COMPLEXITY ANALYSIS OF VVC INTRA CODING

[1]Mário Saldanha, [2]Gustavo Sanchez, [3]César Marcon, [1]Luciano Agostini

[1]Federal University of Pelotas – Pelotas, Brazil
[2]IF Farroupilha – Alegrete, Brazil
[3]Pontifical Catholic University of Rio Grande do Sul – Porto Alegre, Brazil
{mrdfsaldanha, agostini}@inf.ufpel.edu.br, gustavo.sanchez@iffarroupilha.edu.br, cesar.marcon@pucrs.br

ABSTRACT

Versatile Video Coding (VVC) is the next-generation of video coding standards, which was developed to double the coding efficiency over its predecessor High-Efficiency Video Coding (HEVC). Several new coding tools have been investigated and adopted in the VVC Test Model (VTM), whose current version can improve the intra coding efficiency by 24% at the cost of a much higher coding complexity than the HEVC Test Model (HM). Thus, this paper provides a detailed VVC intra coding complexity analysis, which can support upcoming works for finding the most time-consuming tool that could be simplified to achieve a real-time encoder design.

Index Terms**—**Complexity Analysis, Intra Prediction, Versatile Video Coding

1. INTRODUCTION

Video coding techniques have achieved relevant advances in the last few years. These advances lead to the standardization of new video encoders as the most advanced High-Efficiency Video Coding (HEVC) [1] standard, which provides higher coding efficiency than the previous ones. However, the interest of the user experiences and the industry requirements keep growing, including emerging video content such as Ultra-High Definition (UHD), Screen Content Coding (SCC), High-Dynamic Range (HDR) and 360-degree videos. For these kinds of applications, the HEVC has limited efficiency and more advanced techniques are required.

Versatile Video Coding (VVC) [2] has been developed by the Joint Video Experts Team (JVET), in a collaboration of ISO Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG), to meet various requirements from the market associated with the next video applications. JVET designed VVC with a focus on specifying a video coding technology with compression efficiency much higher than the previous standard and being highly versatile for effective use in the emerging applications. For this purpose, a lot of novel techniques have been designed and adopted in the draft of VVC, which has significantly improved the coding efficiency compared to its predecessor.

These techniques include a flexible block partitioning structure called Quadtree with nested Multi-Type Tree (QTMT) [3] that allows more partitioning flexibility in the Coding Unit (CU) shapes. Besides the HEVC translational Motion Compensation (MC) [1], the Affine Motion Compensation (AMC) [4] allows mapping complex movements as scaling, rotation, shearing, and other irregular movements. Matrix-based Intra Prediction (MIP) [5] predicts the intra-coded blocks through matrix multiplications defined by neural networks. Also, several other encoding tools have been developed and contributed to improving the VVC coding efficiency, such as Wide Angular Intra Prediction (WAIP) [6], Multiple Transform Selection (MTS) [7], and Low-Frequency Non-Separable Transform (LFNST) [8]. However, the computational burden of the VVC encoder is significantly increased when adopting these new tools.

Some works have already proposed VVC encoder solutions for complexity reduction, such as [9]-[13]. Furthermore, the VVC Test Model (VTM) 7.0 [3][14] already applies some complexity reduction techniques in the encoding process. However, its computational complexity is still prohibitive for real-time applications.

Topiwala et al. [15] compared VVC, HEVC, AV1 [16], and Essential Video Coding (EVC) [17], considering the bitrate compression performance and subjective assessments. Tissier et al. [18] analyzed opportunities to reduce the complexity of the VVC intra-coder, considering the block partition, the intra prediction mode, and the selection of transformations. However, these works do not perform a detailed complexity study of each VVC intra coding module.

This paper presents a detailed complexity analysis of the VVC intra prediction tools, measuring their execution times. By identifying the complexity of each tool, this paper aims to contribute to outcoming works focusing on reducing the computational complexity of the VVC intra coding.

2. VVC INTRA PREDICTION BACKGROUND

As in HEVC, the encoding process of VVC divides the input frame into blocks, called Coding Tree Units (CTUs). In this case, VVC supports block sizes larger than HEVC, where each frame can be partitioned into CTUs with a maximum size of 128×128 luminance samples. Additionally, each CTU can be recursively partitioned into smaller blocks called CUs.

