



HAL
open science

Fast Intra mode prediction algorithms for SCBs in VVC SCC

Dayong Wang, Yishen Deng, Weisheng Li, Xin Lu, Frédéric Dufaux, Bo Hang,
Ce Zhu

► **To cite this version:**

Dayong Wang, Yishen Deng, Weisheng Li, Xin Lu, Frédéric Dufaux, et al.. Fast Intra mode prediction algorithms for SCBs in VVC SCC. International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2024), IEEE, Apr 2024, Seoul, South Korea. hal-04413636

HAL Id: hal-04413636

<https://hal.science/hal-04413636v1>

Submitted on 20 Mar 2024

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

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

FAST INTRA MODE PREDICTION ALGORITHMS FOR SCBS IN VVC SCC

Dayong Wang^{1,2}, Yishen Deng³, Weisheng Li^{3*}, Xin Lu⁴, Frederic Dufaux⁵, Bo Hang⁶, Ce Zhu⁷

¹Key Laboratory of Big Data Intelligent Computing, Chongqing University of Posts and Telecommunications

²Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University

³Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications

⁴Faculty of Computing, Engineering and Media (CEM), De Montfort University

⁵Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes

⁶the Computer School, Hubei University of Arts and Science

⁷School of Information and Communication Engineering,

University of Electronic Science and Technology of China

wangdayong@cqupt.edu.cn, 739661712@qq.com, liws@cqupt.edu.cn, xin.lu@dmu.ac.uk,

frederic.dufaux@l2s.centralesupelec.fr, bohang@163.com, eczhu@uestc.edu.cn

ABSTRACT

Versatile Video Coding (VVC) now supports Screen Content Coding (SCC) by integrating two efficient coding modes: Intra Block Copy (IBC) and Palette (PLT). However, the numerous modes and the Quad-Tree Plus Multi-Type Tree (QTMT) structure inherent to VVC contribute to a very high coding complexity. To effectively reduce the computational complexity of VVC SCC, we propose a fast Intra mode prediction algorithm for VVC SCC. More specifically, we first use the difference of minimum Sum of Absolute Transformed Differences (SATD) value of four Directional Modes (DMs) of Intra and the SATD value of the IBC-merge mode to determine whether to early skip Intra checking. Subsequently, we use a decision tree to determine whether to early terminate the checking after block differential pulse coded modulation (BDPCM). Finally, we employ a decision tree to determine whether to early skip multiple transform selection (MTS) and low frequency non-separable transform (LFNST) checking. The results demonstrate that our algorithm achieves an average encoding time reduction of 34.34% with a negligible Bjøntegaard delta bitrate increase of 0.46%.

Index Terms— Screen Content Coding (SCC), Versatile Video Coding (VVC), Bayesian model, decision tree, Intra mode

*Corresponding author

This research is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62020106011; in part by the National Key Research and Development Program of China under Grant 2023YFC3806003; in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K202100604; and in part by Guangxi Key Laboratory of Machine Vision and Intelligent Control under Grant 2023B05.

1. INTRODUCTION

In recent years, with the growing prevalence of screen content video applications such as screen sharing, web conferencing, online education, and so on. These types of video content are generated by computers. They typically exhibit repeated patterns, extensive flat regions, limited color values, and a lack of sensor noise. Traditional video coding methods cannot efficiently handle with these videos. In order to address these issues, High Efficiency Video Coding (HEVC) introduces two Screen Content coding modes: intra block copy (IBC) [1] and palette mode (PLT) [2]. These two modes significantly improve performance when dealing with screen content video. As the latest international video coding standard, Versatile Video Coding (VVC) also supports SCC. By introducing new coding tools and the quad-tree plus multi-type tree (QTMT) structure, VVC achieves coding efficiency, but its coding complexity is also very high. With the introduction of IBC and PLT modes, VVC SCC becomes even more complex, which seriously hinders the practical applications of VVC SCC. Therefore, to meet the demands of practical applications, there is an urgent need to significantly improve the coding speed of VVC SCC.

2. RELATED WORK

In recent years, acceleration algorithms for SCC can be roughly categorized into two groups: feature-based acceleration algorithms and content-based acceleration algorithms.

