IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 September 2022, accepted 10 October 2022, date of publication 17 October 2022, date of current version 25 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215163

RESEARCH ARTICLE

Machine Learning-Based Early Skip Decision for Intra Subpartition Prediction in VVC

JEEYOON PARK, (Student Member, IEEE), BUMYOON KIM, (Student Member, IEEE), JEEHWAN LEE, (Student Member, IEEE), AND BYEUNGWOO JEON, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, Sungkyunkwan University, Jangan-gu, Suwon 16410, South Korea

Corresponding author: Byeungwoo Jeon (bjeon@skku.edu)

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) through the Ministry of Science and ICT under Grant NRF-2020R1A2C2007673; and in part by the System LSI Division, Samsung Electronics Company Ltd.

ABSTRACT The recently published video coding standard, Versatile Video Coding (VVC/H.266), has the intra subpartition (ISP) coding mode, which divides an intra-predicted block into smaller blocks called subpartitions, each of which can be predicted using the newly reconstructed subpartition while still sharing the same intra mode. It is a VVC intra prediction tool that brings significant coding gains but also increases its encoding complexity. In this context, this paper addresses how to speed up the ISP encoding process by designing an ISP early skip decision scheme using a simple LightGBM model. The proposed ISP decision expedites the encoding process by early determination of whether or not to skip the ISP mode test. The proposed method uses the mean absolute sum of transform coefficients as a key feature. Our experimental results show an average encoding time saving of 7.2% under the all intra coding configuration with 0.08% BDBR loss. Compared to the state-of-the-art methods, our solution is able to outperform related works in terms of the combined rate-distortion and time saving.

INDEX TERMS VVC, intra prediction, fast intra prediction, H.266/VVC, encoder optimization, intra subpartition (ISP), light gradient boosting machine (LightGBM).

I. INTRODUCTION

Along with the reicent commercial introduction of 5G mobile infrastructure, unconventional media, such as 360-degree video/VR or immersive media providing up to 6 DoF (degrees of freedom), have started to emerge as new business opportunities (in addition to well-known HD, 4K, and 8K video). But all of these types of media carry a large amount of data, causing explosive video traffic. This demands a very powerful video coding technique that can provide very high compression performance.

Versatile Video Coding (VVC) [1], [2], [3] is the latest video coding standard by the Joint Video Experts Team (JVET), jointly formed by the Moving Picture Experts Group (ISO/IEC MPEG) and the Video Coding Experts Group (ITU-T VCEG), and provides more than twice the compression performance compared to the High Efficiency Video Coding (HEVC) standard [4]. It has many advanced coding tools compared to HEVC. It is reported [5], [6], [7] that the coding efficiency of VVC surpasses that of HEVC, with an average bitrate savings of 25.06% (all intra (AI) case), 41.04% (random access (RA) case), and 30.88% (low delay - B (LDB) case) at the same video quality. However, it is also noted that its encoding time has increased significantly by 26, 8, and 6 times against HEVC AI, RA, and LDB, respectively.

Intra coding is a method of encoding a given block through intra prediction referring to samples already reconstructed in the same picture [8]. It is reported [8] that VVC includes many powerful intra-coding tools, such as mode dependent intra smoothing (MDIS) [9], cross-component linear model (CCLM) [10], position dependent intra prediction combination (PDPC) [11], multiple reference line (MRL) [12], [13] intra prediction, intra subpartition (ISP) [14], [15], and

The associate editor coordinating the review of this manuscript and approving it for publication was Chaker Larabi.

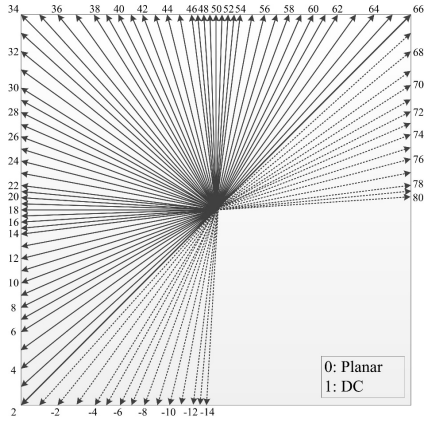

FIGURE 1. Intra prediction modes for luma block in VVC [1].

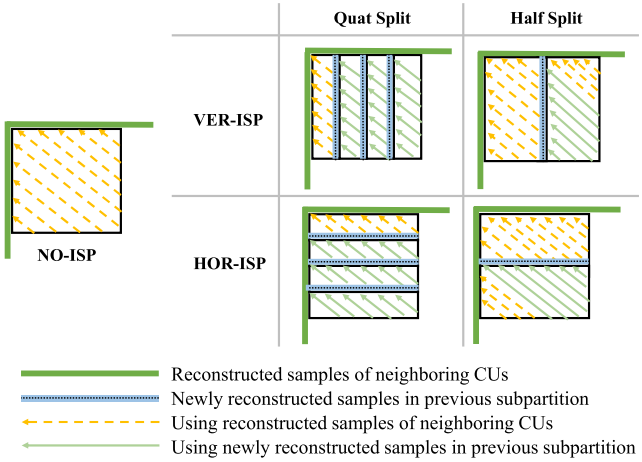

FIGURE 2. ISP mode in VVC [1].

matrix-based intra prediction (MIP) [17]. As shown in Fig. 1, VVC supports up to 95 intra prediction modes, among which 28 modes are referred to as wide angular intra prediction (WAIP) modes [18] while 65 modes are general angular modes. It also has DC and planar modes as non-directional intra modes. For intra prediction, VVC achieves 25.06% of coding efficiency improvements but requires 26 times of encoding time compared to HEVC [6]. The optimal intra prediction modes are determined through a complex search process that involves recursive block partitioning and testing of various predictions for each block, which greatly increases the coding complexity. From a practical point of view, a substantial reduction in coding complexity can help widespread use of the new coding standard. In this regard, many researchers have studied coding complexity reduction of VVC intra prediction for fast VVC encoding.

The ISP [14], [15], [16] is an efficient VVC intra prediction tool. As shown in Fig. 2, the ISP divides a luma intra prediction block equally into two or four smaller blocks. These are called subpartitions each of which is predicted using the same intra mode. [16] describes the ISP scheme implemented in VVC test model (VTM) which also has various early termination strategies to reduce the complexity of the ISP encoding search process. Even after the much enhanced ISP encoder search solution was implemented, however, efforts to minimize ISP complexity while maintaining ISP coding efficiency have continued. Park et al. [19] proposed a fast algorithm that limits the use of ISP by focusing on the reference samples used for each subpartition block when the ISP is applied. In other words, if a block is not predicted using closer reference samples by ISP, its ISP mode test is skipped to make the encoder faster. An optimization scheme for fast ISP coding mode is also proposed based on the CU texture complexity [20], [21], [22]. They measure the CU block texture complexity to determine whether a CU needs to use the ISP mode or not, so as to achieve faster encoding. We note that previous fast ISP decision approaches [19], [20], [21], [22] can reduce the overall encoding time by effectively avoiding unnecessary rate distortion optimization (RDO) processes by fast intra mode decision through characterization of the information of each block. However, we also note that those previous approaches [19], [20], [21], [22] considered only the intra prediction direction and the texture of the block itself; that is, they missed due consideration of the benefit of performing separate transforms for each subpartition.

Meanwhile, since machine learning is a recent viable method to reduce encoder complexity with a small influence on coding efficiency, there are several studies on fast decision making processes by implementing learning-based algorithms [23], [24], [25], [26], [27], [28]. Dong et al. [22] used decision tree (DT) [37] model for designing a fast ISP mode skip method. But only CU texture complexity is considered as in previous works [19], [20], [21]. While the goal of this paper is to design a fast ISP search scheme (the same goal as previous approaches), the proposed ISP Early Skip Decision (ISP-ESD) scheme also makes early determinations on whether or not to test ISP mode in the RDO process by considering the efficiency facilitated by ISP prediction and transforming each subpartition individually. Moreover, the proposed method uses Light Gradient Boosting Machine (LightGBM) classifiers [38]. Therefore, our solution is the first machine learning-based fast ISP search algorithm that takes both aspects of prediction and transform into consideration. In this paper, in comparison with the ISP tool-off test in VTM, the proposed method reduces the encoder run-time of ISP from 13.8% to 7.2% (i.e., about 50% reduction) in exchange for a loss of 0.08% BD-Rate.

The main contributions of this work are:

- New and efficient VVC ISP intra prediction complexity reduction solution.
- Use of efficient LightGBM model to reduce the complexity of the ISP mode test while minimizing the coding efficiency loss.
- Define key features and use them for machine learning classifiers.
- The proposed ISP-ESD implementation is independent of the quantization parameter (QP) setting.

The remainder of this paper is organized as follows. Section II describes the process of the ISP scheme in VVC. Section III explains the motivation for ISP early skip decision method. In Section IV, the proposed machine learning-based ISP-ESD scheme is explained in detail. Subsequently, the simulation results are shown in Section V. Finally, Section VI concludes the paper.

II. ISP PREDICTION SCHEME IN VVC

As the first step of encoding video, each picture is partitioned into coding units (CUs) of various shapes and sizes. How a picture is partitioned into CUs is represented in a tree structure, and the tree information is transmitted to a decoder. CUs represent a group of pixels, which are encoded in the same coding mode. A larger CU is desirable in reducing the signaling overhead of the coding mode and relevant information, but it may cause prediction performance loss unless all the pixels in the CU are either homogeneous (in intra prediction) or well represented by a motion vector (in inter prediction). Especially in intra prediction, a larger CU inevitably means a larger distance from the reference samples in neighboring CU blocks; this tends to decrease the accuracy of intra prediction. In return, a smaller CU can enhance intra-prediction accuracy, but it increases signaling overhead due to the increased number of CUs in a picture. In order to solve this dilemma, under ISP mode, an intra-coded block is subdivided into smaller blocks that still share the same intra prediction mode. ISP performs intra prediction for each subpartition using nearer reconstructed reference samples in already encoded subpartition blocks. In VVC, the regular intra modes, i.e., planar, DC, and all angular modes, can be used with ISP.