VVC adopts the QTMT structure that allows more flexibility in the CU partition shapes compared to the ones obtained in the HEVC partitions. In the QTMT structure, a CU can be partitioned through the quadtree (QT) and Multi-Type Tree (MTT) partitioning structures. The QT structure uses the same concept employed in HEVC, where the current CU is split into four smaller quadratic sub-CUs, while the MTT structure enables CUs with rectangular shapes through Binary Tree (BT) and Ternary Tree (TT).

The BT and TT partitioning, which can be performed in vertical and horizontal directions, split the current CU into two symmetric or three sub-CUs, respectively. The three sub-CUs are composed of a central one and two sides of sub-CUs, having ½ and ¼ of the original CU size, respectively.

QT with nested MTT structure allows more flexible block partition types to adapt the features of several texture patterns, which can improve the encoding efficiency. Thus, VVC removes the separation of CU, Prediction Unit (PU), and Transform Unit (TU) concepts used in HEVC, where each unit may have a different block size. Besides, for intra slices, VVC allows the partitioning process to be performed separately for the luminance and chrominance components.

VVC performs the intra prediction of luminance samples with rectangular CU sizes ranging from 4×4 up to 64×64 and introduces several innovations to improve the encoding efficiency further.

The HEVC angular intra prediction modes are extended from 33 to 65 for representing a more variety of texture patterns and achieve higher accuracy. Besides, VVC allows the intra prediction of rectangular blocks due to the CU shapes obtained with the QTMT partitioning structure. In this case, the VVC intra prediction applies WAIP since the angle of the 33-conventional angular modes were defined only for square blocks. Thus, intra prediction modes with higher amplitude angular than the conventional angular modes are allowed for rectangular CUs.

Multiple Reference Line (MRL) [19] enables the use of more reference lines than the ones used in the HEVC intra prediction since, in some cases, the adjacent reference samples (located at line 0) can significantly differ from the predicting block leading to a meaningful prediction error. Consequently, the VVC intra coding also evaluates the block prediction using reference samples farther than the samples at line 0 (i.e., samples located at lines 1 and 2).

MIP performs the intra prediction through matrix multiplication and samples interpolation, where a set of matrices are evaluated according to the block size, and each matrix represents a prediction mode. The weights of these matrices were defined by neural networks, which were trained with a broad set of data. MIP improves the encoding efficiency allowing predictions that vary in more than one direction (i.e., nonlinear prediction), which is not possible with conventional angular modes.

Intra Subpartition (ISP) [20] improves the encoding efficiency by exploring the correlations among the intra-block samples. For this, ISP divides the current block vertically or horizontally into subpartitions that are sequentially encoded using the same intra prediction mode. Once a subpartition has been encoded, its reconstructed samples are used as reference samples for the next partition. Thus, more correlated reference samples are used for each subpartition compared to the conventional approach, which locates the reference samples at the left and above boundaries of the current block.

Fig. 1 presents the dataflow model for the VVC intra prediction. The encoding process evaluates several possibilities of encoding modes to find the mode with the lowest Rate-Distortion (RD) cost. Thus, the encoding process selects the prediction mode that reaches better visual quality and requires fewer bits to represent the predicted block.

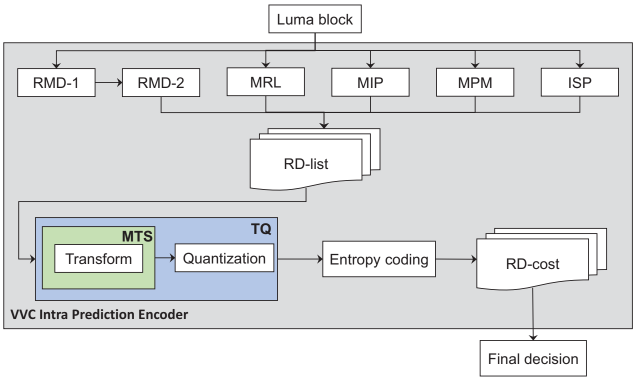


Fig. 1. Dataflow for VVC intra prediction.

