Fast block partitioning scheme for chrominance intra prediction of versatile video coding standard

Mário Saldanha[a,*] **Gustavo Sanchez**[b] **César Marcon**[c] **and Luciano Agostini**[a]

[a]Federal University of Pelotas, Pelotas, Brazil
[b]IF Farroupilha, Alegrete, Brazil
[c]Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

Abstract. We present a fast block partitioning scheme for chrominance intra prediction of the Versatile Video Coding (VVC) standard. VVC adopted several innovative tools to improve the coding performance compared with its predecessor standard, enabling the efficient handling of emerging video content such as ultra-high definition (UHD). These coding performance improvements come at the cost of a significant increase in coding complexity, mainly due to the new block partitioning structure called quadtree with nested multi-type tree (QTMT), which can be applied separately for luminance and chrominance components. Thus, this work proposes a scheme focusing on chrominance intra coding, composed of chrominance splitting early termination based on luminance QTMT and fast chrominance split decision based on the sub-block variances. Experimental results demonstrated that the proposed scheme provides 60.03% and 8.18% of chrominance and total encoding time savings, respectively, with a negligible impact of 0.66% in the YUV-BDBR. [DOI: 10.1117/1.JEI.30.5.053009]

Keywords: complexity reduction; chrominance intra coding; fast block partitioning; versatile video coding.

Paper 210310 received Jun. 3, 2021; accepted for publication Sep. 9, 2021; published online Sep. 23, 2021.

1 Introduction

Remarkable advances in video coding were obtained with the standardization of video codecs such as H.264 advanced video coding (AVC)[1] and H.265 high-efficiency video coding (HEVC),[2] which allowed for encoding efficiently high-resolution videos with great video quality using few bits. However, along with the continuous growth of digital video consumption over the internet, emerging technologies such as ultra-high definition (UHD) and immersive content, including high-dynamic range and 360 deg omnidirectional videos, are increasingly widespread, incurring enormous traffic of digital video data. This became even more noticeable during the COVID-19 pandemic, when the streaming services of different applications increased significantly. Thus, the current video coding standards do not provide satisfactory performance to meet the current market requirements of the next video applications. Hence, this created the demand for next-generation video coding technologies with capabilities beyond the current standards.

The Joint Video Experts Team (JVET) was established in a collaboration between ISO Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) to develop the Versatile Video Coding (VVC) standard,[3] defined as the Final Draft International Standard (FDIS) in July 2020.[3] VVC was designed to obtain a significantly higher compression efficiency than the HEVC standard and have a high versatility for efficient use with different types of video content.

Several novel techniques were developed and enhanced during the VVC standardization, significantly improving the coding efficiency compared with its predecessor standard. These

*Address all correspondence to Mário Saldanha, mrdfsaldanha@inf.ufpel.edu.br

1017-9909/2021/$28.00

improvements include larger block sizes, quadtree with nested multi-type tree (QTMT),[4] a higher number of angular intra prediction modes,[5] dual-tree (DT) coding structure [also known as chroma separate tree (CST)],[5] cross-component linear model (CCLM),[6] multiple transform selection,[7] and low-frequency non-separable transform (LFNST),[8] among others. These tools increased the VVC compression performance at the cost of a significant computational burden increase.

The intra coding effort of the VVC test model (VTM)[9] increased by about 28 times over the HEVC test model (HM) under an all-intra (AI) encoder configuration.[10,11] Consequently, fast coding algorithms and complexity reduction solutions are required to allow for adopting VVC and developing new applications that meet the practical requirements of the market.

Some works focusing on decreasing the complexity of VVC intra coding have been proposed.[12–16] Since the QTMT partitioning structure is the most time-consuming encoding step in VTM,[10] most of these works focused on accelerating this process. Yang et al.[12] proposed a complexity reduction scheme composed of two strategies. The first one uses decision trees to avoid unnecessary block partition evaluations; the second one applies a gradient descendent search to reduce the assessment of some angular intra prediction modes. Fu et al.[13] presented a fast block partitioning algorithm using a classifier based on the Bayesian decision rule to reduce the complexity of the multi-type tree (MTT). In this case, the information derived from the current block and the horizontal binary splitting is used as an input feature for the model. Lei et al.[14] developed a scheme to identify unnecessary block partition evaluations in advance, in which a subset of angular intra prediction modes is evaluated to estimate the cost of partitioning the current block with the horizontal and vertical partition types. Zhao et al.[15] presented a solution based on convolutional neural network (CNN), which is responsible for deciding if the current block should be split based on the residual information of the neighboring blocks. Our previous work[16] proposed a fast block partitioning decision scheme to avoid unnecessary evaluations of MTT partition types. For deciding the direction of MTT partitions to be evaluated, the proposed solution used the best intra prediction mode selected for the current block and the variance of subpartitions in the current block.

Although these works presented outstanding results regarding encoding efficiency and complexity reduction, they all focused on accelerating the encoding of luminance block samples. However, VVC allows the chrominance blocks to have a separate coding tree from luminance blocks, incurring a similar and complex evaluation process in the QTMT structure for chrominance blocks. From our experiments, the average chrominance encoding complexity is about 13% of the total VTM encoder. Therefore, complexity reduction solutions specialized for the chrominance QTMT structure are required to contribute and enable the development of the VVC real-time encoder.

This paper proposes a fast block partitioning scheme for chrominance intra prediction, which explores the correlation between the chrominance and luminance coding tree structures and the statistical information of the chrominance samples in the current block. The main contributions of this work are as follows:

i. Analysis of chrominance block partitioning;
ii. Analysis of the correlation between chrominance and luminance QTMT partitioning structures;
iii. Chrominance splitting early termination based on luminance;
iv. Fast chrominance split decision based on statistical information of the block samples.

The field of video processing has relevant works in different sub-areas, including several works,[17–19] and it is highly active in the challenge of reducing the complexity of video encoders. Therefore, this work brings significant contributions to our community.

The remainder of this paper is organized as follows. Section 2 presents the VVC block partitioning structure and analyzes the chrominance block partitioning for intra coding. Section 3 details the statistical analysis and the fast block partitioning scheme for chrominance. Section 4 presents the experimental results obtained with the proposed solution. Finally, Sec. 5 renders the conclusions of this work.

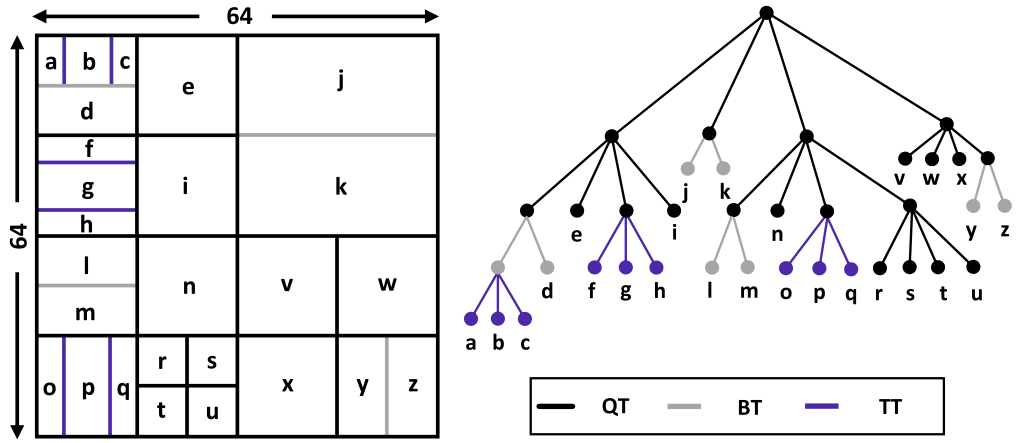


Fig. 1 Illustration of the QTMT partitioning process for a 64×64 chrominance CTU.