A. BLOCK SUBPARTITION IN ISP SCHEME

As shown in Fig. 2, under the ISP mode, a CU can be split into four subpartitions either horizontally (HOR-ISP) or vertically (VER-ISP), where the subpartition direction is indicated by the two ISP flags (Table 1). It should be noted that due to practical considerations of memory access, the partitioning is carried out in such a way that there are at least 16 samples per subpartition [16]. Therefore, ISP is not applied to 4×4 CUs. Additionally, in the case of 4×8 or 8×4 CUs, a CU is divided only into two blocks (called a half split) instead of four. For the other sizes, a CU is divided into four subpartitions of the same shape and size (called a quad split). Furthermore, to avoid writing narrow blocks of data to memory, the minimum width of an intra prediction is four samples. Therefore, when the VER-ISP mode is used for a CU with a width of four, the partition is not made in prediction process, but is still made in transform process [16].

B. TRANSFORM IN ISP SCHEME

ISP is related not only to intra prediction but also to the transform. VVC has two types of transforms. One is the primary transform whose kernel is selected among DCT-II and DST-VII separately for horizontal and vertical directions [29], [30]. The other is the secondary transform, which is the low-frequency non-separable transform (LFNST), obtained by offline training with intra-prediction residuals [29], [31]. While the selection of a primary and secondary transform in VVC is signaled by a CU-level signal, *mts_idx,* and *lfnst_idx*, under the ISP mode, it is signaled implicitly by a CU-level signal, *lfnst_idx*, which indicates the primary and secondary transforms for the CU, as in Table 2. If *lfnst_idx* is 0, a primary transform is selected based on the width (or height) of a subpartition, and the secondary transform is not used. If *lfnst_idx* is either 1 or 2, then, DCT-II is used as the primary transform. In addition, *lfnst_idx* is signaled for a CU block; thus, the same LFNST transform kernel is utilized for all the subpartitions that have a non-zero coded block flag (CBF) [32].

TABLE 1. IntraSubPartitionsSplitType and related flags [2].

IntraSubPartitionsSplitType	NO-ISP	HOR-ISP	VER-ISP
intra_subpartitions_mode_flag	0	1	1
intra_subpartitions_split_flag	N/A	0	1

TABLE 2. Implicit transform selection for ISP.

lfnst_idx	Primary transform	Secondary transform
0	DST-VII (L=4,8,16) DCT-II (L=2,32,64)	Not used
1 or 2	DCT-II	LFNST with idx=1 LFNST with idx=2

*(L: Width or Height)

C. ENCODER SEARCH SCHEME OF ISP MODE

The ISP search is carried out to select the best ISP coding mode for each CU block to encode. This search decides the best intra prediction mode and whether ISP mode is selected or not. If ISP is selected, it also determines whether its split is vertical or horizontal. This ISP test evaluates RD cost of a combination (*mode, split, lfnst*). Here, "*mode*" refers to the intra mode (planar, DC, and all angular modes in Fig. 1); "*split*" the ISP split direction, which are HOR-ISP and VER-ISP; and "*lfnst*" indicates whether or not to use LFNST (whether the index of LFNST is 0, 1, or 2). The RD cost of each combination (*mode, split, lfnst*) is obtained as a cumulative sum of the RD costs of each subpartition. A detailed technical description on how to configure the list for the ISP mode test, the ISP encoder search process, early termination steps, and rules used to skip the ISP test from RDO process can be found in [16].

III. MOTIVATION

The benefits of ISP come not only from better intra prediction but also from better utilization of the correlation between pixels within each subpartition by transform. Since intra prediction can exploit closer reconstructed samples in previous subpartitions, significant accuracy improvement is expected in predictor generation [14]. In this regard, the

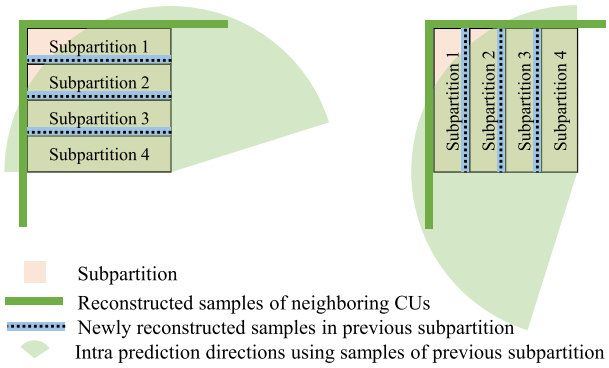

FIGURE 3. ISP mode case where newly reconstructed samples are used.

authors in [19] proposed an ISP-PPC scheme (ISP with pre-pruned candidates), which adds an intra prediction mode to the HOR-ISPList (intra mode candidate list for HOR-ISP) or VER-ISPList (intra mode candidate list for VER-ISP) only when the mode can utilize closer reconstructed samples in consideration of the directions of split and prediction.

However, caution should be taken; apart from the accuracy improvement in prediction, ISP still benefits much from the transform. This motivated us to pay more attention to the advantage of the individual transform on the subpartitions in the ISP mode. Fig. 3 shows the intra angular mode directions in the case of HOR-ISP, VER-ISP where newly reconstructed samples from the previous subpartitions are used respectively. In the state where the intra mode is determined, the encoder can decide whether or not to use an ISP mode for the current CU through additional RD calculation. Furthermore, when the ISP mode is selected, there is a burden to also signal whether it is HOR-ISP or VER-ISP, as mentioned above. If ISP benefited only from better prediction due to the newly reconstructed closer reference samples, it would not be selected in cases of CUs whose intra prediction direction indicates using only neighboring CUs as reference samples. Fig. 4 shows the percentages of two cases in CUs encoded in ISP: one is the case where the intra prediction is made using the samples in neighboring CUs (just like the case where the ISP mode is not used, i.e., NO-ISP as shown in Fig. 2), and the other is the case where the prediction is made using the reconstructed pixels in the previous subpartition. The conditions for using newly reconstructed samples in the previous subpartitions are specified in Table 3. This clearly demonstrates that there are non-trivial cases in which ISP is preferred rather than NO-ISP even if the predictor calculated under ISP mode is the same as that of NO-ISP. For example, in the class E (1280 × 720) sequences, the percentage of intra prediction using reconstructed samples in neighboring CUs is as high as 41%. This means that ISP mode is used irrespective of whether or not to use the reference from the previous subpartition. Furthermore, because of hardware implementation issues, the width of the prediction block is maintained to be not smaller than 4. Therefore, if the width of a subpartition is smaller than four, prediction blocks (PBs)

TABLE 3. Condition of using newly reconstructed samples in previous subpartition.

ISP Mode	IntraMode	BlockSize
HOR-ISP	IntraMode > 18 (HOR)	For all
	Planar	
	DC	$W \geq H/n$
VER-ISP	IntraMode < 50 (VER)	$W > 4$
	Planar	
	DC	$W > 4$ *and* $H \geq \max(4, W/n)$

n: The number of subpartitions

TABLE 4. Percentage (%) of VER-ISP with a CU width of 4 in all video sequences [40].

Video class	A1	A2	B	C	D	E	F
CU width of 4	15.3	13.1	14.6	27.0	34.4	28.1	23.1

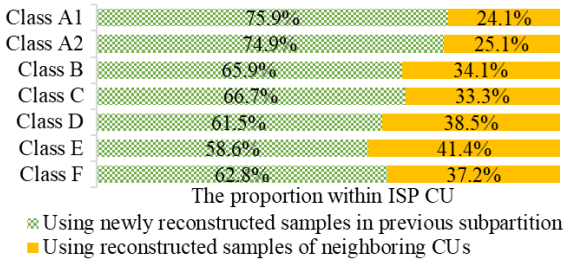

FIGURE 4. ISP distribution by type of reconstructed samples used in all test video sequences [40].

are resized to make the width at least 4. For example, in VER-ISP, when the width of the CU block is four, the whole CU block is predicted at once using the reconstruction reference samples. In this case, a block cannot benefit from prediction by using the ISP mode. Table 4 summarizes the percentage of CU blocks with a width of 4 among CUs encoded in VER-ISP, which is quite non-trivial. This confirms that the ISP mode is selected even when a newly reconstructed reference sample is not structurally available. Thus, it is confirmed that although the newly reconstructed reference samples in the previous subpartition cannot be used for prediction, the ISP mode can still be selected for transform purpose.

Regarding the possible benefit from smaller transform, it is worthwhile to scrutinize Table 5 and Fig. 5. Table 5 shows an example of CBF distribution of CU blocks encoded in the ISP mode. The CBF signals whether a non-zero transform coefficient exists in a given block or not. If the CBF is 0, there is no non-zero transform coefficient in the block thus no residual coding. A better prediction will generate more cases of CBF=0 [32]. Table 5 shows the percentages of CU blocks encoded in the ISP mode that have subpartitions with and without CBF=0 cases (where all subpartitions with CBF=0 are excluded because an ISP CUs should have at least one subpartition with CBF=1). It is observed that there are many cases in which the ISP mode is used even though there are no subpartitions with CBF=0. For example, in classes C and D, more than half of the CU blocks coded in ISP

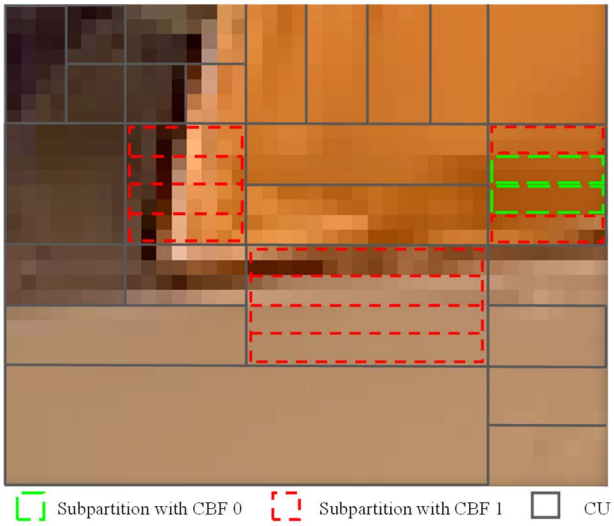

FIGURE 5. Example of CBF distribution of CU blocks encoded in the ISP mode (partyscene video, (QP=27, 1st frame), CUs at (224,392) to (264, 424)).