Feature-based acceleration methods [3–8] aimed to establish relationships between statistical features and mode decision, as well as CU partitioning decisions. Authors in [3, 4] used numerous handcrafted features to construct decision

trees and random forests to predict candidate modes. [5, 6] utilized encoding information and other features to predict candidate CU partitions. In addition to separate feature-based mode prediction and partition prediction, studies in [7, 8] used features to predict both candidate modes and partitions. Content-based acceleration algorithms were presented in [9–12]. In [9, 10], decision trees were employed to classify CUs into screen content blocks (SCBs) and natural image blocks (NIBs), followed by the prediction of candidate modes based on the content type of a CU. In [11], traditional neural networks were used to categorize CUs into SCBs and NIBs, and then candidate modes were assigned to different types of CUs. [12] proposed a CNN-based classifier to categorize CUs into NCBs and SCBs, and subsequently predict likely modes based on their corresponding CU types.

The algorithms mentioned above have been developed to predict candidate modes or partitions, and unlikely ones are skipped to improve coding speed. However, fast Intra mode prediction algorithms for SCBs in VVC SCC have not yet been investigated. If we can develop a fast Intra-mode prediction algorithm for SCBs in VVC SCC, its coding speed can also be improved.

3. THE PROPOSED INTRA OPTIMIZATION METHOD

In VVC, the Intra mode involves a complex process, as illustrated in Fig. 1(a). VVC employs the rough mode decision (RMD) to select a few modes with the least sum of absolute transformed differences (SATD) among the 67 available Intra directional prediction modes. These selected modes are used to construct an RD-list that needs to be subsequently traversed. Furthermore, the RD-list is extended to include various modes such as matrix-weighted intra prediction (MIP) and intra sub-partitions (ISP). In the rate-distortion optimization (RDO) stage, in addition to using the DCT-II transform, VVC also applies low-frequency non-separable transform (LFNST) and multiple transform selection (MTS) to enhance coding performance. In order to improve coding speed, our approach is depicted in Fig. 1(b), where we introduce three strategies into the original Intra mode process. Next, we will elaborate on the specifics of these three strategies.

In order to develop the three strategies, the relative data statistics are investigated based on extensive experiments. To prevent overlap between the training and testing sets, we used text and graphics with motion (TGM) sequences that are not included in CTC [13]. These sequences include PptDocXls, RealTime-Data, WordEditing, ClearTypeSpreadsheet, and PcbLayout, all of which exclusively consist of SCBs. We employed the original encoder VTM-17.0 [14] and set QP values to 22, 27, 32, and 37 to encode each sequence.

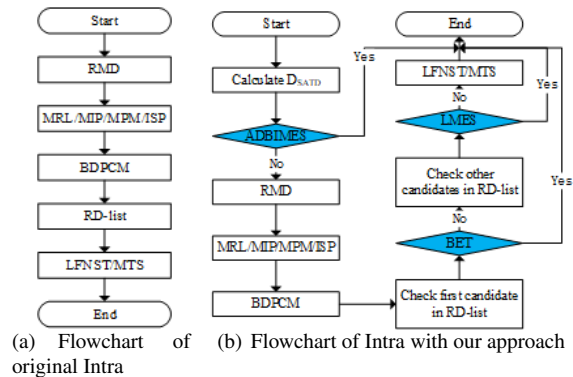


Fig. 1: Flowchart of Intra mode

3.1. SATD Difference Based Intra Mode Early Skip (AD-BIMES)

SATD is calculated as the sum of the absolute values of coefficients resulting from the application of the Hadamard transform to the residual signal. In the Intra mode of the VVC, SATD is employed to perform rough selection of directional modes (DMs), which plays a crucial role in constructing the final RD-list. To explore the distribution of these DMs, their corresponding distribution is depicted in Fig. 2.

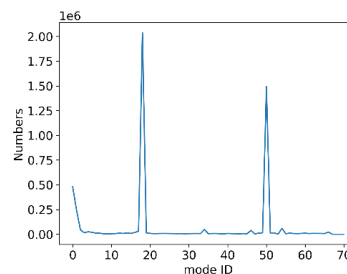


Fig. 2: DM distribution of Intra mode for TGM

In Fig. 2, the horizontal axis represents the index of DMs, and the vertical axis corresponds to the number of DMs. From Fig. 2, we can observe that most of optimal DMs fall into modes 0, 1, 18, or 50. In addition to the Intra mode, VVC needs to further check IBC, IBC-merge, and PLT, and select the mode with the smallest RDcost as the best mode. Since the coding process of the IBC-merge mode is very simple, if the minimum SATD value of these four DMs of the Intra mode is significantly greater than the SATD value of the IBC-merge mode, then it is very unlikely that the Intra mode is the best one and it can be skipped early. Therefore, we can investigate the difference between the minimum SATD value of the four DMs of the Intra mode and the SATD value of the IBC-merge mode to determine whether the Intra mode can be early skipped.