2 VVC Block Partitioning

VVC adopts the same concept of QT employed in the HEVC and inserts the MTT partitioning structure, allowing for square and rectangular coding unit (CU) sizes through QT and binary tree (BT)/ternary tree (TT),[5] respectively. These combined structures compose the QTMT partitioning structure, which allows for six types of partitions: no split, QT, binary tree horizontal (BTH), binary tree vertical (BTV), ternary tree horizontal (TTH), and ternary tree vertical (TTV). A CU can be defined as no split, and the coding process is performed with the current CU size; otherwise, this CU is split using the QT, BT, or TT structure. QT divides a CU into four equal-sized square sub-CUs. BT divides a CU into two symmetric sub-CUs. TT splits a CU into three sub-CUs, where the central sub-CU has 50% of the original CU size and the two sides of sub-CUs have 25%. In addition, BT and TT can be performed in horizontal and vertical directions.

The QTMT structure split each coding tree unit (CTU) recursively through the QT structure, and only the QT leaf nodes can be partitioned in an MTT structure (i.e., no further QT is allowed once a CU is partitioned using MTT). Additionally, MTT can split a QT leaf node recursively through BTH, BTV, TTH, and TTV partitions.

VVC also introduces the DT (or CST) coding tool,[5] enabling the luminance and chrominance components to have a separate QTMT coding tree structure for I-slices (slices that only allow intra prediction) to enhance the compression performance. Thus, encoding the chrominance blocks also performs several block partition evaluations, encompassing QT, BT, and TT structures.

In VVC, a CU can be larger than the CTU size (maximum size of 128×128 samples) or as small as 4×4 samples, considering luminance blocks. Regarding chrominance blocks, a CU can have a maximum size of 64×64 samples and a minimum size of 16 samples, which can be 8×2 or 4×4. However, for the intra prediction, the maximum luminance block size processed is 64×64, representing the maximum transform block size, whereas the maximum chrominance block is 32×32. Consequently, the QTMT structure starts evaluating the intra prediction with a depth level equal to one since the CU is implicitly split by the QT structure in the first level of the QTMT.

Figure 1 exemplifies the QTMT partitioning for a 64×64 chrominance CTU split into several CUs with different QT and MTT levels. Black, gray, and blue lines represent QT, BT, and TT splitting, respectively.

2.1 *Chrominance Block Partitioning Analysis*

This section analyzes the chrominance block partitioning in the intra prediction to understand the complexity reduction possibilities. The simulations were performed using VTM 10.0,[9] with all video sequences and quantization parameter (QP) specified in the JVET common test

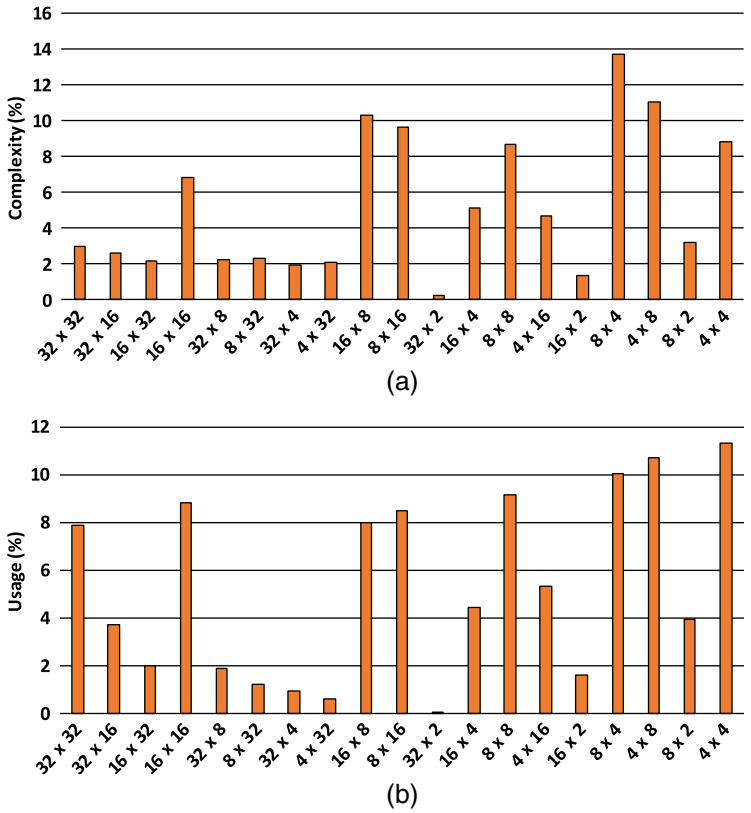


Fig. 2 (a) Encoding complexity and (b) usage distribution for chrominance block sizes.

conditions (CTC).[20] All of the presented analyses considered the average results among the video sequences and QPs.

Figure 2 shows the complexity and usage distributions for all available chrominance block sizes in the VVC intra prediction. Note that the highest encoding complexities are concentrated in smaller block areas, whereas the usage distribution occurs among different block areas. For instance, considering the larger blocks 32×32, 16×16, 16×8, and 8×16 and the smaller blocks 8×4, 4×8, and 4×4, the total encoding complexity is lower for larger blocks (29.8%) than for smaller blocks (33.7%). However, the total selection of these larger blocks (33.2%) is slightly higher than the smaller ones (32.1%), indicating that most of the chrominance coding complexity is related to evaluating smaller block sizes, but in several cases, the encoder decides to use larger block sizes.

The following analysis presents the impact of limiting the QTMT depth level and removing the horizontal or vertical directions of the MTT structure. We performed both modifications only for the chrominance QTMT structure to analyze the chrominance behavior. This evaluation considered the encoding efficiency measured through the Bjontegaard Delta bitrate (BDBR)[21] and the encoding time saving (ETS), regarding both chrominance ETS (C-ETS) and Total ETS (T-ETS). C-ETS measures only the chrominance coding time, excluding the time consumed in luminance coding, loop filters, and control operations.

BDBR expresses the percentage variation of the bitrate between a test configuration and a reference configuration, considering the same objective video quality measured through the peak signal noise ratio (PSNR).[21] Hence, a positive BDBR result indicates that the test configuration provides a worse compression efficiency than the reference configuration, and a negative BDBR result refers to a better compression efficiency of the test configuration compared with its reference. BDBR can be calculated individually for each channel (Y, U, and V) or combined for the three components. For all cases, the overall bitrate is used to compute BDBR (i.e., the bits considering the three components), whereas different PSNR values are used to compute the luminance (Y), chrominance (U), and chrominance (V) BDBR results, according to the corresponding component. YUV-BDBR also uses the overall bitrate, but it computes the PSNR

Table 1 Encoding efficiency and time saving of limiting the QTMT depth for chrominance blocks.

	QTMT depth 2					QTMT depth 3				
Class	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS (%)	T-ETS (%)	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS (%)	T-ETS (%)
A1	2.88	12.22	13.19	85.90	10.41	0.90	3.88	3.84	64.02	9.03
A2	4.25	20.09	17.28	91.78	20.12	1.54	7.45	6.63	75.88	19.03
B	2.56	15.41	17.97	89.23	8.88	0.90	5.88	6.67	70.33	7.62
C	5.42	20.06	24.76	93.06	13.00	1.99	8.55	10.33	79.21	12.40
D	5.42	20.36	24.74	94.21	13.34	2.31	9.46	11.76	81.75	11.11
E	2.50	11.91	12.23	86.66	7.95	0.71	3.49	3.68	65.30	6.36
Avg.	**3.84**	**16.67**	**18.36**	**90.14**	**12.28**	**1.39**	**6.45**	**7.15**	**72.75**	**10.92**

considering the three components through a weighted PSNR average,[22] providing the coding efficiency result considering all components together.