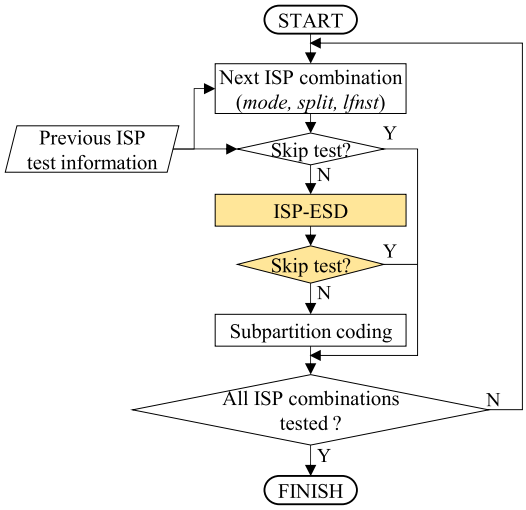

FIGURE 6. Flowchart of the ISP-ESD operation incorporated into the [16] ISP test flowchart (yellow color signifies ISP-ESD).

with a half split do not have any subpartitions with CBF=0. In general, the transform gain increases as the block size increases. However, this is true only if the block is not too heterogeneous. If all subpartitions are with CBF=1, the advantage of not performing residual coding cannot be taken, and the transform efficiency may decrease under ISP mode (because the block is divided into blocks smaller than the CU block). It is not favorable to ISP mode. Nevertheless, there are many cases of ISP decision even when all subpartitions have CBF=1. It is because too-heterogeneous CU blocks can be transformed after they are divided into homogeneous subpartitions as much as possible by ISP. In this case, the process of transforming each subpartition is more advantageous in coding efficiency. The ISP mode can benefit both from availability of newly reconstructed near reference samples and transforming each subpartition separately to better exploit pixel correlation within each subpartition. Based on these observations, we design an ISP-ESD scheme that determines whether or not to test a given ISP mode by taking into account the benefits of prediction and transform together. More information about the design of ISP-ESD scheme and the selection of input features will be explained in detail in the next section.

IV. PROPOSED METHOD

The proposed ISP-ESD scheme is modeled as a binary classification where the decision is either 0 (which means the corresponding ISP mode test should be conducted) or 1 (which means the corresponding ISP mode test should be skipped). Its solution can be obtained by applying various approaches, including deep neural networks [33], K-nearest neighbors [34], stochastic gradient descent [35], or support vector machines [36]. The LightGBM [38] is a strong machine learning algorithm that has demonstrated significant success in a wide range of applications. It is built on DT classifiers [37], which provide high accuracy and hardware affinity. Since the goal of this paper is to reduce the coding complexity, we choose to take a very efficient LightGBM approach so that the classifier handles a vast amount of data quickly. In this sense, we propose a statistical learning solution of LightGBM-based ISP-ESD scheme to remove unnecessary ISP mode testing for a combination (*mode, split, lfnst*) one by one that is not likely to be chosen as the best. Also, once the tree is trained, there is no need to include additional libraries or graft new networks to implement the solution in VVC encoder. Therefore, it can be easily implemented in software and hardware VVC encoders [38]. Fig. 6 illustrates how the proposed ISP-ESD scheme can be easily implemented, for an example, by incorporating it on existing ISP encoder search process in VTM. It shows the step of the ISP-ESD scheme inserted into the ISP test flowchart in [16], where the yellow boxes indicate the process of ISP-ESD scheme. When a combination (*mode, split, lfnst*) under ISP test is passed to the ISP-ESD, its ISP test can be skipped depending on the output of ISP-ESD, that is, "skip" (i.e., 1) or "not skip" (i.e., 0). The proposed ISP-ESD consists of an ISP-ESD-H classifier for HOR-ISP and an ISP-ESD-V classifier for VER-ISP. The effectiveness of the classifier is highly related to the relevance of the training data set [38]. In this section, several features that have strong influence on whether the ISP mode should be tested or not are considered and checked based on statistical analysis. By applying

TABLE 5. CBF distribution (%) inside ISP-encoded cu blocks.

ISP Split Type	CBF	Video Sequences [40]						
		A1	A2	B	C	D	E	F
Quad Split	All "CBF = 1"	25	17	25	35	38	21	47
	"CBF = 0" exists	75	83	75	65	62	79	53
Half Split	All "CBF = 1"	37	32	45	57	61	40	44
	"CBF = 0" exists	63	68	55	43	39	60	56

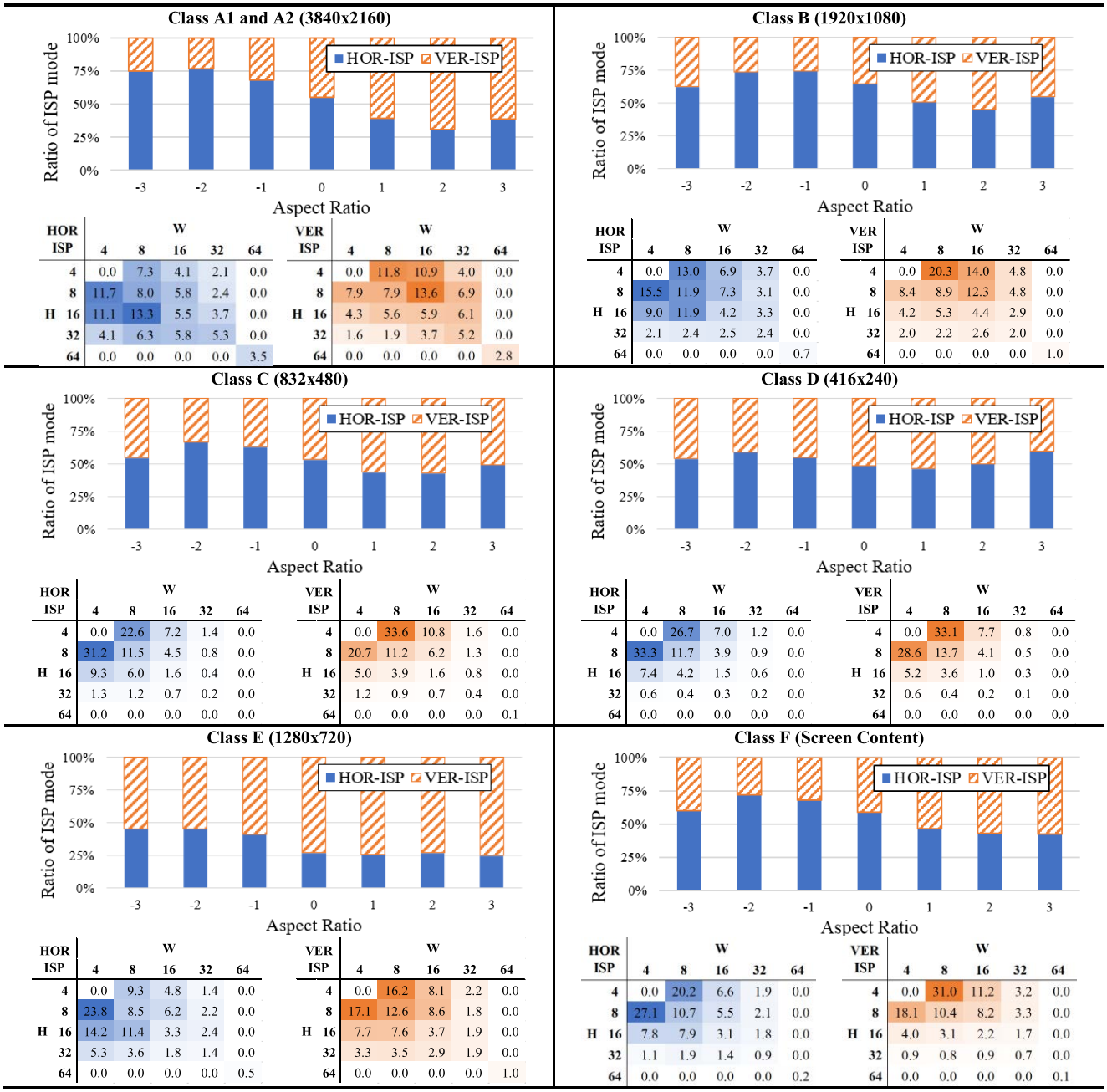

FIGURE 7. Statistical distribution of ISP modes with respect to block size and block aspect ratio. Top: ISP mode distribution by aspect ratio, bottom: ISP mode distribution by block size (%).

the transform to the subpartition block in advance, features that can check the transform effect in each subpartition are calculated. All experimental analyses followed the standard common test conditions [40]. In addition, the result of each class is obtained by averaging those results by QP 22, 27, 32, and 37.