Like HEVC, the VVC intra prediction employs Rough Mode Decision (RMD) and Most Probable Modes (MPM). VVC divides RMD into two steps (RMD-1 and RMD-2) to avoid an exhaustive evaluation of the 67 intra prediction modes. RMD-1 evaluates planar, DC, and 33 angular modes inherited from HEVC, through Sum of Absolute Transformed Differences (SATD) between the original block and the intra-predicted block, and inserts a few modes with the lowest SATDs ordered into the RD-list. RMD-2 performs a refinement step to evaluate the SATDs of angular modes that are adjacent to the angular modes already included in the RD-list. Thus, a reduced set of the new VVC angular modes are evaluated to try finding a more accurate prediction. MPM gets the default modes (the most used ones), and the modes in the neighbor blocks left and above encoded. The MPM list is composed of six candidates.

The residual coding encompasses Transform and Quantization (TQ) steps. In addition to DCT-II, which has been employed in HEVC, VVC also includes MTS to improve the residual coding efficiency by using multiple selected transforms. MTS introduces primary transform matrices DST-VII and DCT-VIII that can be applied in the horizontal and vertical directions. Besides, LFNST is evaluated as a set of secondary transforms to reduce the correlations of transform coefficients [8]. Finally, the best combination of primary and secondary transform is chosen through the lowest RD-cost.

3. EXPERIMENTS AND ANALYSIS

Our first experiment compares the execution time increase of the latest VVC reference software (VTM-7.0) in comparison to the latest HEVC reference software (HEVC Test Model - HM) 16.20 [21] when running in all-intra encoder configuration. Accordingly, we encoded 40 frames of five video sequences with video resolutions ranging from 240p to 4K according to the Common Test Conditions (CTC) [22], using 22, 27, 32, and 37 as Quantization Parameters (QPs), and temporal subsample ratio factor of 8.

Fig. 2 shows the encoding time increase rate of VTM for each video sequence and QP value. One can notice that QP 22 provides the highest encoding time increase rate, where VTM is 56 times slower than HM, on average. In the best scenario, reached with QP 37, VTM takes 21 times more than the HM encoding time for *ParkRunning3* video sequence.

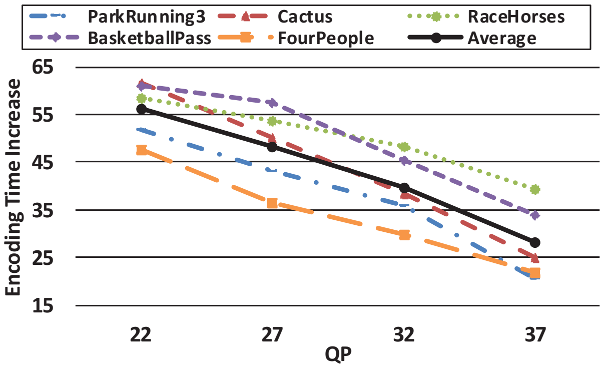
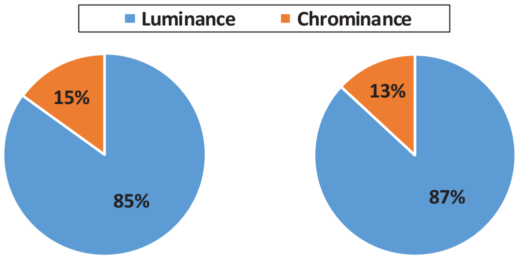


Fig. 2. Encoding time increase rate of VTM-7.0 in comparison to the HM-16.20.

This analysis shows that the execution time of VTM reference software is expressively higher than HM, considering all scenarios evaluated. Moreover, detailed information about the complexity of each intra coding tool and block sizes are essential when designing new complexity reduction solutions to achieve better results.

The next evaluations consider the same configuration of the previous analysis, including CTC encoder parameters and QPs; however, we increased the number of encoded frames to 160 and encoded all CTC video sequences. Fig. 3 presents the encoding time distribution of luminance and chrominance components for the corner cases of QP (22 and 37).

This analysis is interesting in the VVC intra prediction since the encoder can partition both components separately and this enables us to verify the impact of each channel in the total encoding time. Nevertheless, slight variations in the complexity distribution are noticed when QP varies. The luminance encoding process represents the highest encoding complexity of up to 87% of the total encoding time when considering QP 37, and 85% of the total encoding complexity for QP 22. One can conclude that solutions targeting the luminance channel tend to achieve more impressive complexity reduction results for the VVC encoder.