This work presents the YUV-BDBR results to show the coding efficiency impact of our scheme and the U-BDBR and V-BDBR results to show the impacts of our scheme in each chrominance channel since this work focuses on chrominance coding.

Table 1 shows the obtained results when limiting the maximum chrominance QTMT depth level in 2 and 3 for the intra prediction. Considering that QTMT starts with a depth level equal to one for the intra prediction, QTMT depth 2 indicates that the 32×32 chrominance CUs can be split using only one level of QT or MTT structure. In contrast, QTMT depth 3 allows the encoder to split the CUs with two levels of QT, two levels of MTT, or one level of QT followed by one level of MTT structure.

Limiting the QTMT depth can provide expressive C-ETS and T-ETS. For instance, restricting the QTMT depth in 2 provides a C-ETS and T-ETS of up to 94.21% (Class D) and 20.12% (Class A2), respectively. When limiting the QTMT depth in 3, a C-ETS and T-ETS of up to 81.75% (Class D) and 19.03% (Class A2) are obtained. It is important to highlight that the class that attained the highest C-EST is not necessarily the same class with the highest T-ETS because the total encoding complexity distribution between luminance and chrominance components varies according to the QP and video sequence encoded.

As expected, the fewer possibilities there are for block partitioning to evaluate, the greater the time saving is; QTMT with depth 2 attained on average 90% and 12% of chrominance and total encoding time savings, respectively. However, these time savings come with a significant BDBR increase of 3.84%, 16.67%, and 18.36% in YUV-BDBR, U-BDBR, and V-BDBR, respectively.

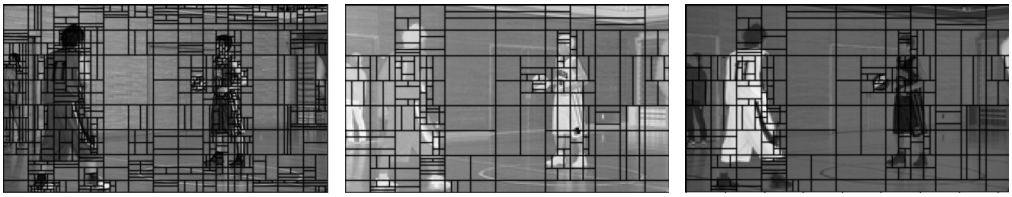
Table 2 presents the BDBR impact, C-ETS, and T-ETS when removing horizontal partitioning (BTH and TTH partitions) and vertical partitioning (BTV and TTV partitions) for chrominance QTMT structure. On average, when the chrominance horizontal partitioning is disabled, C-ETS and T-ETS are 82.90% and 12.50%, respectively, whereas with disabling the chrominance vertical partitioning, 78.65% and 11.71% of C-ETS and T-ETS are achieved, respectively. The average BDBR impact when removing horizontal partitioning is 2.48%, 12.06%, and 12.43% for YUV-BDBR, U-BDBR, and V-BDBR, respectively. Regarding the vertical partitioning, the BDBR increases by 2.03%, 9.21%, and 9.80% for YUV-BDBR, U-BDBR, and V-BDBR, respectively.

This analysis demonstrated that, although horizontal and vertical partitioning increases the encoding efficiency, most of the chrominance encoding complexity is related to the BT/TT partition direction.

Based on the presented analyses, one can notice that QT with nested MTT partition structure enables more flexible block partition types that adapt to a wide range of video characteristics, allowing for better compression performance. But, evaluating this high amount of splitting

Table 2 Encoding efficiency and time saving of removing the horizontal or vertical MTT partitions for chrominance blocks.

	Horizontal partitioning less					Vertical partitioning less				
Class	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS (%)	T-ETS (%)	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS (%)	T-ETS (%)
A1	3.91	16.23	18.17	82.06	11.48	1.98	9.00	9.52	76.78	11.41
A2	2.44	13.67	10.36	84.36	20.77	2.70	12.99	11.72	81.18	18.53
B	2.29	14.34	15.12	82.47	7.69	1.43	8.44	9.90	77.14	7.20
C	2.32	9.63	11.12	84.51	13.89	2.23	8.91	10.88	80.33	12.89
D	1.89	8.90	10.19	85.79	13.50	1.98	7.52	9.01	80.25	12.13
E	2.02	9.58	9.60	78.23	7.67	1.83	8.39	7.75	76.20	8.08
Avg.	**2.48**	**12.06**	**12.43**	**82.90**	**12.50**	**2.03**	**9.21**	**9.80**	**78.65**	**11.71**



(a) **Y** (b) **U** (c) **V**

Fig. 3 Block size distributions for (a) luminance; (b) chrominance (U); and (c) chrominance (V).

possibilities requires a high computational effort. However, the simple removal of some block partitioning evaluations is not appropriate for reducing the encoding effort since this decision can significantly degrade the encoding efficiency. Thus, solutions capable of adaptively limiting the chrominance QTMT depth and deciding the direction of BT/TT partitions are promising alternatives to reducing the encoding complexity and enabling real-time video encoding.

Figures 3(a), 3(b), and 3(c) present the block size distributions for luminance, chrominance (U), and chrominance (V), respectively, regarding the first frame of the BasketballPass video sequence encoded with AI configuration and QP 37. Since VVC encodes chrominance channels U and V together, the same block size distribution is obtained for both components.

One can notice that, for most cases, chrominance is encoded with larger blocks than luminance since chrominance has more homogeneous regions. However, the chrominance QTMT structure evaluates all splitting possibilities to find the best one, demanding a high computational effort. Since the best coding structure of luminance QTMT is obtained earlier than the chrominance QTMT coding, the luminance QTMT structure can be employed as a predictor to decide early termination of the chrominance QTMT evaluation.

3 Statistical Analysis and Proposed Solution

For I-slices, luminance and chrominance QTMT structures are obtained separately for each CTU; however, the chrominance CTU encoding is performed after encoding the associated luminance CTU. As these channels represent the same scenario, one can explore some correlations between the luminance and chrominance QTMT structures. Furthermore, the characteristics of the chrominance CU samples can also be explored to identify the behavior of the texture and predict the best chrominance block partitioning type. Section 3.1 presents the methodology used to develop the proposed solution. Sections 3.2 and 3.3 detail the proposed fast block partitioning scheme, which is composed of two solutions: (i) chrominance CU splitting early termination

(CSETL) based on Luminance QTMT and (ii) fast chrominance split decision based on variance of sub-blocks (FCSDV). Section 3.4 presents the overall fast block partitioning scheme and discusses the definition of the thresholds.

3.1 *Methodology*

All statistical analyses presented in this section followed the CTC encoder configurations and QP values specified by JVET. The experiments considered the AI configuration, where only intra prediction tools are available. The experiments were executed using the VTM software (version 10.0), which serves as a reference implementation of all encoding features defined in the VVC standard.

In the experiments used to collect the data used to define the proposed techniques, five video sequences with different resolutions were selected: TrafficFlow (3840×2160), ParkScene (1920×1080), Flowervase (416×240) (HEVC CTC[23]), Netflix_DrivingPOV (1280×720) (AV1 CTC[24]), and BasketballDrill (832×480) (VVC CTC[20]). Each video sequence is encoded with 80 frames for each QP.

These video sequences were selected based on spatial information (SI) and temporal information (TI).[25] SI estimates the amount of details within the video sequence, whereas TI measures the quantity of motion in the video sequence. Since the effectiveness of the analyses and developed solutions are highly linked to the diversity and relevance of the dataset, we selected video sequences with varied SI/TI values and different resolutions. This ensures a substantial variation in the information generated and increases the potential for developing a robust solution. Figure 4 shows the SI and TI for these video sequences used to collect data.