Based on the above analysis, we first examine IBC-merge mode and the four DMs of the Intra mode. Subsequently, we

employ their difference, denoted as D_{SATD} , as the sole feature for constructing a Bayesian decision model. To be more specific, we extract D_{SATD} and calculate the posterior probability $P_s(\omega_{i=Intra}|D_{SATD})$, which represents the likelihood that the Intra mode is the optimal mode, given the CU size s and D_{SATD} . According to Bayes' rule, we can calculate the probability of the Intra mode being chosen, denoted as $P_s(\omega_{i=Intra}|D_{SATD})$, as follows:

$$P_s(\omega_{i=Intra}|D_{SATD}) = \frac{P_s(D_{SATD}|\omega_{i=Intra})P_s(\omega_{i=Intra})}{P_s(D_{SATD})} \quad (1)$$

where s can encompass all partition sizes except for 64×64 , as the number of 64×64 CUs is relatively small and lacks strong regularity. We do not consider this size for processing. $P_s(\omega_{i=Intra})$ represents the prior probability of the Intra mode, $P_s(D_{SATD})$ represents the total probability density of D_{SATD} at that size, and $P_s(D_{SATD})$ is obtained using the following equation:

$$P_s(D_{SATD}) = \sum_{\omega_i} P_s(D_{SATD}|\omega_i)P_s(\omega_i) \quad (2)$$

where $i \in \{Intra, non - Intra\}$, and all parameters can be obtained through offline statistics. However, representing $P_s(D_{SATD}|\omega_i)$ with a specific probability function is challenging. Therefore, we employ a non-parametric estimation method to estimate the likelihood function:

$$P_s(D_{SATD}|\omega_i) = \frac{N_s^{D_{SATD}, \omega_i}}{N_s^{\omega_i}} \quad (3)$$

where $N_s^{D_{SATD}, \omega_i}$ represents the number of CUs with the difference D_{SATD} and mode belonging to ω_i under size s , and $N_s^{\omega_i}$ represents the count of all CUs with a mode belonging to ω_i under size s . Suppose α is the best threshold, if the Intra mode can be early skipped, the following condition should be satisfied:

$$P_s(\omega_{i=Intra}|D_{SATD}) < \alpha \quad (4)$$

The selection of the best threshold, α , will be illustrated in the section of experiment result.

3.2. BDPCM Early Termination (BET)

In VVC SCC, block differential pulse-coded modulation (BDPCM) is integrated into the Intra mode and executed prior to RD-list traversal. BDPCM supports two prediction directions: horizontal and vertical. Furthermore, prediction residuals are quantized directly, bypassing the transformation process. BDPCM proves highly effective in screen content coding. We conducted statistical analysis on the final selection proportion of BDPCM within Intra modes for SCBs, as shown in Table 1.

From Table 1, we can observe that the proportion of BDPCM is close to half for SCBs. Therefore, if we can deter-

Table 1: BDPCM distribution in Intra mode for screen content

BDPCM (%)	Non-BDPCM (%)
40.98	59.02

mine that BDPCM is the best mode and skip the subsequent complex RDO process, coding speed can be significantly improved. In order to determine whether BDPCM is the best mode, we have examined the first mode in the RD-list and BDPCM. We then utilize decision trees as the classification tool for predicting early terminations of BDPCM. The accuracy of decision trees relies on feature selection. We use the CU size information and intermediate coding information, which includes the size of CU (f_{SC}), quantization parameter (QP), the RD cost (f_{RDC}), and the coding bit (f_{CB}) of the BDPCM mode, and the mode of the parent CU (f_{MPC}). Both f_{RDC} and f_{CB} can indicate the encoding performance of BDPCM, and if the parent CU's optimal choice within Intra is BDPCM, the likelihood of the CU selecting BDPCM is very high. Once these features are obtained, we input them into a decision tree to obtain the probability of using Non-BDPCM, denoted as P_1 . Suppose β is the best threshold. If the BDPCM mode can be terminated, the following condition should be satisfied:

$$P_1 < \beta \quad (5)$$

The selection of the best threshold, β , will be illustrated in the section of experiment result.

