li, and F. Dufaux, “Fast depth and inter mode prediction for quality scalable high efficiency video coding,” *IEEE Trans. Multimedia.*, vol. 22, no. 4, pp. 833-845, Apr. 2020.

[7] D. Wang, Y. Sun, C. Zhu, W. Li, F. Dufaux, and J. Luo, "Fast Depth and Mode Decision in Intra Prediction for Quality SHVC," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 6136 - 6150, Dec. 2020.

[8] 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.

[9] I.-M. Pao and M.-T. Sun, “Modeling DCT coefficients for fast video encoding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.

[10] 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.

[11] D. Wang, Y. Sun, J. Liu, F. Dufaux, X. Lu and B. Hang, “Probability-Based Fast Intra Prediction Algorithm for Spatial SHVC,” *IEEE Trans. Broadcast.*, vol. 68, no. 1, pp. 83-96, Mar. 2022.

[12] 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.

[13] W. Jiang, H. Ma, and Y. Chen, “Gradient based fast mode decision algorithm for intra prediction in HEVC,” *in Proc.2nd Int. Conf. Consumer. Electron.*, Commun. Netw., Apr.2012, pp.1836-1840.

[14] T. Zhang, M.T. Sun, D. B. Zhao, and W. Gao, “Fast Intra-Mode and CU Size Decision for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1714–1726, Aug. 2017.

[15] M. Jamali, S. Coulombe, and F. Caron, “Fast HEVC intra mode decision based on edge detection and SATD costs classification,” *in Proc. Data Compress. Conf.*, Snowbird, UT, USA, Apr. 2015, pp. 43-52.

[16] H. Zhang, and Z. Ma, “Fast Intra Mode Decision for High-Efficiency Video Coding (HEVC),” *IEEE Trans. Circuits Syst. Video Technol.*, vol.24, no.4, pp.660-668, Apr. 2014.

[17] M. Jamali, and S. Coulombe, “Fast HEVC Intra Mode Decision Based on RDO Cost Prediction,” *IEEE Trans. Broadcast.*, vol.65, no.1, pp.109-122, Mar. 2019.

[18] Y. Li, G.B Yang, Y. Song, H.L Zhang, X.L Ding, D.Y Zhang, “Early Intra CU Size Decision for Versatile Video Coding Based on a Tunable Decision Model,” *IEEE Trans. Broadcast.*, vol. 67, no. 3, pp. 710-720, Sep. 2021.

[19] J. J Lei, D.Y Li, Z.Q Pan; Z.Y Sun; S. Kwong; C.P Hou, “Fast Intra Prediction Based on Content Property Analysis for Low Complexity HEVC-Based Screen Content Coding,” *IEEE Trans. Broadcast.*, vol. 63, no. 1, pp. 48-58, Mar. 2017.

[20] M. Xu, T.Y Li, Z.L Wang, X. Deng, R. Yang, and Z.Y Guan, "Reducing Complexity of HEVC: A Deep Learning Approach," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5044-5059, Oct. 2018.

[21] 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.

[22] L. W Zhu, Y. Zhang, S. Kwong, X. Wang, T.S Zhao, “Fuzzy SVM-Based Coding Unit Decision in HEVC,” *IEEE Trans. Broadcast.*, vol. 64, no. 3, pp. 681-694, Oct. 2018.

[23] L. Shen and G. Feng, “Content-based adaptive SHVC mode decision algorithm,” *IEEE Trans. Multimedia.*, vol.21, no. 11, pp. 2714 -2725, Apr. 2019.

[24] L. Shen, G. Feng and P. An, "SHVC CU Processing Aided by a Feedforward Neural Network", *IEEE Trans. Industrial Informatics.*, vol. 15, no. 11, pp. 5803-5815, Apr. 2019.

[25] X. Lu, C. Yu, Y.F Gu, and G. Martin, "A fast Intra coding algorithm for spatial scalability in SHVC,” *in Proc. IEEE Int. Conf. Image Processing (ICIP)*, Athens, Greece, Oct. 2018, pp. 1792-1796.