A. ANALYSIS ON INPUT FEATURES: BLOCK SIZE AND ASPECT RATIO

The ISP mode distribution with respect to block size and block aspect ratio in Fig. 7 reveals that there is a certain likelihood of choosing the ISP mode. It shows three distributions for each video sequence, including the ISP mode distribution among all CU blocks encoded in the ISP mode. The bar graph shows the ratio of the ISP mode according to the aspect ratio of a block (here the aspect ratio is calculated as $\log_2 width - \log_2 height$). If a block aspect ratio is greater than 1 (that is, horizontally elongated), VER-ISP is more likely than HOR-ISP. This observation is even clearer as the aspect ratio increases. It is also seen that if a block is vertically elongated, HOR-ISP is more likely than VER-ISP. This becomes also clearer as the aspect ratio further decreases. Therefore,

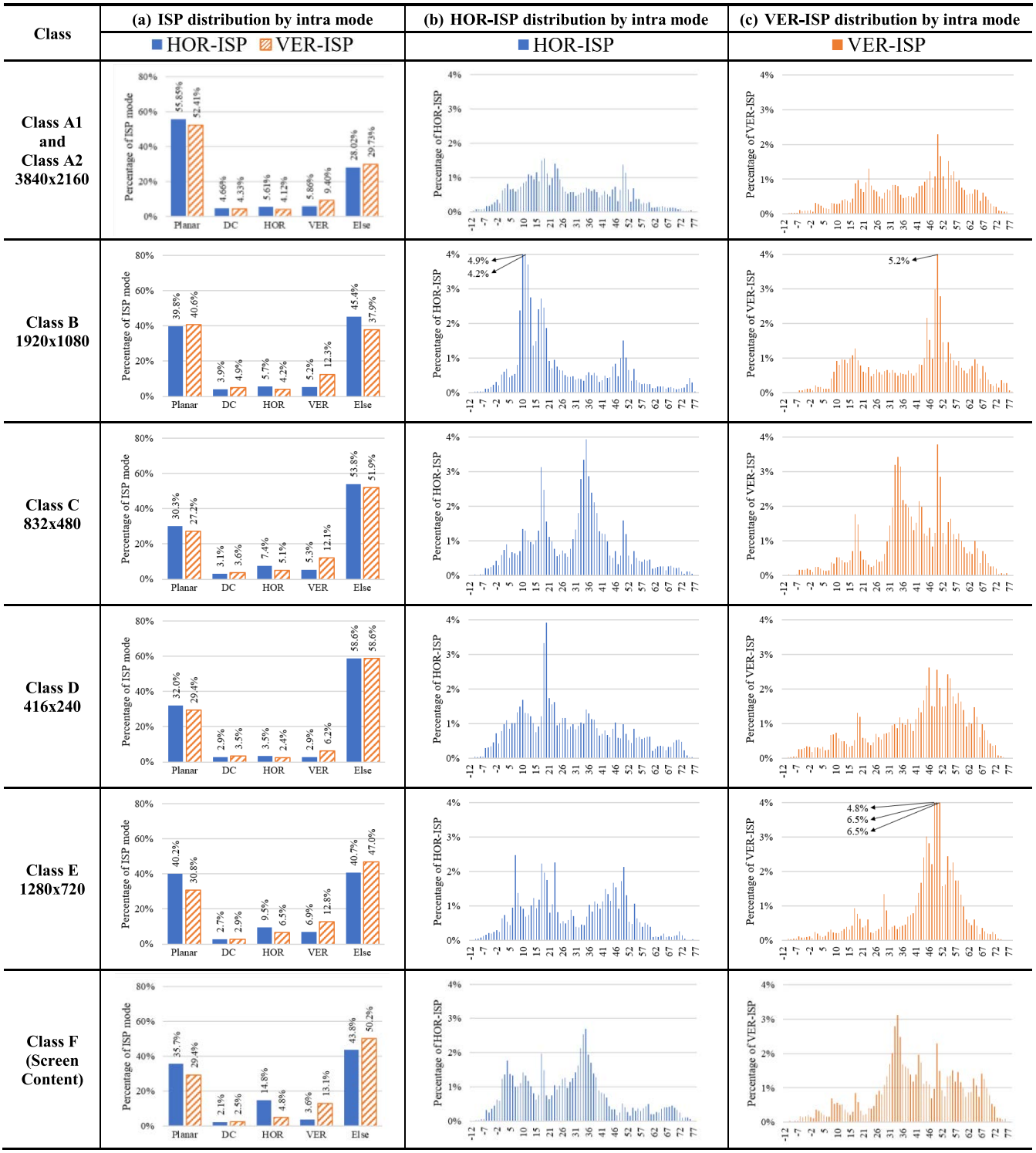

FIGURE 8. Statistical distribution of ISP modes with respect to intra prediction mode. (a) ISP distribution by default intra mode, (b) HOR-ISP distribution by intra mode, and (c) VER-ISP distribution by angular intra mode.

the aspect ratio certainly affects the ISP mode decision. Two other tables show the ISP mode distribution color map for HOR-ISP and VER-ISP of each class. A dark color indicates that the number of corresponding CUs encoded in ISP mode is large. It can be seen that these are more concentrated in small blocks irrespective of HOR-ISP and VER-ISP. As a result, it can be said that both the block size and the aspect ratio have a non-trivial influence on whether ISP mode is

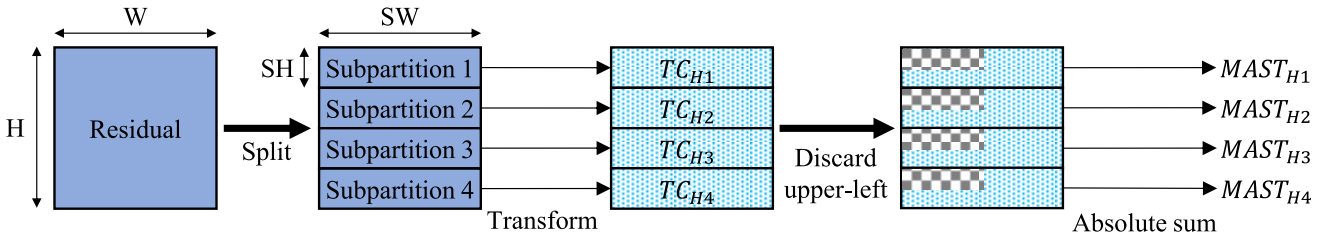

FIGURE 9. Process for calculating MAST for HOR-ISP with quad split ($MAST_H$).

selected or not. Therefore, the block size and its aspect ratio are considered to be the basic features to predict the ISP mode decision.

B. ANALYSIS ON INPUT FEATURES: INTRA PREDICTION MODE

The statistical distribution of intra mode for CUs encoded in ISP mode (HOR-ISP and VER-ISP) is provided in Fig. 8. Since the four default modes (planar, DC, HOR (horizontal), VER (vertical)) are dominantly selected modes, it is observed that the distribution of the intra prediction mode in the ISP mode is skewed to them. In general, the planar mode is one of the most frequently selected modes, and since newly reconstructed samples can be used for prediction when ISP is applied (regardless of the ISP mode), as shown in Fig. 8 (a), the planar mode is the most common intra mode for ISP mode. Next DC, HOR, and VER follow. One thing to note is that, regardless of the video class, HOR-ISP is more concentrated in the case of HOR intra mode, while VER-ISP is more concentrated in the case of VER intra mode. Fig. 8(b) shows HOR-ISP distribution by angular intra mode. Additionally, Fig. 8(c) shows VER-ISP distribution by angular intra mode. Note that the angular mode is the intra angular prediction mode indicated in Fig. 1. In the case of HOR-ISP, it is observed that in most of the video classes, the intra prediction mode are distributed in the horizontal directions. Also, in the case of VER-ISP, it is observed that the intra prediction mode is mainly distributed in the vertical direction (35-66) regardless of the video class. Both HOR-ISP and VER-ISP have non-uniform distribution in the intra mode. Therefore, the intra prediction mode can be also considered to be an effective feature for deciding whether the ISP mode will be used or not.

C. ANALYSIS ON INPUT FEATURES: MAST

The energy compaction property of transform generally tends to make the high-frequency transform coefficients have small values. It suggests that the absolute sum of transform coefficients at high frequency locations can reflect how well the energy is compacted by transform. Many high-frequency coefficients of non-trivial magnitude indicate that the energy inside the block may not be well compacted [29]. As mentioned in Section III, ISP tightly interacts not only with intra prediction but also with the transform. If benefit in coding efficiency is much expected from transforming each partition, ISP mode can be early chosen. Therefore, we try to investigate the coding efficiency of the ISP mode in advance by performing prediction on an entire CU block and then simply applying the primary transform for each subpartition following Table 2. To do this, we use the mean absolute sum of transform coefficients ($MAST$) based on subpartitions as a key feature for deciding whether the ISP mode will be used or not. The formulas for calculating $MAST$ for each ISP mode are as following:

$$MAST_{Hn} = \begin{cases} \frac{4}{3N} \cdot \left(\sum_{j=1}^{SH} \sum_{i=1}^{SW} |TC_{Hn}(i,j)| - \sum_{j=1}^{\frac{SH}{2}} \sum_{i=1}^{\frac{SW}{2}} |TC_{Hn}(i,j)| \right), & SH \neq 1 \\ \frac{2}{N} \cdot \sum_{i=SW/2+1}^{SW} |TC_{Hn}(i,1)|, & SH = 1 \end{cases} \tag{1}$$

$$MAST_{Vn} = \begin{cases} \frac{4}{3N} \cdot \left(\sum_{j=1}^{SH} \sum_{i=1}^{SW} |TC_{Vn}(i,j)| - \sum_{j=1}^{\frac{SH}{2}} \sum_{i=1}^{\frac{SW}{2}} |TC_{Vn}(i,j)| \right), & SW \neq 1 \\ \frac{2}{N} \cdot \sum_{j=SH/2+1}^{SH} |TC_{Vn}(1,j)|, & SW = 1 \end{cases} \tag{2}$$

SW and SH are the width and height of a subpartition, respectively, and $TC_{Hn}(i,j)$ is a transform coefficient of the n_{th} subpartition for HOR-ISP at (i,j). In the same way, $TC_{Vn}(i,j)$ is a transform coefficient of the n_{th} subpartition for VER-ISP at (i,j). i and j ($1 \leq i \leq SW, 1 \leq j \leq SH$) are the two indices of the transform coefficients inside a subpartition. N is the number of pixels inside a subpartition. Fig. 9 shows the process for calculating the $MAST$ value for the HOR-ISP case. $MAST$ is calculated on the residual obtained when a non-ISP prediction is applied to the whole block. After obtaining the residual, it is split according to the ISP mode. After that, the mean absolute sum of the transform coefficient excluding the upper left 1/4 of the subpartition is $MAST$.