3.3. LFNST and MTS Early Skip(LMES)

In VVC, the main transform supports five different combinations of transform types. Aside from the traditional (DCT-2, DCT-2) combination, four new combinations using multiple transform selection (MTS) were introduced: (DST-7, DST-7), (DST-7, DCT-8), (DCT-8, DST-7), and (DCT-8, DCT-8). In specific scenarios, MTS can effectively reduce the correlation in prediction residuals. For instance, in Intra mode prediction residuals, prediction errors often increase as the distance from boundary samples increases. In such cases, MTS could yield better results. Additionally, the low-frequency non-separable transform (LFNST) was introduced as a secondary transform, combined with DCT-2, to further reduce spatial redundancy in residuals. However, for SCBs, the newly introduced transforms might not be as effective due to their distinctive texture structures. We conducted statistical analysis on the optimal transform selections for Intra modes within SCBs, as shown in Table 2.

From Table 2, we can observe that the percentages of using DCT-II, LFNST, and MTS are 79.35%, 7.38%, and 13.27%, respectively. Obviously, the possibility of requiring multiple transforms or secondary transforms for residuals is

Table 2: Transform distribution in Intra mode for screen content

DCT-II (%)	LFNST (%)	MTS (%)
79.35	7.38	13.27

relatively low for SCBs. As a result, after checking DCT-2, we use a decision tree model to calculate the probability of requiring LFNST or MTS, denoted as P_2 . Suppose γ is the best threshold, we can skip checking the subsequent LFNST and MTS processes if the following conditions are met:

$$P_2 < \gamma \quad (6)$$

The selection of best threshold, γ , will be illustrated in the section of experiment results. In terms of feature selection, we have also incorporated additional intermediate coding features, including the size of the transform Unit (f_{STU}), QP, the variance of residual coefficients (f_{VRC}), the RD cost (f_{RDC}), and the coding bit (f_{CBD}) of DCT-2, the transform type of parent CU (f_{TTC}), and the RD cost of ISP (f_{RDI}).

4. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed Intra mode optimization algorithm on screen content, we rigorously adhere to the common test conditions (CTC) and utilize the VTM-17.0 encoder with IBC and PLT enabled. We use it to encode the recommended TGM sequences with QP values of 22, 27, 32, and 37 under All Intra configuration. The coding efficiency was measured using BDBR [15] method, and the percentage of time savings is denoted as ΔT .

In Section 3, it is evident that the performance of our proposed algorithm depends on thresholds. To determine the appropriate threshold, we conducted extensive individual experiments for each technique. The results are depicted in Fig. 3. For each strategy, as the threshold increases, the reduction in encoding time becomes more pronounced, and correspondingly, the BDBR also increases. However, in general, substantial time reductions were observed for each strategy, with a negligible increase in BDBR. Among these strategies, the ADBIMES provided the most significant reduction in encoding time. This can be attributed to the fact that, nearly the entire Intra mode can be effectively skipped by this strategy. As α varies from 0.05 to 0.25, ADBIMES offers an encoding time reduction ranging from 10.98% to 27.83%, with a BDBR increase of only 0.06% to 0.55%. Compared to ADBIMES, BET and LMES exhibited relatively modest time reductions under the same BDBR growth.

In order to strike a favorable balance between computational complexity and coding performance, we chose $\alpha = 0.15$, $\beta = 0.2$ and $\gamma = 0.1$ as our final settings, which ensures a minimal increase in BDBR while yielding higher time reduction benefits.

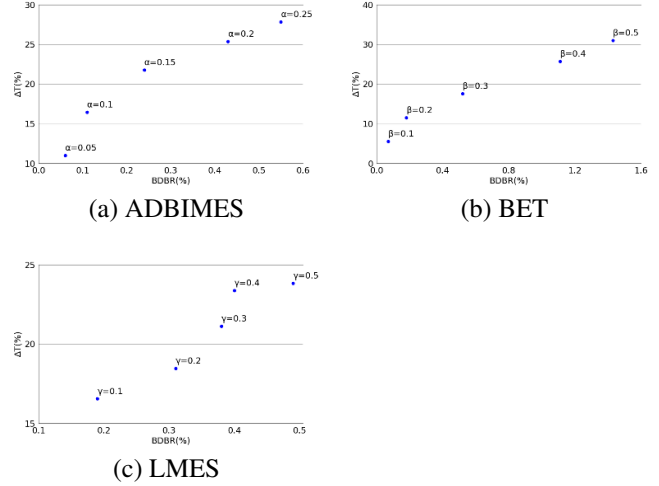


Fig. 3: Average BDBR and ΔT for different threshold values.