[26] X. Lu, C.Yu and G.R.Martin, "Fast intra-and inter-coding algorithms for the spatially scalable extension of H.265/HEVC," *Multimedia Tools and Applications*, vol. 79, pp. 26447–26465, Sep. 2020

[27] H. Li Tan, C. C Ko, and S. Rahardja, “Fast Coding Quad-Tree Decisions Using Prediction Residuals Statistics for High Efficiency Video Coding (HEVC),” *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 128-133, Mar. 2016.

[28] H. Yuan, C. Guo, J. Liu, X. Wang, S. Kwong, “Motion-homogeneous-based fast transcoding method from H.264/AVC to HEVC,” *IEEE Transactions on Multimedia.*, vol. 19, no. 7, pp. 1416-1430, Jul. 2017.

[29] H. Yuan, Y. Chang, Z. Lu, M. Li, “Fast and efficient intra prediction method for H.264/AVC,” *Optical Engineering.*, vol. 49, no. 4, pp. 1-3, Apr. 2010.

[30] Common SHM Test Conditions and Software Reference Configurations, document JCTVC-Q1009, ITU-T SG 16 WP 3 ISO/IEC JTC1/SC29/WG., Mar. 2014.

[31] G. Bjontegaard, “Calculation of average PSNR difference between RD-curves,” *in Proc. 13th VCEG-M33 Meeting*, Austin, TX, USA, Apr. 2001, pp. 2-4.

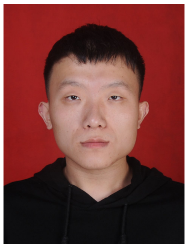
[32] Shiliang Sun, Ziang Dong, and Jing Zhao, “Conditional Random Fields for Multiview Sequential Data Modeling,” *IEEE Transactions on Neural Networks and Learning Systems.*, vol. 33, no. 3, pp. 1242-1253, Mar. 2022.

[33] C. Li, S. Li, Y. Gao, X. Zhang, and W. Li, “A Two-stream Neural Network for Pose-based Hand Gesture Recognition,” *in IEEE Transactions on Cognitive and Developmental Systems.*, doi: 10.1109/TCDS.2021.3126637.

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. His main research interest is video coding.

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.

Weisheng Li received the B.S. and M.S. degrees from the School of Electronics and Mechanical Engineering, Xidian University, Xi’an, China, in 1997 and 2000, respectively, and the Ph.D. degree from the School of Computer Science and Technology, Xidian University, in 2004. He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research focuses on intelligent information processing and pattern recognition.



Lele Xie Lele Xie received the B.Sc. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2018. He is currently pursuing the M.Sc. degree in Chongqing University of Posts and Telecommunications. His research interest is video coding.

Xin Lu received the B.Sc. and M.Sc. degrees from the Harbin Institute of Technology, Harbin, China, in 2008 and 2010, respectively, and the Ph.D. degree in computer science from the University of Warwick, Coventry, U.K., in 2013. He is currently a lecturer with the School of Computer Science and Informatics, De Montfort University. His current research interests include video coding standards, data compression, and image or video processing.

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. Frederic is a Fellow of IEEE. He was Chair of the IEEE SPS Multimedia Signal Processing (MMSP) Technical Committee in 2018 and 2019. He is a member of the IEEE SPS Technical Directions Board. He was Vice General Chair of ICIP 2014, General Chair of MMSP 2018, and Technical Program co-Chair of ICIP 2019. He will be Technical Program co-Chair of ICIP 2021. He is also a founding member and the Chair of the EURASIP Technical Area Committee on Visual Information Processing.

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 CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON BROADCASTING, 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 a few special issues in international journals, including as a Guest Editor in the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is an APSIPA Distinguished Lecturer (2021-2022), and was also an IEEE Distinguished Lecturer of Circuits and Systems Society (2019-2020). He is a co-recipient of multiple paper awards at international conferences, including the most recent Best Demo Award in IEEE MMSP 2022, and the Best Paper Runner Up Award in IEEE ICME 2020.