It is worth noting that, to provide a more robust solution, we selected different video sequences to collect data from those used in VVC CTC to obtain the experimental results of the proposed solution. Only one video sequence of VVC CTC was used in the data collection step (BasketballDrill). The remaining VVC CTC video sequences were used to evaluate the proposed solution, as presented in Sec. 4. Thus, this avoids overfitting the proposed solution.

During the evaluations, the success rate of each predictor was analyzed according to the QP and video sequence encoded. Equation (1) describes the success rate, which is the number of cases for which the proposed predictor had success divided by the total number of cases.

$$Success\ Rate = \frac{Predictor\ have\ success}{Total\ cases}(\%) \tag{1}$$

3.2 *Chrominance CU Splitting Early Termination based on Luminance QTMT*

We propose the following idea to explore the correlation of the chrominance and luminance QTMT structures. For a given chrominance CU being evaluated in the QTMT depth d, it is

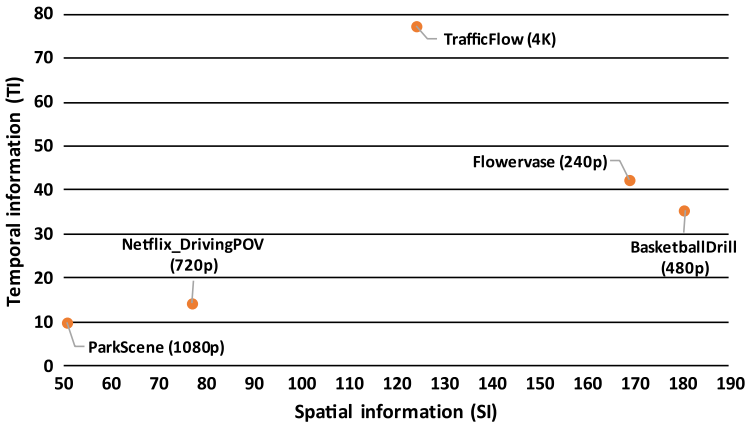


Fig. 4 Spatial and temporal information distribution of the training video sequences.

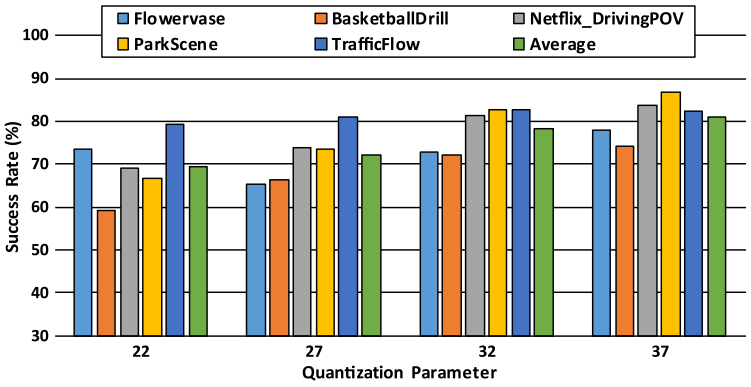


Fig. 5 Success rate of the best chrominance block size found in the QTMT depth lower or equal to luminance QTMT.

possible to verify the split type selected in the associated luminance CU for the same QTMT depth d.

Considering this fact, if the selected split type for the luminance CU in this depth is not split, a high probability of the best chrominance CU has been found at a depth lower or equal to d. Then, the process of dividing the chrominance CU ends; otherwise, the execution flow remains unchanged.

Figure 5 shows the success rate for the video sequences and each QP described in Sec. 3.1. The success rate was calculated according to Eq. (1), and in this case, the total number of cases is the number of evaluated chrominance blocks. The predictor has success when the chrominance block is encoded with a QTMT depth lower or equal to the QTMT depth of the associated luminance block. On average, this predictor has a success rate higher than 70% for all QPs evaluated. In addition, the success rate also increases as QP increases, reaching up to 81% for QP 37. On average, a success rate of 75% was obtained, indicating that the luminance QTMT structure can be a good predictor for the chrominance QTMT coding.

Along with this analysis, Fig. 6 shows the probability density functions for "not split" and "split" curves according to the rate-distortion (RD) cost divided by the block area since larger blocks tend to have larger RD-costs. The block area refers to the number of samples inside the block, and it is calculated by multiplying the block width by the block height. The chrominance RD-cost was collected for both curves when the colocated luminance CU was defined as not split. The not split indicates that the chrominance CU was not split when the colocated luminance CU also was defined as not split. The split indicates that the chrominance CU was split when the colocated luminance CU was defined as not split.

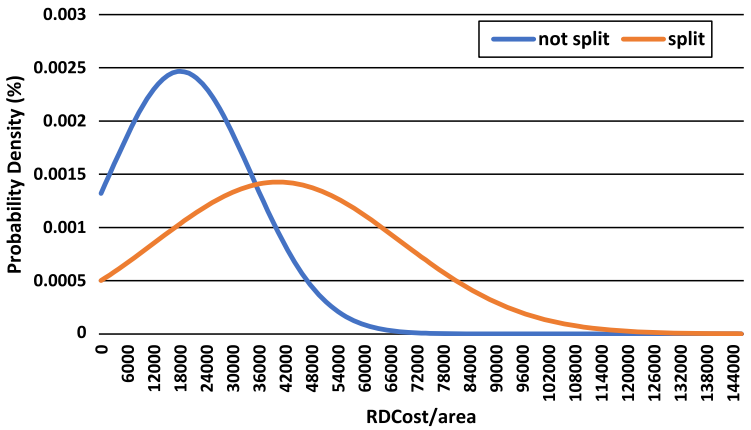


Fig. 6 Probability density function of splitting or not splitting the chrominance CU using RD-cost based on the luminance QTMT.

These functions, which follow a Gaussian distribution, were achieved by encoding the video sequences defined in Sec. 3.1. These results demonstrate a high probability of the luminance QTMT predictor having success for low RD-costs while having almost no chance of having success for larger values. Thus, this approach can be explored to perform an early termination decision according to a threshold criterion, increasing the success rate and decreasing the encoding efficiency loss. This strategy is explored in Sec. 3.4.1.

3.3 *Fast Chrominance Split Decision based on Variance of Sub-Blocks*

As discussed in the analysis of Sec. 2.1, horizontal and vertical evaluations of the BT/TT partitions increase the chrominance encoding time significantly. Therefore, a solution capable of deciding the direction of BT/TT partitions is a promising strategy for reducing the encoding complexity efficiently. Typically, the best BT/TT partition direction is highly linked to the texture direction.[13,16] This link happens because the best block partitioning is achieved with the partition type that results in a more homogeneous region or a region with similar behavior since it increases the accuracy of the prediction tools and reduces the residual error, incurring a smaller RD-cost.

Thus, we propose evaluating the variance of sub-blocks in the current encoding chrominance CU as a predictor to decide the BT/TT partition directions. Figure 7 demonstrates the calculation of horizontal and vertical variances for an 8×8 chrominance CU in Figs. 7(a) and 7(b), respectively.

To calculate the horizontal variance, we consider that the current CU is horizontally subdivided into two equal-sized sub-blocks (in a similar way as the BTH partitioning), resulting in two 8×4 sub-blocks (in the case of the example of Fig. 7). The var_{upper} and var_{lower} are obtained by calculating the variances of the highlighted regions in red and blue, respectively. Then, var_{hor} refers to the sum of the variances of the upper (var_{upper}) and lower (var_{lower}) partitions. Similarly, var_{ver} considers that the current CU is vertically subdivided (in a similar way as the BTV partitioning), where var_{left} and var_{right} are added to obtain the var_{ver}.

Considering that the lowest sum of variances indicates that the partition type provides more homogeneous regions, in the example of Fig. 7, the vertical partitioning could be skipped since the horizontal partitioning is the most promising for achieving a better coding efficiency.