To demonstrate the overall performance, we compare the proposed algorithm with those suggested in [3], [7], and [12]. Since this algorithm is only developed for Intra mode of screen content, only TGM sequences are selected to test. The comparison of coding performance and coding speed is shown in Table 3.

Table 3: Overall performance comparison

Sequences	[3]		[7]		[12]		Proposed	
	BDBR(%)	ΔT (%)	BDBR(%)	ΔT (%)	BDBR(%)	ΔT (%)	BDBR(%)	ΔT (%)
Console	2.33	27.37	0.49	16.74	1.98	23.69	0.24	36.36
Desktop	1.56	30.89	0.50	20.72	2.24	32.29	0.18	51.02
FlyingGraphics	1.71	24.28	0.57	13.19	6.66	18.51	0.55	34.94
Map	1.65	20.56	1.03	10.84	2.28	26.87	0.43	21.02
Programming	1.29	23.09	0.79	9.73	3.64	24.55	0.56	35.71
SlideShow	1.74	17.86	0.85	11.01	3.77	23.19	1.26	18.44
WebBrowsing	1.87	26.35	0.46	17.35	7.01	17.62	-0.04	42.90
Average	1.74	24.34	0.67	14.22	3.94	23.82	0.46	34.34

From Table 3, we can observe that for all tested TGM sequences, the algorithm consistently ensures robust coding performance. Compared to other acceleration methods, our approach excels in both encoding performance and computational complexity. Not only our method achieves a more substantial reduction in encoding time but it also keeps the increase in BDBR significantly lower.

5. CONCLUSIONS

In this paper, we present an Intra mode optimization algorithm tailored for screen content with the aim of reducing the computational complexity of SCC in VVC. Experimental results demonstrate that, in comparison to conventional SCC in VVC the algorithm achieve an average encoding time reduction of 34.34% at a neglect increase of 0.46% in BDBR.

6. REFERENCES

- [1] 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 Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 409–419, 2016.
- [2] Z. Ma, W. Wang, M. Xu, and H. Yu, "Advanced screen content coding using color table and index map," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4399–4412, 2014.
- [3] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Machine learning-based fast intra mode decision for hevc screen content coding via decision trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1481–1496, 2020.
- [4] S.-H. Tsang, Y.-L. Chan, and W. Kuang, "Mode skipping for hevc screen content coding via random forest," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2433–2446, 2019.
- [5] Y. Xue, X. Wang, L. Zhu, Z. Pan, and S. Kwong, "Fast coding unit decision for intra screen content coding based on ensemble learning," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8543–8547.
- [6] F. Duanmu, Z. Ma, and Y. Wang, "Fast cu partition decision using machine learning for screen content compression," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4972–4976.
- [7] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Online-learning-based bayesian decision rule for fast intra mode and cu partitioning algorithm in hevc screen content coding," *IEEE Transactions on Image Processing*, vol. 29, pp. 170–185, 2020.
- [8] W. Kuang, Y.-L. Chan, S.-H. Tsang, and W.-C. Siu, "Deepsc: Deep learning-based fast prediction network for screen content coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1917–1932, 2020.
- [9] H. Yang, L. Shen, and P. An, "An efficient intra coding algorithm based on statistical learning for screen content coding," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 2468–2472.
- [10] F. Duanmu, Z. Ma, and Y. Wang, "Fast mode and partition decision using machine learning for intra-frame coding in hevc screen content coding extension," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 517–531, 2016.
- [11] C. Huang, Z. Peng, F. Chen, Q. Jiang, G. Jiang, and Q. Hu, "Efficient cu and pu decision based on neural network and gray level co-occurrence matrix for intra prediction of screen content coding," *IEEE Access*, vol. 6, pp. 46 643–46 655, 2018.
- [12] S.-H. Tsang, N.-W. Kwong, and Y.-L. Chan, "Fastscnet: Fast mode decision in vvc screen content coding via fully convolutional network," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, pp. 177–180.
- [13] Y.-H. Chao, Y.-C. Sun, J. Xu, and X. Xu, *JVET common test conditions and software reference configurations for non-4:2:0 colour formats*, Doc. JVET-R2013-v1, Apr. 2020.
- [14] *VVC Test Model Version 17.0 VTM-17.0*, [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM.
- [15] G. Bjontegaard, *Calculation of Average PSNR Differences Between RDCurves*, document VCEG-M33, VCEG, Austin, TX, USA, Mar. 2001.