(a) (b)

Fig. 3. Encoding time distribution of luminance and chrominance for (a) QP 22 and (b) QP 37.

Table I shows the encoding time reduction results when removing the BT and TT structures of the partitioning of luminance and chrominance components. This table allows us to assess the impact of the QTMT partitioning structure. On average, when the VTM encoder disables the BT partitioning, the execution time is reduced by 75%, whereas disabling the TT partitioning reduces about 48% of the encoding time. Besides, when both BT and TT partitioning are disabled (i.e., only QT partitions are available), the encoding time is decreased by about 92%, on average.

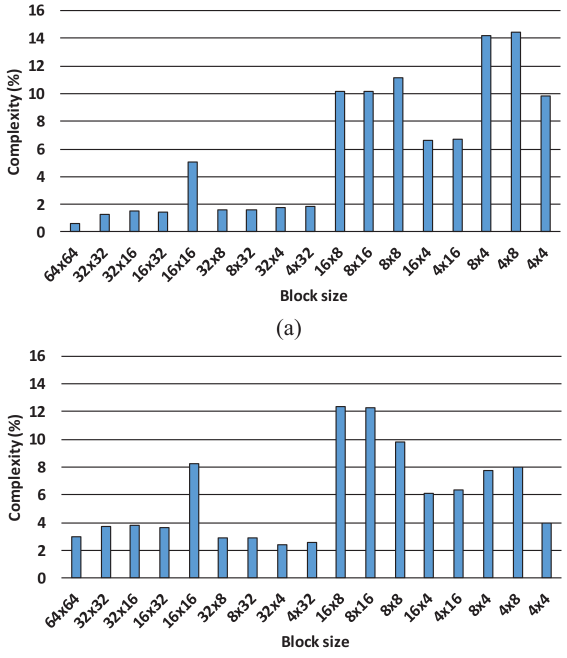
Table I. Encoding time reduction when removing BT or TT.

Class	Resolution	Encoding Time Reduction		
		No BT	No TT	No BT+TT
A1	3840×2160	67.9%	39.9%	85.5%
A2	3840×2160	77.1%	48.1%	92.7%
B	1920×1080	75.7%	47.5%	92.8%
C	832×480	79.3%	51.9%	95.1%
D	416×240	77.5%	51.2%	92.9%
E	1280×720	74.4%	47.2%	91.1%
Average		**75.3%**	**47.6%**	**91.7%**

This analysis demonstrates that the quadtree nested with MTT partitioning structure increases the total encoding time of the VTM significantly. Then, solutions focusing on reducing the evaluations of the BT/TT partitioning can carry remarkable results for the encoding complexity reduction, enabling the real-time encoder design. However, these solutions should be able to efficiently predict unnecessary evaluations of BT/TT to maintain the coding efficiency.

Since the luminance channel produces the highest complexity of the VVC encoder, the next evaluations are considering the block sizes and prediction mode complexity of this channel. Fig. 4 displays the average encoding complexity spent during luminance coding on each block size for the QP corner cases, i.e., QP 22 in Fig. 4(a) and QP 37 in Fig. 4(b). The x-axis represents each block size, which is sorted from the largest block area to the smallest one, ranging from 64×64 to 4×4.

The QP variation produces a different complexity distribution. Lower QPs have a complexity distribution concentrated in the block sizes with smaller areas, whereas higher QPs have a more heterogeneous complexity distribution. It occurs because QP indicates the compression rate and affects the image quality directly. Therefore, lower QPs tend to evaluate smaller blocks most frequently to maintain the image details, whereas higher QPs can finish early the evaluation of smaller block sizes to achieve more compression rate. The most complex block area for QP 22 is 32 (i.e., block sizes 4×8 and 8×4) and for QP 37 is 128 (i.e., block sizes 16×8 and 8×16).