The VVC encoder minimizes the RD-cost considering both chrominance components (U and V) to define the best block size and prediction mode combination. Consequently, the proposed predictor should also consider the features of both chrominance components to provide more accurate decisions. In this case, we calculate the var_{hor} and var_{ver} for the block samples of each chrominance component and only skip a given BT/TT direction if both components agree to skip that direction. For instance, the horizontal splitting is skipped only if var_{hor} (U) is lower than var_{ver} (U) and var_{hor} (V) is lower than var_{ver} (V). To simplify, we use only var_{hor} and var_{ver} to represent the condition of chrominance U and V.

Figure 8 shows the success rate of skipping the horizontal partitioning (BTH and TTH) when var_{ver} is lower than var_{hor} and skipping the vertical partitioning (BTV and TTV) when var_{hor} is lower than var_{ver}. Considering Eq. (1), the predictor has success when it decides to skip

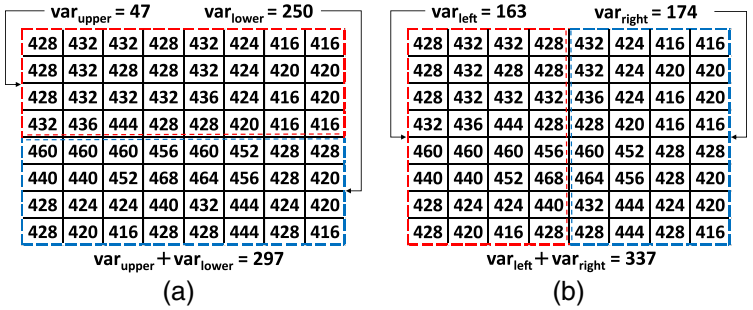


Fig. 7 Illustration of the calculation of (a) horizontal and (b) vertical variances for an 8×8 chrominance block.

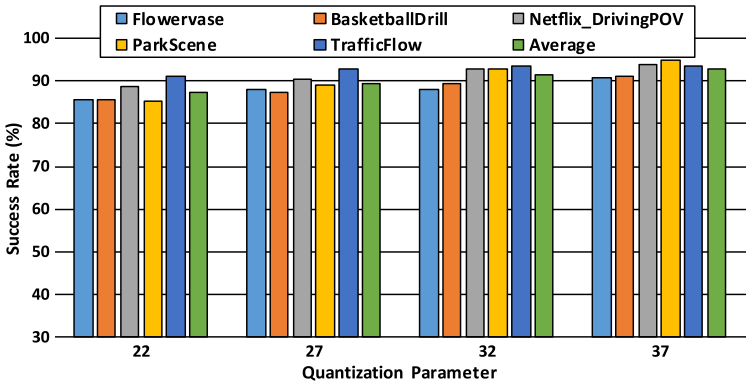


Fig. 8 Success rate of skipping the horizontal or vertical splitting of BT/TT partitions using the variance of sub-blocks.

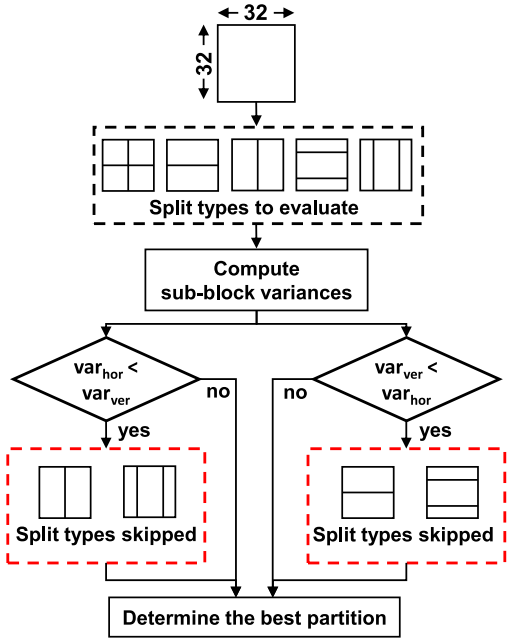


Fig. 9 Flowchart of the split decision based on the variance of sub-blocks.

the horizontal direction and the best split type selected is neither BTH nor TTH or if the predictor decides to skip the vertical direction and the best split type selected is neither BTV nor TTV. The total number of cases is the total number of chrominance blocks that evaluated at least one horizontal split and one vertical split. This evaluation shows that this predictor provides a more than 85% success rate for all cases evaluated. On average, this predictor presents a success rate of 90%, demonstrating that this solution can provide interesting results regarding encoding efficiency and complexity reduction.

Figure 9 shows the flowchart of the split decision based on the variance of sub-blocks for a 32×32 CU. In this case, the evaluation of vertical split types is skipped if var_{hor} is lower than var_{ver}; otherwise, the encoding flow follows without modifications. Analogously, the horizontal split types are skipped if var_{ver} is lower than var_{hor}.

3.4 *Fast Block Partitioning Scheme*

Figure 10 shows the flowchart of the proposed fast block partitioning scheme, encompassing the chrominance splitting early termination (CSETL) and the fast chrominance split decision (FCSDV).

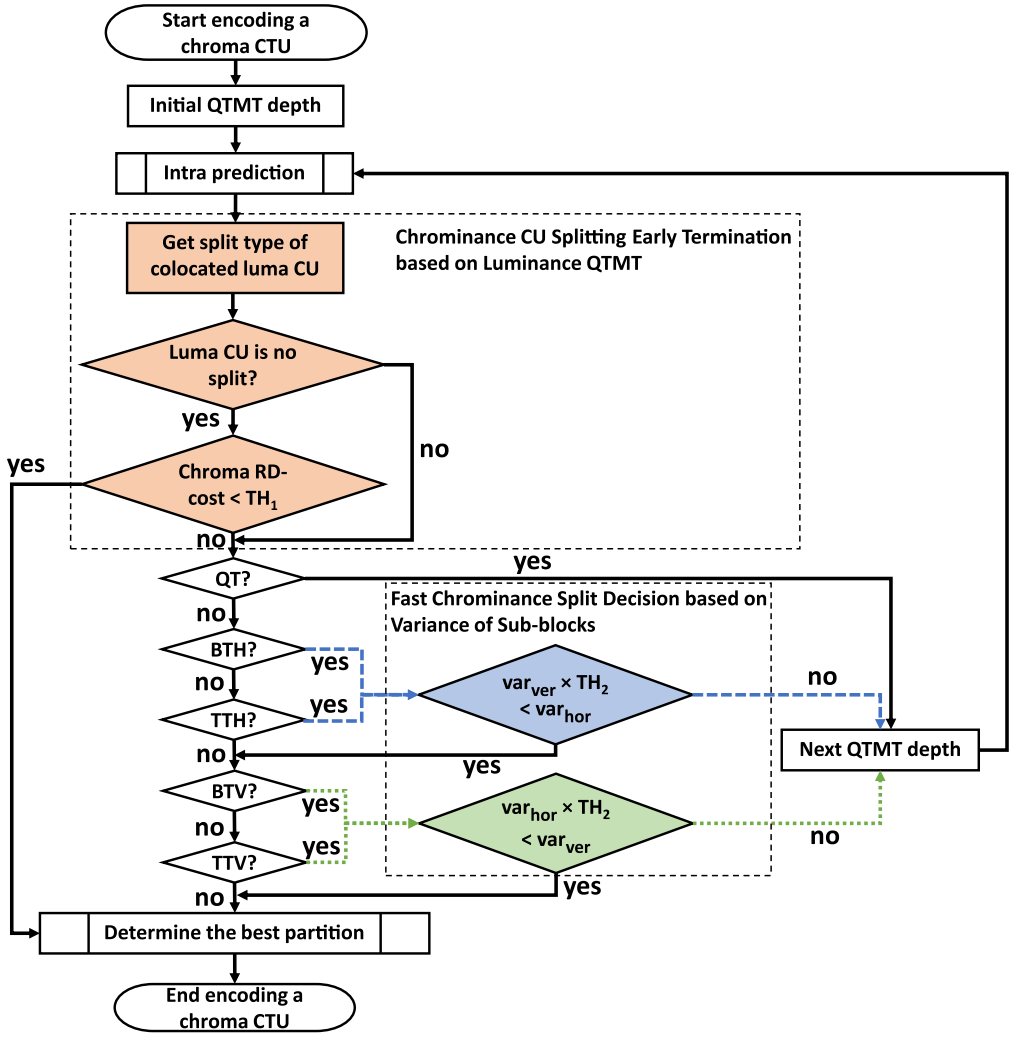


Fig. 10 Flowchart of the proposed fast block partitioning scheme for chrominance coding.