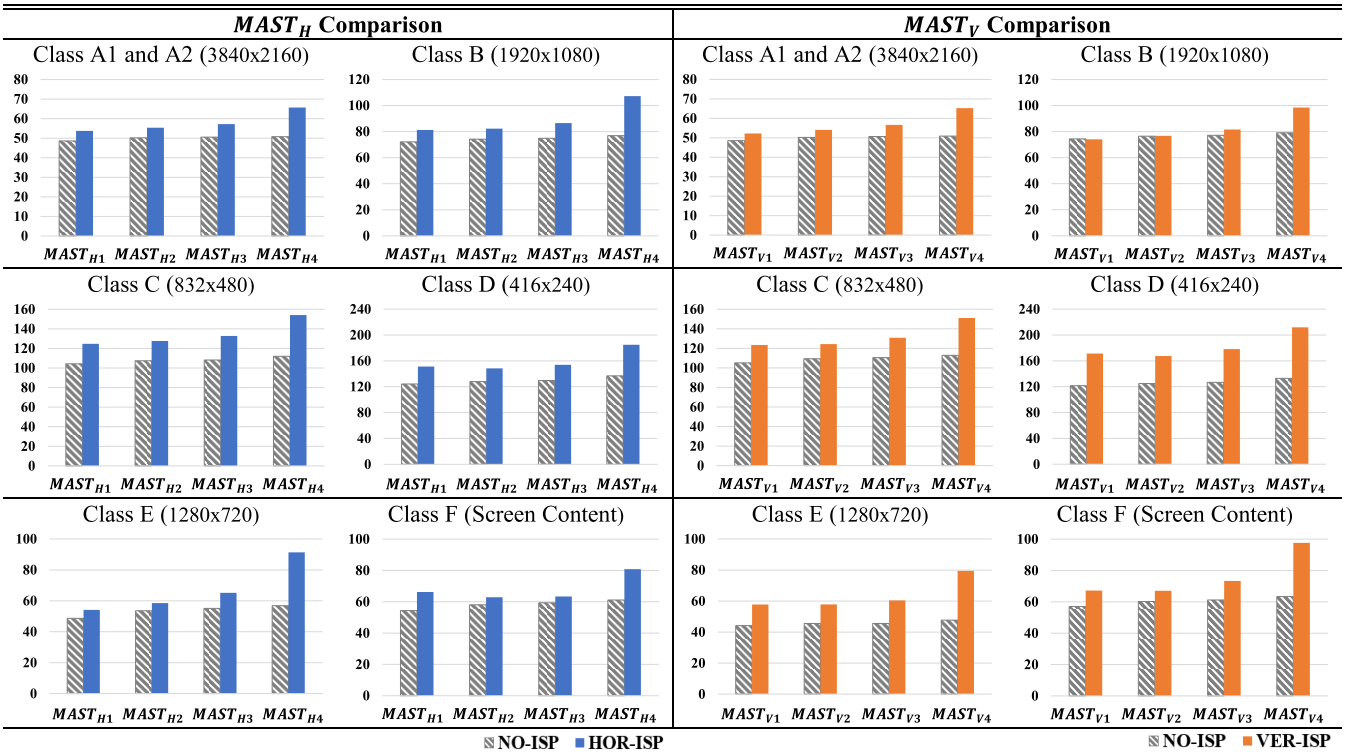

FIGURE 10. MAST comparison in all test video sequences [40].

Therefore, there can be four *MASTs* in the case of an ISP quad split, and two *MASTs* in the case of a half split. *MAST* for VER-ISP is calculated in the same way as a vertical split. The method of calculating *MAST* by dividing a block in the horizontal direction is called $MAST_H$, and the method of calculating *MAST* by dividing a block in the vertical direction is called $MAST_V$. Fig. 10 compares $MAST_H$ values for each CU block coded with NO-ISP and HOR-ISP. In the same way, $MAST_V$ values for each CU block coded with NO-ISP and VER-ISP. In both $MAST_H$ and $MAST_V$, it is observed that there is no significant difference by subpartition in NO-ISP. On the other hand, in the case of HOR-ISP in $MAST_H$, it is observed that the $MAST_H$ values gradually increase from $MAST_{H1}$ to $MAST_{H4}$. Also, in the case of VER-ISP in $MAST_V$, it is observed that the $MAST_V$ values gradually increase from $MAST_{V1}$ to $MAST_{V4}$. Therefore, it confirms that there is a clear difference in the *MAST* values of NO-ISP and each ISP mode. In general, in intra prediction, the residual increases as the distance from the reference samples increase, so the residual of each subpartition gradually increases from subpartition 1 to subpartition 4. At this time, when the prediction is performed by dividing a block, the difference between $MAST_{H1}$ to $MAST_{H4}$($MAST_{V1}$ to $MAST_{V4}$) can be reduced because the residual in a block can be minimized by narrowing the distance to the reference sample for each subpartition. Therefore, the ISP mode is effective for CU blocks whose values tend to increase from $MAST_{H1}$ to $MAST_{H4}$ ($MAST_{V1}$ to $MAST_{V4}$). To make a fast estimation, the *MAST* of first subpartition ($MAST_F$) and the last subpartition ($MAST_L$) are considered as key features. Table 6 shows the *MAST* ratio ($MAST_R$) for each class. Here, the $MAST_R$ is obtained by dividing $MAST_L$ by $MAST_F$. Regardless of whether it is HOR-ISP or VER-ISP, the $MAST_R$ can clearly determine the NO-ISP or ISP mode. Since $MAST_F$, $MAST_L$, and the $MAST_R$ have very strong correlations with the ISP mode decision, these are also considered to be the most effective features for deciding whether the ISP mode will be I$MAST_H MAST_V$ used or not.

TABLE 6. MAST ratio in all test video sequences [40].

MAST	Class	A1	A2	B	C	D	E	F	Overall
$MAST_H$	NO-ISP	1.03	1.06	1.08	1.07	1.09	1.18	1.12	1.09
	HOR-ISP	1.21	1.19	1.23	1.06	1.06	1.57	1.22	1.22
$MAST_V$	NO-ISP	1.04	1.07	1.08	1.06	1.07	1.12	1.11	1.08
	VER-ISP	1.20	1.27	1.29	1.22	1.25	1.53	1.45	1.32

D. ISP-ESD TRAINING

Here, the proposed ISP-ESD training is introduced in detail. As discussed previously, it has been confirmed that the block size and aspect ratio, intra prediction mode, $MAST_F$ and $MAST_L$, and $MAST_R$ strongly influence the decision of whether or not to use the ISP mode. In order to train the proposed ISP-ESD, four test sequences, which are common

TABLE 7. Video sequences used for training.

Anchor	VTM11.0
CFG	All Intra (AI)
Condition	All frames, QP {22, 27, 32, 37}
Sequence	Kimono, ParkScene, PeopleOnStreet, Traffic

TABLE 8. Specification of ISP-ESD models.

Method	ISP-ESD design (ISP-ESD-H, ISP-ESD-V)
Input features	Size (W, H)
	Aspect Ratio ($\log_2$ W- $\log_2$H)
	Intra Prediction Mode
	MAST
Output	1: Skip ISP mode test or 0: Test ISP mode
boosting_type	*gbdt*
max_depth	10
num_leaves	15
metric	*binary_logloss*
learning_rate	0.01

TABLE 9. Accuracy (%), Precision (%), and recall (%) of ISP-ESD model using different combinations of $MAST_F$, $MAST_L$, and $MAST_R$.

ISP-ESD	Hit ratio	*MAST*		
		$MAST_F, MAST_L$	$MAST_R$	$MAST_F, MAST_L, MAST_R$
ISP-ESD-H	Accuracy	82	84	**87**
	Precision	85	87	**86**
	Recall	81	82	**87**
ISP-ESD-V	Accuracy	80	81	**84**
	Precision	81	83	**88**
	Recall	79	81	**81**

$*MAST_R = MAST_L / MAST_F$ ($if\ MAST_F = 0, MAST_R = \frac{MAST_L}{0.0001}$)

test video sequences of HEVC [39] but not common test video sequences of VVC [40], are encoded using VVC reference software VTM11.0 [41]. The training datasets are collected by encoding all frames of the four test sequences with four QP values (22, 27, 32, and 37) under all intra (AI) configuration as shown in Table 7. The dataset is further divided into a training set (70%) and a validation set (30%). The ISP-ESD scheme employs two classifiers, one for ISP-ESD for HOR-ISP (ISP-ESD-H) and the other for ISP-ESD for VER-ISP (ISP-ESD-V). If an output of ISP-ESD-H (or ISP-ESD-V) is 0, a CU will undergo testing corresponding ISP combination *(mode, split, lfnst)*. In the same sense, the output of ISP-ESD-H (or ISP-ESD-V) being 1 indicates skipping the corresponding test of a given ISP combination *(mode, split, lfnst)*. The selected input features and the associated outputs are described in Table 8 with the specification of ISP-ESD design. *Boosting_type* is gradient boosting decision tree (*gbdt*). *Num_leaves* denotes the maximum number of leaves in one tree, and *max_depth* denotes the maximum depth for each tree. As the number of branches in the tree increases and the depth increases, the accuracy increases, but the execution time also increases. Therefore, to ensure a low complexity encoder, the optimal parameter settings for the

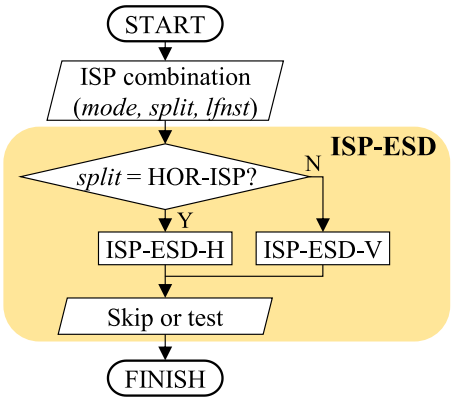

FIGURE 11. Flow chart of ISP-ESD scheme with LGBM classifiers.

two ISP-ESD models were set to 10 for *max_depth* and 15 for *num_leaves* in consideration of complexity and accuracy. Log loss (*binary_logloss*) is applied as the loss function to our binary classification problem. The *Learning_rate*, which corresponds to the rate at which errors are corrected from each iteration to the next, is set to 0.01. After the hyperparameters were determined, each classifier of ISP-ESD scheme is evaluated using 10-fold cross-validation. Table 9 shows the hit ratio (Accuracy (%), Precision (%), and Recall (%)) of three MAST feature combinations: both $MAST_F$ and $MAST_L$ are used, only the $MAST_R$ is used, and all three are used. It is observed that the hit ratio is the highest when all three $MAST_F$, $MAST_L$, $MAST_R$ are used. Therefore, for ISP-ESD, $MAST_F$, $MAST_L$, and $MAST_R$ are considered as *MAST* feature. Fig. 11 shows the flowchart of ISP-ESD scheme with two classifiers ISP-ESD-H, and ISP-ESD-V. A given ISP combination (*mode, split, lfnst*) undergoes through one of the classifiers depending on "*split*". As a result, whether to skip the ISP mode test for the corresponding ISP combination (*mode, split, lfnst*) is determined according to the output of each ISP-ESD-H and ISP-ESD-V.