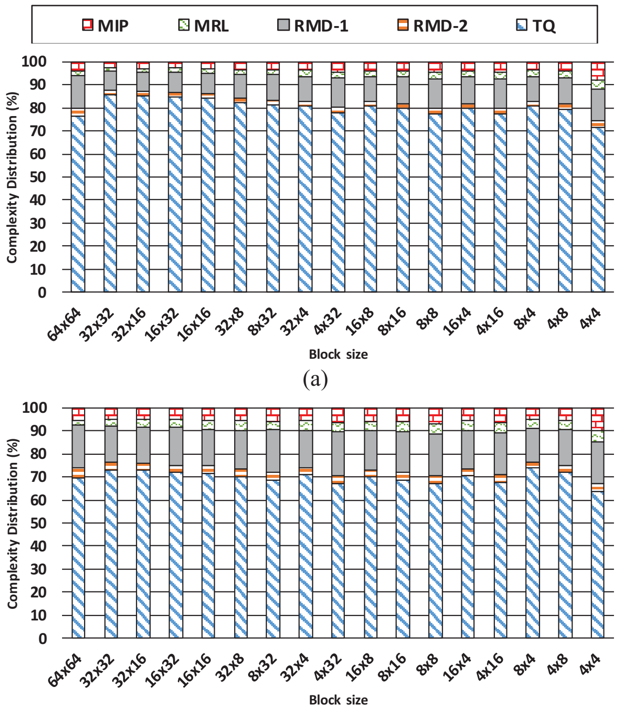
(a)

(b)

Fig. 4. Encoding complexity of each luminance block size for (a) QP 22 and (b) QP 37.

Fig. 5 displays the average complexity distribution of the encoding intra prediction steps according to the block size and the QP corner cases. This analysis considers RMD-1, RMD-2, MRL, and MIP as prediction tools and TQ as the residual coding flow regarding transform and quantization steps. The entropy coding time is included in TQ flow. Since ISP and MPM prediction tools have a negligible processing time, they were not considered in this analysis.

Firstly, one can notice that the complexity distribution variates slightly according to the block sizes. Also, for both corner QPs, the TQ evaluation (residual coding with transform and quantization) is the most time-demanding for all block sizes. On average, the TQ flow represents 80% and 70% of the encoding complexity distribution for QPs 22 and 37, respectively. This behavior demonstrates that MTS and LFNST evaluations increased the encoding complexity of residual coding significantly. Considering the prediction tools, RMD-1 is the task that presented the highest percentage of processing time (11% for QP 22 and 17% for QP 37, on average). Then, simplifying the RMD-1 search and reducing the number of modes inserted in the RD-list to be evaluated in the TQ flow should produce meaningful complexity reduction results.



(a)

(b)

Fig. 5. Encoding complexity distribution according to the block size for (a) QP 22 and (b) QP 37.

5. CONCLUSIONS

This work presented a VVC intra coding prediction complexity analysis, which can guide upcoming works focusing on complexity reduction. Several new tools have been developed and introduced in the VVC intra coding during the standardization, and this work demonstrated the complexity of those tools when running the VTM-7.0 under CTCs using the all-intra encoder configuration. This paper showed that the execution time of the VTM encoder is much higher than HM, and solutions for complexity reduction are required to enable a real-time design. The analyses demonstrated that complexity reduction solutions focusing on luminance channel, MTT partitioning structure, RMD-1, and residual coding could provide significant gains in terms of time-saving. Finally, to the best of our knowledge, this is the first work in the literature that provides a detailed complexity analysis considering the steps of the VVC intra coding.

ACKNOWLEDGMENT

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

6. REFERENCES

[1] G. Sullivan, J. Ohm, W. Han, T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *Transactions on Circuits and Systems for Video Technology (TCSVT)*, v. 22, n. 12, pp. 1649-1668, 2012. DOI: 10.1109/TCSVT.2012.2221191

[2] B. Bross, J. Chen, S. Liu, Y. Wang. “Versatile Video Coding (Draft 7),” JVET 16th Meeting, JVET-P2001, Geneva, Oct. 2019.

[3] J. Chen, Y. Ye, S. Kim. “Algorithm description for Versatile Video Coding and Test Model 7 (VTM 7),” JVET 16th Meeting, JVET-P2002, Geneva, Oct. 2019.