CSETL aims to terminate the chrominance QTMT evaluation early based on the luminance CU split type and the current chrominance RD-cost. When the luminance CU is not split and the chrominance RD-cost is lower than TH_1, the chrominance CU is not divided and the QTMT evaluation is finished; otherwise, the execution proceeds to evaluate other splitting possibilities.

FCSDV works to decide the direction of BT/TT partitions. On the one hand, when var_{ver} is lower than var_{hor}, the horizontal partitions are skipped; on the other hand, when var_{hor} is lower than var_{ver}, the vertical partitions are skipped. However, we decided to introduce another threshold criterion (TH_2) because very close variance values have no obvious texture direction, hampering the definition of the best split direction. Therefore, TH_2 is a value that defines the percent that a variance value of a given direction should be less than the variance value of the other direction to avoid the evaluation.

3.4.1 *Definition of thresholds*

We employed a detailed threshold analysis that evaluates five scenarios for each solution of the proposed scheme, considering C-ETS and YUV-BDBR since the threshold values impact the complexity reduction meaningfully. The obtained results in this section also consider the video sequences presented in Sec. 3.1. In addition, the obtained U-BDBR and V-BDBR results are also presented in this analysis.

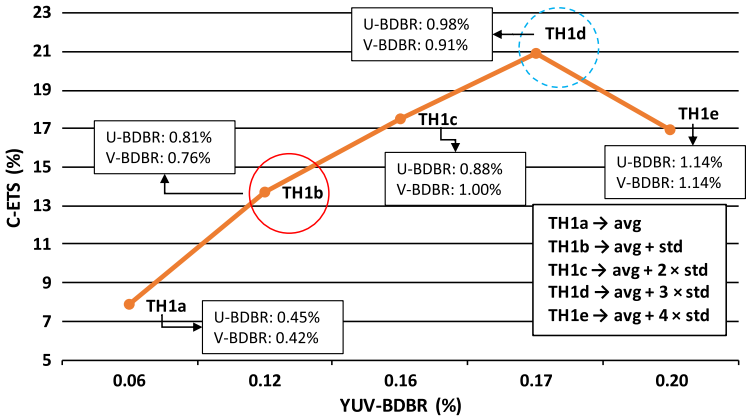


Fig. 11 Chrominance encoding time saving and BDBR impact for CSETL solution according to some threshold values.

Figure 11 shows the C-ETS and YUV-BDBR results collected for different threshold values (TH_1) used in CSETL. In the Sec. 3.2 analysis, the average and the standard deviation of the RD-cost per area have been saved to define the evaluation scenarios.

The evaluation scenarios for each threshold (TH_1) use Eq. (2), where k is empirically selected, ranging from 0 to 4, and μ and σ indicate the average and the standard deviation values for the TH computation, respectively.

$$TH_1 = \mu + k \times \sigma \tag{2}$$

CSETL allows for various operation points to provide better coding efficiency or higher complexity reduction. We have separated two interesting limits to evaluate as a case study (Sec. 4). The TH1b in the solid red line provides a good trade-off between YUV-BDBR and C-ETS, while the TH1d in the dotted blue line denotes the highest C-ETS.

Figure 12 exhibits the C-ETS and YUV-BDBR results according to TH_2 for the FCSDV solution. Since the analysis in Sec. 3.3 presented a high success rate for this solution, we decided to verify the impact of using some threshold values and using no threshold ("noTH" in Fig. 12). This threshold defines the percent that one variance must be less than the other to skip the BT/TT direction. For instance, $TH_2 = 110\%$ indicates that a given direction variance must be <10% of the other to skip a direction.

In this case, the higher the threshold is, the lower the C-EST and YUV-BDBR impact is. Following the same idea as the previous experiment, we selected two case studies. We chose the highlighted threshold with a solid red line (TH2a) since it can avoid the cases where the

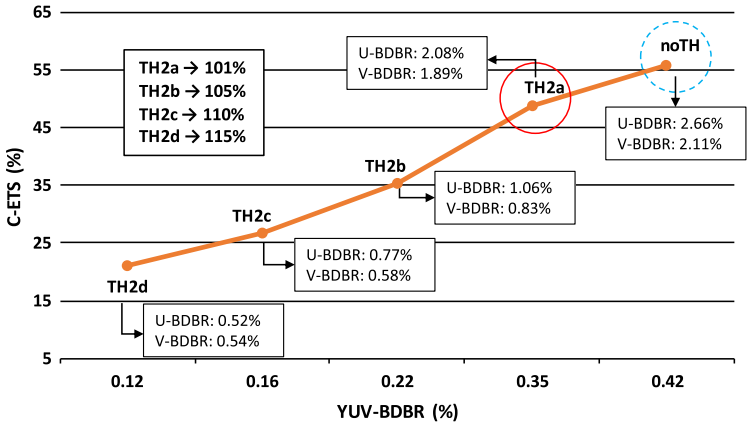


Fig. 12 Chrominance encoding time saving and BDBR impact for FCSDV solution according to some threshold values.

variance values are very close, reducing the YUV-BDBR loss while maintaining a high C-ETS. The noTH evaluation provides the highest C-ETS reduction; this evaluation associated with TH1d of the CSETL solution achieves the highest ETS of the proposed scheme.

Therefore, we selected two case studies considering two combinations of thresholds to be evaluated in the next section. Case 1 refers to the combination of TH1b (CSETL) and TH2a (FCSDV), which provides the best trade-off between coding efficiency and ETS. Case 2 indicates the combination of TH1d (CSETL) and noTH (FCSDV), which allows for the highest ETS of the proposed scheme.

4 Experimental Results

This section presents the results of the fast block partitioning scheme for chrominance intra prediction, which encompasses CSETL and FCSDV solutions. All experiments conducted in this paper were evaluated using VTM 10.0 and followed the JVET CTC for standard dynamic range (SDR) video sequences.[20] The evaluations were performed in the AI encoder configuration, which allows for only intra prediction tools. Six classes of video sequences are specified in CTC. Classes A1 and A2 are UHD video sequences (3840×2160) and class B encompasses Full-High Definition (FHD) video sequences (1920×1080). Classes C and D enclose 832×480 and 416×240 video sequences, respectively. Finally, class E covers 1280×720 video sequences. Each video sequence is encoded for each QP specified in CTC: 22, 27, 32, and 37.

We measured the encoding performance in terms of ETS and BDBR. Since the proposed solution focuses on the chrominance coding and most of the bits are used to encode the luminance component, the individual chrominance BDBR can become difficult to interpret.[22] Consequently, in addition to the chrominance BDBR results, we used YUV-BDBR to evaluate the quality impact in all video sequence channels and the total bitrate.