V. EXPERIMENTAL RESULT

All the experiments in this paper are carried out by implementing the methods under test on top of VVC reference software VTM 11.0 with four QP settings (22, 27, 32, and 37). Total 26 video sequences from the common test conditions are encoded under all intra (AI) coding configurations [40], and the BDBR [42] and the average time saving (ATS) in (3) are evaluated against the anchor which is the method in VTM 11.0 (except Table 11). A negative BDBR value represents a better compression performance of the tested method than the anchor. A positive ATS value indicates that the encoding time is saved by the tested method compared to the anchor. $T_{PROPOSED}$ is the total encoding time taken by the proposed method and T_{Anchor} is the time taken by the anchor.

$$Average\ Time\ Saving\ (ATS) = \frac{T_{Anchor} - T_{PROPOSED}}{T_{Anchor}} \times 100\ (\%) \quad (3)$$

TABLE 10. Performance comparison against anchor (under all intra configuration).

Class	Videos	ISP-OFF				HARD-HOR-ISP				HARD-VER-ISP				ISP-ESD (proposed)			
		BDBR (%)			ATS	BDBR (%)			ATS	BDBR (%)			ATS	BDBR (%)			ATS
		Y	Cb	Cr	(%)	Y	Cb	Cr	(%)	Y	Cb	Cr	(%)	Y	Cb	Cr	(%)
A1	Tango2	0.07	-0.36	-0.33	12	0.05	-0.62	-0.14	5	0.04	-0.17	0.05	5	0.09	-0.42	-0.12	8
	FoodMarket4	0.05	0.00	-0.25	11	0.03	0.08	0.02	3	0.04	0.06	-0.22	3	0.05	-0.02	-0.03	8
	Campfire	0.10	-0.14	-0.61	15	0.03	-0.13	-0.54	6	0.05	0.00	0.10	5	0.04	-0.08	-0.27	12
A2	CatRobot	0.33	0.09	0.08	11	0.12	-0.01	0.00	5	0.21	0.13	0.08	5	0.15	0.06	0.01	13
	DaylightRoad2	0.45	0.35	0.29	14	0.18	0.24	0.01	6	0.30	0.23	0.10	6	0.17	0.27	0.12	5
	ParkRunning3	0.07	0.05	0.02	11	0.03	0.01	0.01	5	0.05	0.04	0.00	5	0.05	0.05	0.00	5
B	MarketPlace	0.13	0.00	0.06	13	0.06	0.09	0.02	5	0.06	0.05	0.06	6	0.02	-0.05	0.00	5
	RitualDance	0.32	0.26	0.01	13	0.14	0.30	0.06	5	0.16	0.17	0.00	5	0.04	0.21	-0.01	7
	Cactus	0.52	0.32	0.48	18	0.22	0.05	0.32	5	0.27	0.20	0.30	6	0.05	0.04	0.14	11
	BasketballDrive	0.71	0.50	0.56	13	0.29	0.09	-0.02	5	0.45	0.47	0.23	7	0.16	0.16	0.08	8
	BQTerrace	0.56	0.77	0.78	15	0.10	0.07	0.09	6	0.42	0.61	0.74	7	0.06	0.06	0.06	4
C	RaceHorses	0.39	0.31	0.08	16	0.08	-0.01	0.02	7	0.26	0.36	-0.05	8	-0.03	0.07	0.05	7
	BQMall	1.00	0.75	1.01	15	0.38	0.35	0.56	7	0.58	0.47	0.57	7	0.16	0.07	0.19	10
	PartyScene	0.53	0.46	0.58	17	0.26	0.13	0.17	7	0.24	0.08	0.31	7	0.09	0.04	0.11	9
	BasketballDrill	1.00	0.35	0.30	19	0.37	0.07	0.13	8	0.47	0.16	0.31	8	0.02	0.01	-0.26	6
E	FourPeople	0.74	0.48	0.49	12	0.42	0.30	0.31	7	0.30	0.07	0.11	6	0.10	0.07	0.05	6
	Johnny	0.74	0.33	0.01	12	0.52	-0.05	0.11	7	0.23	-0.10	-0.03	6	0.09	-0.19	-0.07	3
	KristenAndSara	0.91	0.58	0.51	12	0.64	0.51	0.61	5	0.23	0.18	-0.14	4	0.13	0.13	0.13	3
	Overall	**0.48**	**0.28**	**0.23**	**13.8**	**0.22**	**0.08**	**0.10**	**5.5**	**0.24**	**0.17**	**0.14**	**5.8**	**0.08**	**0.03**	**0.01**	**7.2**
D	RaceHorses	0.44	0.50	0.11	15	0.19	0.23	0.49	7	0.20	0.25	-0.04	6	0.10	0.07	0.06	7
	BQSquare	0.68	0.28	0.76	16	0.25	-0.04	0.31	7	0.40	0.11	0.74	8	0.03	-0.07	-0.19	8
	BlowingBubbles	0.66	0.37	0.70	17	0.40	0.39	0.45	7	0.27	0.14	0.23	6	0.07	0.07	0.25	5
	BasketballPass	0.69	0.54	0.70	15	0.34	0.13	0.28	2	0.40	0.17	0.32	2	0.04	-0.13	-0.03	3
F	ArenaOfValor	0.70	0.40	0.34	6	0.26	0.19	0.15	3	0.40	0.24	0.23	3	0.02	0.06	0.04	1
	BasketballDrillText	0.97	0.39	0.35	8	0.46	0.03	0.03	3	0.45	0.44	0.11	4	0.06	0.20	-0.03	2
	SlideEditing	0.39	0.06	0.02	7	0.09	0.04	0.06	1	0.28	0.06	0.21	2	-0.01	0.02	-0.06	5
	SlideShow	0.82	0.19	-0.11	6	0.44	0.09	0.01	2	0.35	-0.17	-0.03	3	-0.02	0.02	-0.01	2

Before evaluating ATS, it is meaningful to investigate the maximum possible ATS by the fast ISP algorithm under test. In this context, an experiment comparing ISP-OFF (that is, VTM11.0 with ISP disabled), HARD-HOR-ISP, and HARD-VER- ISP is carried out as shown in Table 10. HARD-HOR-ISP is a method of disabling VER-ISP always, thereby allowing only HOR-ISP (i.e., it compares the RD costs of HOR-ISP and NO-ISP only). In the same way, HARD-VER-ISP disables HOR-ISP (i.e., it compares the RD costs of VER-ISP and NO-ISP only). The results in Table 10 show that when ISP itself is disabled, an average of 13.8% of the encoding time can be saved; this is the highest attainable ATS by a fast ISP scheme in the general. Note that ISP-OFF suffers significantly with a BDBR loss of around 0.48%. When only one mode is allowed (either HARD-HOR-ISP or HARD-VER-ISP), an average ATS of 5.5% to 5.8% is achieved at a BDBR loss of around 0.22% to 0.24%. When the proposed ISP-ESD is used, one can achieve ATS of 7.2% with a small BDBR loss of 0.08%.

For the test sequence Tango2 of class A1, the BDBR increase of the proposed method exceeds that of ISP-OFF. This is because the decisions of intra mode are made according to the RD cost, where the rate is estimated, so it may not be the same as actual RD cost after entropy coding [43]. In case of the test sequence CatRobot of class A2, ATS of the proposed method exceeds that of ISP-OFF. Turning off such coding tools may result in more residuals after prediction, which can consequently increase the processing time at quantization and entropy coding stages later. Therefore, sometimes, total encoding time can be increased when such coding tools are turned off [44]. CatRobot in class A2, the proposed ISP-ESD scheme achieves an ATS of up to 13%. In class B (1920 × 1080), the proposed ISP-ESD brings an ATS of 7% with a 0.07% BDBR loss. In addition, for the RaceHorses(832 × 480) sequence, the proposed ISP-ESD brings an ATS of 7% with no loss. The experiment shows that the proposed ISP-ESD successfully simplifies the ISP encoder search process without practically problematic BDBR loss. In the class F, which is screen content videos,

TABLE 11. Performance comparison of each ISP-ESD model; ISP-ESD-H, ISP-ESD-V. (under all intra configuration).

Class	HARD-HOR-ISP vs ISP-ESD-H				HARD-VER-ISP vs ISP-ESD-V			
	BDBR (%)			ATS	BDBR (%)			ATS
	Y	Cb	Cr	(%)	Y	Cb	Cr	(%)
A1	0.05	0.12	-0.07	6	0.06	-0.15	-0.18	5
A2	0.13	0.06	0.10	6	0.08	0.08	0.14	4
B	0.03	0.06	0.04	3	0.04	-0.01	0.02	3
C	0.06	0.01	0.03	2	0.06	0.06	0.08	2
E	0.04	0.14	-0.02	2	0.13	0.08	0.17	3
Overall	**0.06**	**0.08**	**0.02**	**4**	**0.07**	**0.01**	**0.05**	**3**
D	0.02	0.04	-0.03	2	0.06	0.04	0.07	2
F	0.74	0.61	0.50	1	0.10	0.05	-0.01	1

TABLE 12. Performance comparison of the proposed method with existing works [19], [20] against anchor VTM11.0.