[4] K. Zhang, Y. Chen, L. Zhang, W. Chien, M. Karczewicz, “An Improved Framework of Affine Motion Compensation in Video Coding,” *IEEE Transactions on Image Processing (TIP)*, v. 28, n. 3, pp. 1456-1469, Mar. 2019. DOI: 10.1109/TIP.2018.2877355

[5] M. Schäfer et al., “An Affine-Linear Intra Prediction with Complexity Constraints,” *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 1089-1093. DOI: 10.1109/ICIP.2019.8803724

[6] L. Zhao et al., “Wide Angular Intra Prediction for Versatile Video Coding,” *Data Compression Conference (DCC)*, 2019, pp. 53-62. DOI: 10.1109/DCC.2019.00013

[7] X. Zhao et al. “Enhanced Multiple Transform for Video Coding,” *Data Compression Conference (DCC)*, 2016, pp. 73–82. DOI: 10.1109/DCC.2016.9

[8] M. Koo, M. Salehifar, J. Lim, S. Kim, “Low Frequency Non-Separable Transform (LFNST),” *Picture Coding Symposium (PCS)*, 2019, pp. 1-5. DOI: 10.1109/PCS48520.2019.8954507

[9] A. Wieckowski, J. Ma, H. Schwarz, D. Marpe, T. Wiegand, “Fast Partitioning Decision Strategies for The Upcoming Versatile Video Coding (VVC) Standard,” *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4130-4134. DOI: 10.1109/ICIP.2019.8803533

[10] Z. Zhang, X. Zhao, X. Li, Z. Li, S. Liu, “Fast Adaptive Multiple Transform for Versatile Video Coding,” *IEEE Data Compression Conference (DCC)*, 2019, pp. 63-72. DOI: 10.1109/DCC.2019.00014

[11] H. Yang et al., “Low Complexity CTU Partition Structure Decision and Fast Intra Mode Decision for Versatile Video Coding,” *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, early access, pp. 1-14, 2019. DOI: 10.1109/TCSVT.2019.2904198

[12] T. Fu, H. Zhang, F. Mu, H. Chen, “Fast CU Partitioning Algorithm for H.266/VVC Intra-Frame Coding,” *IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 55-60. DOI: 10.1109/ICME.2019.00018

[13] M. Lei, F. Luo, X. Zhang, S. Wang, S. Ma, “Look-Ahead Prediction Based Coding Unit Size Pruning for VVC Intra Coding,” *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4120-4124. DOI: 10.1109/ICIP.2019.8803421

[14] VTM-7.0 [online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-7.0

[15] P. Topiwala, M. Krishnan, W. Dai, “Performance comparison of VVC, AV1, and HEVC on 8-bit and 10-bit content,” *SPIE Applications of Digital Image Processing XLI*, v. 10752, pp. 305-314, 2018. DOI: 10.1117/12.2322024

[16] Y. Chen et al., “An Overview of Core Coding Tools in the AV1 Video Codec,” *Picture Coding Symposium (PCS)*, 2018, pp. 41-45. DOI: 10.1109/PCS.2018.8456249

[17] K. Choi et al., “MPEG-5: essential video coding standard,” *SPIE Applications of Digital Image Processing XLII*, v. 11137, pp. 250-258, 2019. DOI: 10.1117/12.2530429

[18] A. Tissier et al., “Complexity Reduction Opportunities in the Future VVC Intra Encoder,” *IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, 2019, pp. 1-6. DOI: 10.1109/MMSP.2019.8901754

[19] B. Bross et al. “CE3: Multiple reference line intra prediction (Test 1.1.1, 1.1.2, 1.1.3 and 1.1.4),” JVET 12th Meeting, JVET-L0283, Macao, Oct. 2018.

[20] S. Hernández, V. George, J. Ma, T. Nguyen, H. Schwarz, D. Marpe, T. Wiegand, “An Intra Subpartition Coding Mode for VVC,” In IEEE International Conference on Image Processing (ICIP), 2019, pp. 1203-1207. DOI: 10.1109/ICIP.2019.8803777

[21] HM-16.20 [online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20/

[22] F. Bossen, J. Boyce, X. Suehring, X. Li, V. Seregin, “JVET common test conditions and software reference configurations for SDR video,” JVET 14th Meeting, JVET-N1010, Mar. 2019.