Table 3 shows the results acquired with the proposed solution for case 1 and case 2. "Average" refers to the average regarding of all video sequences, and "average (without TS *)" indicates the average of all video sequences excluding BasketballDrill, which was used in the data collection step (Sec. 3).

Case 1 is the configuration of thresholds that provides a good trade-off between compression performance and ETS. In this case, the proposed scheme can save the chrominance and total encoding time by about 60.03% and 8.18%, respectively, with a negligible impact of 0.66% in YUV-BDBR, on average (without considering BasketballDrill). Negligible variation in the average results is obtained when the training sequence is considered, demonstrating that the proposed scheme achieves excellent results regardless of the characteristics of the video sequence. On the one hand, the Campfire and ParkRunning3 video sequences produce the highest (72.97%) and the lowest (42.90%) C-ETS, respectively. On the other hand, Campfire and BQSquare result in the highest (1.79%) and lowest (0.20%) YUV-BDBR increase, respectively. These results demonstrated that the scheme could reduce more than 60% of the chrominance encoding time, on average, with minor impacts on the coding efficiency.

Furthermore, it is important to highlight that the highest C-ETS will not always result in the highest T-ETS since the luminance and chrominance complexity distribution varies according to the video sequence and QP encoded. For instance, ParkRunning3 obtained 42.90% of C-ETS and 14.04% of T-ETS, whereas MarketPlace attained 60.49% of C-ETS and 6.97% of T-ETS.

Considering case 2, which allows our scheme to achieve a higher ETS than other threshold combinations, the obtained C-ETS and T-ETS were 66.38% and 9.40%, respectively. This high complexity reduction impacts 0.81% in YUV-BDBR. In this case, a higher ETS was obtained at the cost of a YUV-BDBR increase compared with case 1. Regarding only the chrominance channels, case 1 increases the U-BDBR and V-BDBR by 3.42% and 3.58%, respectively, whereas case 2 impacts 4.10% and 4.22% in U-BDBR and V-BDBR, respectively.

Figure 13 shows the CU size distribution obtained with VTM without modifications (baseline) and the proposed scheme (case 1), considering the chrominance (U) channel of the first frame in the RaceHorses video sequence (class C) encoded with AI configuration and QP 32. One can notice that, even with the proposed scheme reducing the number of CUs evaluated, the final distribution of CUs chosen has a similar behavior compared with the baseline

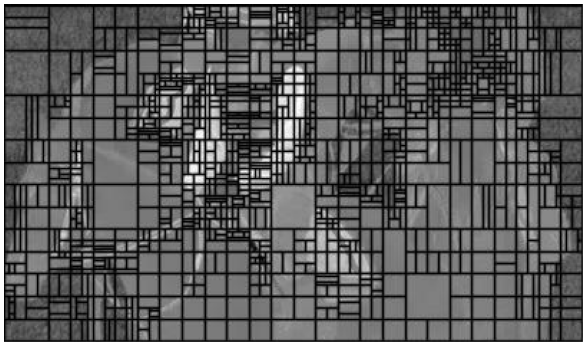
Table 3 Experimental results obtained with the proposed fast block partitioning scheme for chrominance coding under AI configuration.

		Case 1					Case 2				
					Time saving (%)					Time saving (%)	
Class	Sequence	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS	T-ETS	YUV BDBR (%)	U BDBR (%)	V BDBR (%)	C-ETS	T-ETS
A1	Tango2	1.07	7.91	8.40	64.29	4.87	1.11	8.26	8.69	65.75	5.72
	FoodMarket4	0.76	3.37	3.71	62.84	6.85	0.82	3.69	3.99	66.23	6.93
	Campfire	1.79	5.96	8.35	72.97	14.73	2.33	8.54	9.30	81.18	17.15
A2	CatRobot	1.54	7.70	7.32	68.73	11.76	1.93	9.51	9.12	74.77	13.16
	DaylightRoad2	0.39	4.98	3.74	58.52	4.77	0.46	6.07	4.32	64.08	6.56
	ParkRunning3	0.45	0.90	0.85	42.90	14.04	0.90	1.94	1.75	66.73	21.59
B	MarketPlace	0.55	3.90	2.95	60.49	6.97	0.72	5.15	3.56	68.86	8.95
	RitualDance	0.62	3.53	3.91	59.37	5.79	0.70	3.98	4.50	64.29	5.88
	Cactus	0.55	2.97	3.67	58.03	5.59	0.70	3.76	4.67	65.02	5.07
	BasketballDrive	0.64	3.17	3.70	58.42	4.11	0.75	3.78	4.35	63.71	6.23
	BQTerrace	0.21	3.12	2.92	53.95	6.24	0.28	3.72	3.54	61.61	7.09
C	BasketballDrill*	1.06	3.91	4.18	59.76	9.63	1.29	4.70	5.01	66.07	10.83
	BQMall	0.62	3.34	3.46	63.08	9.64	0.71	3.68	4.11	67.11	9.78
	PartyScene	0.39	2.35	2.53	57.50	10.02	0.50	2.95	3.26	63.94	11.59
	RaceHorsesC	0.52	1.61	2.43	64.20	10.78	0.67	1.81	2.97	70.57	12.28
D	BasketballPass	0.88	3.47	3.08	59.75	8.05	1.08	4.02	3.76	65.91	9.60
	BQSquare	0.20	1.67	2.42	53.18	5.16	0.22	1.76	2.69	58.80	5.77
	BlowingBubbles	0.42	1.80	2.33	57.61	9.74	0.49	2.48	2.61	62.86	9.18
	RaceHorses	0.71	2.18	3.04	62.47	10.28	0.85	2.56	3.82	69.80	11.82
E	FourPeople	0.36	1.74	1.91	57.23	6.22	0.41	1.90	2.33	60.96	6.25
	Johnny	0.72	3.73	2.26	62.51	6.51	0.76	3.92	2.74	65.99	8.07
	KristenAndSara	0.52	2.46	2.21	62.65	9.64	0.58	2.70	2.59	65.82	8.79
Average		0.68	3.44	3.61	60.02	8.24	0.83	4.13	4.26	66.37	9.47
Average (without TS*)		**0.66**	**3.42**	**3.58**	**60.03**	**8.18**	**0.81**	**4.10**	**4.22**	**66.38**	**9.40**

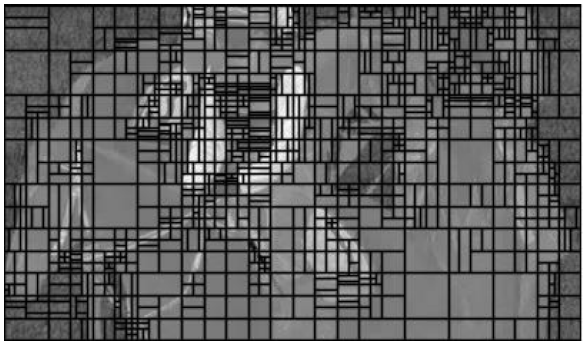
Note: * refers to the Training Sequence (TS) BasketballDrill.

implementation, indicating that our scheme achieves a high ETS while keeping the encoding efficiency.

To the best of our knowledge, this is the first solution to reduce the encoding complexity of chrominance block partitioning in the VVC encoder, presenting the results of chrominance time saving and compression performance. Although other works [12–16] also proposed fast block partitioning solutions for the VVC encoder and achieved excellent complexity reduction results, they focused only on luminance blocks, hampering a fair comparison. Nonetheless, it is important to highlight that our scheme can be combined with solutions focusing on luminance blocks



(a) **Baseline**



(b) **Proposed scheme (Case 1)**

Fig. 13 CU size distribution obtained with (a) baseline and (b) proposed scheme (case 1) for chrominance (U) channel.

to achieve even higher complexity reductions. Furthermore, since DT coding structure is a recent tool adopted in the video coding standard, this comparison cannot be performed with predecessor standards.