Class	Videos	Existing work [19]				Existing work [20]				ISP-ESD (proposed)			
		BDBR (%)			ATS	BDBR (%)			ATS	BDBR (%)			ATS
		Y	Cb	Cr	(%)	Y	Cb	Cr	(%)	Y	Cb	Cr	(%)
A1	Tango2	0.05	-0.01	0.11	8	0.06	-0.32	-0.26	9	0.09	-0.42	-0.12	8
	FoodMarket4	0.03	0.10	0.02	7	0.02	0.04	-0.06	6	0.05	-0.02	-0.03	8
	Campfire	0.10	0.02	0.01	7	0.00	-0.14	-0.62	7	0.04	-0.08	-0.27	12
A2	CatRobot	0.17	0.10	0.05	9	0.12	0.04	-0.04	8	0.15	0.06	0.01	13
	DaylightRoad2	0.23	0.20	0.05	10	0.14	0.30	0.05	6	0.17	0.27	0.12	5
	ParkRunning3	0.03	0.01	0.02	10	0.04	0.04	-0.03	5	0.05	0.05	0.00	5
B	MarketPlace	0.04	0.08	-0.04	5	0.05	0.05	-0.15	4	0.02	-0.05	0.00	5
	RitualDance	0.12	0.21	0.10	4	0.10	0.25	-0.08	4	0.04	0.21	-0.01	7
	Cactus	0.23	0.22	0.36	5	0.07	0.02	0.11	4	0.05	0.04	0.14	11
	BasketballDrive	0.42	0.42	0.27	5	0.22	0.07	-0.12	5	0.16	0.16	0.08	8
	BQTerrace	0.37	0.58	0.75	7	0.06	0.09	0.03	3	0.06	0.06	0.06	4
C	RaceHorses	0.18	0.28	-0.10	3	0.01	-0.05	-0.14	3	-0.03	0.07	0.05	7
	BQMall	0.41	0.27	0.40	7	0.12	0.04	0.05	2	0.16	0.07	0.19	10
	PartyScene	0.26	0.12	0.29	7	0.02	-0.04	0.00	2	0.09	0.04	0.11	9
	BasketballDrill	0.67	0.33	0.52	7	0.30	-0.11	0.15	7	0.02	0.01	-0.26	6
E	FourPeople	0.38	0.24	0.22	4	0.16	-0.01	0.05	4	0.10	0.07	0.05	6
	Johnny	0.41	0.01	-0.04	5	0.15	-0.10	-0.03	4	0.09	-0.19	-0.07	3
	KristenAndSara	0.54	0.58	0.44	5	0.12	0.06	0.09	3	0.13	0.13	0.13	3
	Overall	**0.27**	**0.21**	**0.22**	**7.0**	**0.10**	**0.01**	**-0.06**	**4.8**	**0.08**	**0.03**	**0.01**	**7.2**
D	RaceHorses	0.26	0.40	0.31	5	-0.03	0.12	0.12	2	0.10	0.07	0.06	7
	BQSquare	0.36	0.33	0.55	4	0.07	-0.31	0.03	4	0.03	-0.07	-0.19	8
	BlowingBubbles	0.33	0.19	0.32	6	0.06	-0.15	-0.13	5	0.07	0.07	0.25	5
	BasketballPass	0.35	-0.15	0.33	7	0.15	-0.13	0.02	2	0.04	-0.13	-0.03	3
F	ArenaOfValor	0.38	0.17	0.20	2	0.18	0.11	0.02	2	0.02	0.06	0.04	1
	BasketballDrillText	0.67	0.28	0.25	4	0.26	0.08	-0.18	1	0.06	0.20	-0.03	2
	SlideEditing	0.30	0.08	0.06	3	-0.02	-0.15	0.07	-1	-0.01	0.02	-0.06	5
	SlideShow	0.71	0.46	-0.08	-2	0.06	-0.18	-0.33	1	-0.02	0.02	-0.01	2

many blocks select Intra-Block Copy (IBC) and Transform skip Mode which are not integrated with ISP [45]. Thus, the proposed method has relatively small ATS and BDBR loss.

Furthermore, the coding performance of two models, ISP-ESD-H and ISP-ESD-V is given respectively in Table 11. To confirm the effect of ISP-ESD-H, the anchor was set to HARD-HOR-ISP, and the test was set to HARD-HOR-ISP + ISP-ESD-H. In the same sense, to confirm the effect of ISP-ESD-V, the anchor was set to HARD-VER-ISP, and the test was set to HARD-VER-ISP + ISP-ESD-V. As shown in Table 11, both ISP-ESD-H and ISP-ESD-V provide complexity reduction.

To understand the performance of our proposed method more fairly, we also compare it with the existing works [19], [20]. Table 12 shows the overall performance comparison between our work in this paper and the existing methods in [19], [20], which were implemented on VTM11.0. It should be noted that the method described in [19] was implemented on VTM9.0 and the method in [20] was implemented on VTM8.0, originally. In [19], a fast ISP mode skip decision was proposed by considering only the prediction mode and the ISP mode. To be more precise, encoding time is saved by disabling the ISP mode for cases where the reconstructed reference samples are used in the same way as in the case of NO-ISP, rather than using the newly reconstructed reference samples. In the method of [19], regardless of the location of the reference samples, there are many cases where the ISP mode cannot be used, even if the ISP is efficient in the current block. For example, as mentioned in Fig. 4, in the case of class E, 41% of the CU blocks that selected ISP were not using the newly reconstructed samples in the previous subpartition but were using reconstructed samples of neighboring CUs. However, due to the ISP fast decision scheme described in [19], those CU blocks cannot use the ISP mode. Therefore, a BDBR loss of 0.67% (BasketballDrill (832 × 480) in class C), which is the worst case, occurs. In [20], they aim to determine whether a CU needs to use ISP mode in advance by calculating CU texture complexity so as to reduce the computation complexity of ISP; if the value of the CU texture complexity is above a certain threshold (the decision threshold is set to 20), the ISP mode is not tested. It is noted that the method in [20] generally performs better on high-resolution video sequences, such as UHD/FHD (class A1, class A2, and class B). This is because the fast ISP mode decision scheme in [20] has better performance on simple texture blocks. Simple texture blocks often make up a larger portion of UHD/FHD videos. However, sufficient encoding time reduction cannot be achieved for sequences with complex textures. For example, [20] can achieve an ATS of 9% with a mere 0.06% BDBR loss at Tango2 (3840 × 2160). But in BasketballPass (416 × 240), [20] showed a poor encoding time reduction of 2.0% with 0.15% BDBR loss. In [22], which is the most recent related work, texture features were also defined for early termination of the ISP mode search. However, also the IBC mode search is terminated early by enabling the IBC mode to be used in all CUs

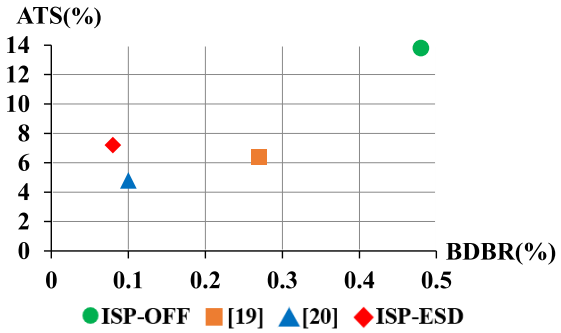

FIGURE 12. BDBR vs. ATS comparison of existing works [19], [20], ISP-OFF, and ISP-ESD under the AI configuration.

(according to [40], IBC is turned off by default in natural sequences). Therefore, it is difficult to compare [22] to our method because the experimental results for early termination of only the ISP mode search are not available. For our method, we proposed an early skip decision method that limits the ISP mode search by defining *MAST*. Here, the prediction efficiency within a block and also the transform effect were checked in advance to determine whether to use or not the given ISP mode. Therefore, the overall time reduction can be confirmed and the result of maximizing the effect of the ISP can be obtained. Fig. 12 shows the BDBR and ATS comparison of existing works [19], [20], ISP-OFF, and ISP-ESD under the AI configuration. The x-axis is BDBR (%) and the y-axis is ATS (%). It is shown that ISP-ESD achieves significant encoding time saving with the least BDBR loss.

VI. CONCLUSION

In this paper, we addressed a machine learning-based ISP-ESD scheme in a VVC encoder which makes an early skip decision of ISP mode considering the split direction. Our solution takes ISP mode test decision as a binary classification problem, where LightGBM classifiers (ISP-ESD-H, ISP-ESD-V) are trained offline for each ISP mode, and these classifiers are responsible for deciding whether to skip the test of ISP mode or not, thus skipping the estimation of the RDO process in specific cases. The proposed machine learning-based ISP-ESD determines whether or not to skip the corresponding test of ISP combination considering the intra mode, block size, aspect ratio, and *MAST* of subpartition. Therefore, it is noted that our solution is the first machine learning-based ISP early skip decision algorithm taking into account the two coding efficiency benefits of ISP in the general case: independent transform of each subpartition and efficient intra prediction at the same time. Through our experiments, the proposed ISP-ESD scheme is shown to provide a reduction in total encoding time with a very marginal coding efficiency loss compared to the existing methods.

REFERENCES

[1] *Versatile Video Coding*, Standard ISO/IEC 23090-3, ISO/IEC JTC 1, Jul. 2020.
[2] B. Bross, J. Chen, S. Liu, and Y. Wang, *Versatile Video Coding Editorial Refinements on Draft 10*, document Joint Video Experts Team (JVET), 20th Meeting, teleconference, TR JVET-T2001-v2, Oct. 2020.
[3] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021.
[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[5] F. Bossen, X. Li, K. Suehring, K. Sharman, V. Seregin, and A. Tourapis, *AHG Report: Test Model Software Development (AHG3)*, document Joint Video Experts Team (JVET), 21st Meeting, Teleconference, JVET-U0003-v1, Jan. 2021.
[6] W. Chien, *JVET AHG Report: Tool Reporting Procedure (AHG13)*, document Joint Video Experts Team (JVET), 20th Meeting, Teleconference, JVET-T0013, Oct. 2020.
[7] A. Mercat, A. Makinen, J. Sainio, A. Lemmetti, M. Viitanen, and J. Vanne, "Comparative rate-distortion-complexity analysis of VVC and HEVC video codecs," *IEEE Access*, vol. 9, pp. 67813–67828, 2021.
[8] J. Pfaff, A. Filippov, S. Liu, X. Zhao, J. Chen, S. De-Luxán-Hernández, T. Wiegand, V. Rufitskiy, A. K. Ramasubramonian, and G. Van der Auwera, "Intra prediction and mode coding in VVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3834–3847, Oct. 2021.
[9] Y. Wang, X. Fan, D. Zhao, and W. Gao, "Mode dependent intra smoothing filter for HEVC," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 539–543.
[10] S. Lee and N. Cho, "Intra prediction method based on the linear relationship between the channels for YUV 4:2:0 intra coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 1037–1040.
[11] A. Said, X. Zhao, M. Karczewicz, J. Chen, and F. Zou, "Position dependent prediction combination for intra-frame video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 534–538.
[12] Y.-J. Chang, H.-J. Jhu, H.-Y. Jiang, L. Zhao, X. Zhao, X. Li, S. Liu, B. Bross, P. Keydel, H. Schwarz, D. Marpe, and T. Wiegand, "Multiple reference line coding for most probable modes in intra prediction," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2019, p. 559.
[13] B. Bross, *CE3: Multiple Reference Line Intra Prediction (Test 1.1.1, 1.1.2, 1.1.3 and 1.1.4)*, document JVET 12th Meeting, JVET-L0283, Macao, China, Oct. 2018.
[14] S. De-Luxán-Hernández, "An intra subpartition coding mode for VVC," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 1203–1207.
[15] S. De-Luxán-Hernández, *Non-CE3: Proposed ISP Cleanup*, document Joint Video Experts Team (JVET), 15th Meeting, Gothenburg, Sweden, JVET-O0502-v3, Jul. 2019.
[16] S. De-Luxán-Hernández, V. George, G. Venugopal, J. Pfaff, H. Schwarz, D. Marpe, and T. Wiegand, "Design of the intra subpartition mode in VVC and its optimized encoder search in VTM," in *Proc. Appl. Digit. Image Process.*, Aug. 2020, pp. 165–180.
[17] M. Schafer, B. Stallenberger, J. Pfaff, P. Helle, H. Schwarz, D. Marpe, and T. Wiegand, "A data-trained, affine-linear intra-picture prediction in the frequency domain," in *Proc. Picture Coding Symp. (PCS)*, Ningbo, China, Nov. 2019, pp. 1–5.
[18] L. Zhao, X. Zhao, S. Liu, X. Li, J. Lainema, G. Rath, F. Urban, and F. Racape, "Wide angular intra prediction for versatile video coding," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2019, pp. 53–62.
[19] J. Park, B. Kim, and B. Jeon, "Fast VVC intra prediction mode decision based on block shapes," in *Proc. Appl. Digit. Image Process.*, vol. 11510, Aug. 2020, pp. 581–593.
[20] Z. Liu, M. Dong, X. H. Guan, M. Zhang, and R. Wang, "Fast ISP coding mode optimization algorithm based on CU texture complexity for VVC," *EURASIP J. Image Video Process.*, vol. 2021, no. 1, pp. 1–14, Dec. 2021.
[21] M. Saldanha, G. Sanchez, C. Marcon, and L. Agostini, "Learning-based complexity reduction scheme for VVC intra-frame prediction," in *Proc. Int. Conf. Vis. Commun. Image Process. (VCIP)*, Munich, Germany, Dec. 2021, pp. 1–5.
[22] X. Dong, L. Shen, M. Yu, and H. Yang, "Fast intra mode decision algorithm for versatile video coding," *IEEE Trans. Multimedia*, vol. 24, pp. 400–414, 2022.
[23] Q. Zhang, Y. Wang, L. Huang, and B. Jiang, "Fast CU partition and intra mode decision method for H.266/VVC," *IEEE Access*, vol. 8, pp. 117539–117550, 2020.

[24] M. Xu and B. Jeon, "Improved hard-decision quantization with decision tree for HEVC video compression," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2020, p. 401.
[25] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1668–1682, Jun. 2020.
[26] S.-H. Park and J.-W. Kang, "Context-based ternary tree decision method in versatile video coding for fast intra coding," *IEEE Access*, vol. 7, pp. 172597–172605, 2019.
[27] M. Saldanha, G. Sanchez, C. Marcon, and L. Agostini, "Configurable fast block partitioning for VVC intra coding using light gradient boosting machine," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3947–3960, Jun. 2022.
[28] M.-J. Chen, C.-A. Lee, Y.-H. Tsai, C.-M. Yang, C.-H. Yeh, L.-J. Kau, and C.-Y. Chang, "Efficient partition decision based on visual perception and machine learning for H.266/Versatile video coding," *IEEE Access*, vol. 10, pp. 42141–42150, 2022.
[29] X. Zhao, S.-H. Kim, Y. Zhao, H. E. Egilmez, M. Koo, S. Liu, J. Lainema, and M. Karczewicz, "Transform coding in the VVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3878–3890, Oct. 2021.
[30] Z. Zhang, X. Zhao, X. Li, L. Li, Y. Luo, S. Liu, and Z. Li, "Fast DST-VII/DCT-VIII with dual implementation support for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 355–371, Jan. 2021.
[31] M. Koo, M. Salehifar, J. Lim, and S.-H. Kim, "Low frequency non-separable transform (LFNST)," in *Proc. Picture Coding Symp. (PCS)*, Ningbo, China, Nov. 2019, pp. 1–5.
[32] H. Schwarz, M. Coban, M. Karczewicz, T.-D. Chuang, F. Bossen, A. Alshin, J. Lainema, C. R. Helmrich, and T. Wiegand, "Quantization and entropy coding in the versatile video coding (VVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3891–3906, Oct. 2021.
[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[34] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.
[35] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist. (COMPSTAT)*, Paris, France, Aug. 2010, pp. 177–186.
[36] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification: Fast feature extraction and SVM training," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 1689–1696.
[37] D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991.
[38] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 3146–3154.
[39] F. Bossen, *Common HM Test Conditions and Software Reference Configuration*, document Joint Collaborative Team on Video Coding (JCT-VC), Geneva, Switzerland, JCTVC-L1100, Jul. 2012.
[40] F. Bossen, J. Boyce, K. Suehring, X. Li, and V. Seregin, *JVET Common Test Conditions and Software Reference Configurations for SDR Video*, document Joint Video Experts Team (JVET), 14th Meeting, Geneva, Switzerland, JVET-N1010-v1, Mar. 2019.
[41] J. Chen, Y. Ye, and S. Kim, *Algorithm Description for Versatile Video Coding and Test Model 11 (VTM 11)*, document Joint Video Experts Team (JVET), 20th Meeting, Teleconference, JVET-T2002-v2, Oct. 2020.
[42] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RDCurves*, document ITU-T SG16, VCEG-AH21, Antalya, Turkey, Jan. 2008.
[43] H. Schwarz, T. Nguyen, D. Marpe, and T. Wiegand, "Hybrid video coding with trellis-coded quantization," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2019, pp. 182–191.
[44] M. Xu and B. Jeon, "User-priority based AV1 coding tool selection," *IEEE Trans. Broadcast.*, vol. 67, no. 3, pp. 736–745, Sep. 2021.
[45] X. Xu, S. Liu, T.-D. Chuang, Y.-W. Huang, S.-M. Lei, K. Rapaka, C. Pang, V. Seregin, Y.-K. Wang, and M. Karczewicz, "Intra block copy in HEVC screen content coding extensions," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 409–419, Dec. 2016.

JEEYOON PARK (Student Member, IEEE) was born in Seoul, South Korea, in 1993. She received the bachelor's degree from the Department of Computer Engineering, Kookmin University, in 2016. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Sungkyunkwan University, South Korea. Her research interest includes image/video coding.

BUMYOON KIM (Student Member, IEEE) was born in Seoul, South Korea, in 1994. He received the bachelor's degree in electronic and electrical engineering from Sungkyunkwan University, in 2020, where he is currently pursuing the Ph.D. degree. His research interests include video compression, image processing, and neural networks.

JEEHWAN LEE (Student Member, IEEE) received the B.S. degree in electronic and electrical engineering from Sungkyunkwan University, Suwon, South Korea, in 2021, where he is currently pursuing the M.S. degree. His research interests include HEVC and VVC video coding standards, in particular, intra prediction and residual coding of chroma channel.

BYEUNGWOO JEON (Senior Member, IEEE) received the B.S. degree *(Magna Cum Laude)*, the M.S. degree from the Department of Electronics Engineering, Seoul National University, Seoul, South Korea, in 1985 and 1987, respectively, and the Ph.D. degree from the School of Electrical Engineering, Purdue University, West Lafayette, USA, in 1992. From 1993 to 1997, he was at the Signal Processing Laboratory, Samsung Electronics, South Korea, where he worked for research and development of video compression algorithms, design of digital broadcasting satellite receivers, and other MPEG-related research for multimedia applications. Since September 1997, he has been with Sungkyunkwan University (SKKU), South Korea, where he is currently a Full Professor. His research interests include multimedia signal processing, video compression, statistical pattern recognition, and remote sensing. He was the Project Manager of Digital TV and Broadcasting at the Korean Ministry of Information and Communications from March 2004 to February 2006, where he has supervised all digital TV-related Research and Development in South Korea. From January 2015 to December 2016, he was the Dean of the College of Information and Communication Engineering, SKKU. In 2019, he was the President of Korean Institute of Broadcast and Media Engineers. He is a member of SPIE. He was a recipient of the 2005 IEEK Haedong Paper Award in Signal Processing Society in South Korea, and the 2012 Special Service Award and the 2019 Volunteer Award both from the IEEE Broadcast Technology Society. In 2016, he was conferred Korean President's Commendation for his key role in promoting international standardization for video coding technology in South Korea. He is an Associate Editor of IEEE TRANSACTIONS ON BROADCASTING and has been an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.