5 Conclusions

This paper presented a fast chrominance block partitioning scheme to reduce the VVC intra coding complexity. The proposed scheme includes CSETL and FCSDV for early termination of the chrominance QTMT evaluation and deciding the best direction of BT/TT partitions, respectively. The CSETL solution correlates the chrominance and luminance QTMT to determine early termination of the chrominance evaluation based on a threshold criterion. FCSDV explores the variance of the sub-blocks of both chrominance components to identify the texture direction and decide on splitting the block horizontally or vertically. The thresholds have been defined under extensive experimentation. The proposed scheme was evaluated according to the JVET CTC under AI encoder configuration, obtaining a chrominance encoding time saving of 60.03% with a drawback of 0.66% in YUV-BDBR. These results demonstrate that the solution could contribute to the development of the VVC encoder for real-time processing.

Acknowledgments

The authors thank the FAPERGS, CNPq, and CAPES (Finance Code 001) Brazilian research support agencies that financed this investigation.

References

1. D. Marpe, T. Wiegand, and G. J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Commun. Mag.* **44**(8), 134–143 (2006).

2. G. J. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012).
3. ITU-T and ISO/IEC, "Versatile Video Coding," ITU-T Rec. H.266 and ISO/IEC 23090-3 (2020).
4. Y.-W. Huang et al., "A VVC proposal with quaternary tree plus binary-ternary tree coding block structure and advanced coding techniques," *IEEE Trans. Circuits Syst. Video Technol.* **30**(5), 1311–1325 (2019).
5. J. Chen, Y. Ye, and S. Kim, "Algorithm description for versatile video coding and test model 10 (VTM 10)," JVET-S2002 v1 (2020).
6. K. Zhang et al., "Multi-model based cross-component linear model chroma intra-prediction for video coding," in *IEEE Vis. Commun. Image Process.*, IEEE, pp. 1–4 (2017).
7. X. Zhao et al., "Enhanced multiple transform for video coding," in *Data Compression Conf.*, IEEE, pp. 73–82 (2016).
8. M. Koo et al., "Low frequency non-separable transform (LFNST)," in *Picture Coding Symp.*, IEEE, pp. 1–5 (2019).
9. JVET, "VVC Test Model (VTM)," 2020, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/releases/VTM-10.0 (accessed Jan. 2021).
10. M. Saldanha et al., "Complexity analysis of VVC intra coding," in *IEEE Int. Conf. Image Process.*, IEEE, pp. 3119–3123 (2020).
11. F. Pakdaman et al., "Complexity analysis of next-generation VVC encoding and decoding," in *IEEE Int. Conf. Image Process.*, pp. 3134–3138 (2020).
12. H. Yang et al., "Low-complexity ctu partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.* **30**(6), 1668–1682 (2020).
13. T. Fu et al., "Fast Cu partitioning algorithm for H.266/VVC intra-frame coding," in *IEEE Int. Conf. Multimedia and Expo*, IEEE, pp. 55–60 (2019).
14. M. Lei et al., "Look-ahead prediction based coding unit size pruning for VVC intra coding," in *IEEE Int. Conf. Image Process.*, IEEE, pp. 4120–4124 (2019).
15. J. Zhao, Y. Wang, and Q. Zhang, "Adaptive CU split decision based on deep learning and multifeature fusion for H.266/VVC," *Sci. Prog.* **2020**, 1–11 (2020).
16. M. Saldanha et al., "Fast partitioning decision scheme for versatile video coding intra-frame prediction," in *IEEE Int. Symp. Circuits and Syst.*, IEEE, pp. 1–5 (2020).
17. S. S. Sengar and S. Mukhopadhyay, "Motion segmentation-based surveillance video compression using adaptive particle swarm optimization," *Neural Comput. Appl.* **32**(15), 11443–11457 (2020).
18. S. Zebhi, S. M. AlModarresi, and V. Abootalebi, "Video classification by fusing two-stream image template classification and pretrained network," *J. Electron. Imaging* **29**(5), 053011 (2020).
19. N. Mansouri et al., "Bayesian model for pedestrian's behavior analysis based on image and video processing," *J. Electron. Imaging* **30**(1), 013019 (2021).
20. F. Bossen et al., "VTM common test conditions and software reference configurations for SDR video," Joint Video Experts Team (JVET) of ITU-T SG 16 (2020).
21. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," VCEG-M33 VCEG-M33 (2001).
22. J. Ström et al., "Working practices using objective metrics for evaluation of video coding efficiency experiments," HSTP-VID-WPOM Technical Report, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG, 2020, http://www.itu.int/pub/T-TUT-ASC-2020-HSTP1/ (accessed 16 September 2021).
23. K. Sharman and K. Suehring, "Common test conditions (JCTVC-z1100)," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 (2017).
24. T. Daede, A. Norkin, and I. Brailovskiy, "Video codec testing and quality measurement," draft-ietf-netvc-testing-08 (work in progress) 8, 23 (2019).
25. ITU-T, "Recommendation P.910: Subjective video quality assessment methods for multimedia applications," International Telecommunication Union E 33690 (2008).

Mário Saldanha received his BS degree in computer science from the Federal University of Pelotas (UFPel), Pelotas, RS, Brazil, in 2016. In 2018, he received his MSc degree in computer

science from the UFPel. He is currently pursuing his PhD in computer science at the UFPel. He is a member of the Video Technology Research Group (ViTech) at the UFPel. His research interests include complexity reduction and hardware-friendly algorithms for 2D/3D video coding.

Gustavo Sanchez has been a professor at the IFFarroupilha, Brazil, since 2014. He received his electrical engineer degree from the Sul-Rio-Grandense Federal Institute of Education, Science and Technology in 2013 and his BS degree in computer science from the Federal University of Pelotas (UFPel) in 2012. In 2014, he received his MSc degree in computer science from the UFPel. Sanchez obtained his PhD in computer science at the Pontifical Catholic University of Rio Grande do Sul in 2019. He has more than 10 years of research experience in algorithms and hardware architectures for video coding. His research interests include complexity reduction algorithms, hardware-friendly algorithms, and dedicated hardware design for 2D/3D video coding.

César Marcon has been a professor at the Polytechnic School of Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil, since 1995. He received his PhD in computer science from Federal University of Rio Grande do Sul, Brazil, in 2005. He is a senior member of the Institute of Electrical and Electronics Engineers (IEEE) and of the Brazilian Computer Society (SBC). He is a distinguished Brazilian researcher with a CNPq PQ-2 grant. He is an advisor of MSc and PhD graduate students at the Graduate Program in Computer Science (PPGCC) of PUCRS. He has published more than 150 papers in prestigious journals and conference proceedings. Since 2005, he has coordinated several research projects in the areas of telecom and healthcare. His research interests are embedded systems in the telecom domain, MPSoC architectures, partitioning and mapping application tasks, fault tolerance, and real-time operating systems.

Luciano Agostini is a distinguished researcher through a CNPq PQ-1D grant. He received his MS degree and PhD from Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 2002 and 2007, respectively. He has been a professor since 2002 at Federal University of Pelotas (UFPel), Brazil, where he leads the Video Technology Research Group (ViTech). He is an advisor at the UFPel master and doctorate in computer science programs. He was the executive vice president for Research and Graduate Studies of UFPel from 2013 to 2017. He has published more than 300 papers in journals and conference proceedings. His research interests include 2D and 3D video coding, algorithmic optimization, arithmetic circuits, and digital systems. Currently, he is an associate editor of two IEEE CAS journals: TCSVT and OJCAS. He is a senior member of IEEE and ACM. He is a member of the IEEE CAS, CS, and SPS societies. At the IEEE CAS, he is a member of the MSATC. He is also member of the SBC and SBMicro Brazilian societies.