Adaptive Block-Level Quality Parameter Adjustment Towards Low Video Bit-Rate Fluctuation

Longtao Feng[1], Qian Yin[1], Huiwen Ren[1], Zhao Wang[1,2*], Siwei Ma[1,2] and Yuwen He[3]

[1]National Engineering Research Center of Visual Technology,
School of Computer Science, Peking University, China

[2]Advanced Institute of Information Technology, Peking University, China

[3]Bytedance Inc., San Diego, CA, USA

{ltfeng,hwren,zhaowang,swma}@pku.edu.com, yinqian_xixi@stu.pku.edu.com, yuwen.he@bytedance.com

***Abstract*—Existing quantization parameter (QP) adjustment methods in video coding often focus solely on coding efficiency and ignore the impact of bit-rate fluctuations on video transmission and bandwidth waste. This is mainly because intra pictures, in a hierarchical coding structure, are allocated smaller QP and thus consume more bits. To address this issue, we propose an adaptive block-level QP adjustment method. Specifically, intra picture importance (IPI) is first introduced to evaluate the adjustability of intra picture QP. For intra pictures whose QP can be adjusted, we further propose block importance (BI) to determine their optimal block-level QP adjustment. Experimental results show that our proposed method reduce the bit-rate fluctuations while basically maintaining the coding performance. Notably, significant improvements can be observed in high-resolution videos, with a reduction of approximately 11% in bit-rate fluctuations.**

***Index Terms*—Video Coding, Quantization Parameter, Bit-Rate Fluctuation.**

I. Introduction

The rapid growth of diverse video applications has brought in sheer volume of video data, and poses huge challenges to limited bandwidth and storage space. To address this issue, the joint video experts team (JVET) has been committed to promoting the development of video coding standards. Starting from the advanced video coding (AVC) [1], the coding order and display order of pictures are completely decoupled, and any picture can be designated as a reference picture, which provides great flexibility for applying various prediction structures. It is worth mentioning that the hierarchical B pictures [2], [3] prediction structure in AVC significantly improve the coding efficiency. Therefore, subsequent coding standards, high efficiency video coding (HEVC) [4] and versatile video coding (VVC) [5], have also adopted and improved this hierarchical coding structure (HCS).

In recent years, extensive works [6]–[13] have been proposed to apply QP adjustment for the HCS. For example, a low-complexity adaptive QP offset selection method was proposed by relating QP offset to the type of texture content. [6]. Similarly, Zhao *et al.* [7] proposed an adaptive QP cascading (QPC) scheme using features of natural videos. In terms of the low delay (LD) configuration, Gong *et al.* [8] introduced an efficient QPC technique for surveillance video coding by using all inter reference pictures. Additionally, Tang *et al.* [9] proposed a robust hierarchical QP setting method for the screen content coding.

The above-mentioned QP adjustment methods have improved coding performance by making full use of HCS. However, they usually cause significant QP differences between pictures in lowest and highest temporal layers. As a result, the bit consumption of different pictures varies considerably, which is not conducive to video transmission and bandwidth utilization. To mitigate bit-rate fluctuations, Ren *et al.* [14] proposed to increase QP of intra pictures for AVS3 by analyzing intra and inter prediction cost [15]. However, this method only simply adjusts the picture-level QP and fails to consider the impact of different blocks on subsequent pictures. Moreover, it is only effective for high-resolution content.

In this paper, we further investigate QP adjustments aimed at reducing video bitrate fluctuations. We are the first to propose an adaptive block-level QP adjustment method in VVC for reducing bit-rate fluctuations without substantially affecting performance. Specifically, we introduce the intra picture importance (IPI) to denote the influence of the quality of intra pictures on subsequent pictures to be encoded. This influence is modeled using the motion information between these pictures, which determines whether their QPs can be increased or not. For intra pictures that can increase QPs, we further propose the block importance (BI) to represent the impact of the quality of blocks on subsequent pictures to be encoded. The BI is modeled using residuals after the motion compensation, and employed to determine the optimal block QP adjustments. Experimental results show that our method significantly reduces the bit-rate fluctuations without compromising the coding efficiency.

II. Proposed Method

Fig. 1 shows the overall framework of the proposed adaptive block-level QP adjustment for intra picture, which mainly consists of three stages: 1) downsampling the intra picture and four adjacent pictures twice to generate the corresponding

$*$ means corresponding author. This work was supported in part by the “Leading Goose” R&D Program of Zhejiang Province under grants 2024C01101, in part by the National Natural Science Foundation of China no.62031013.

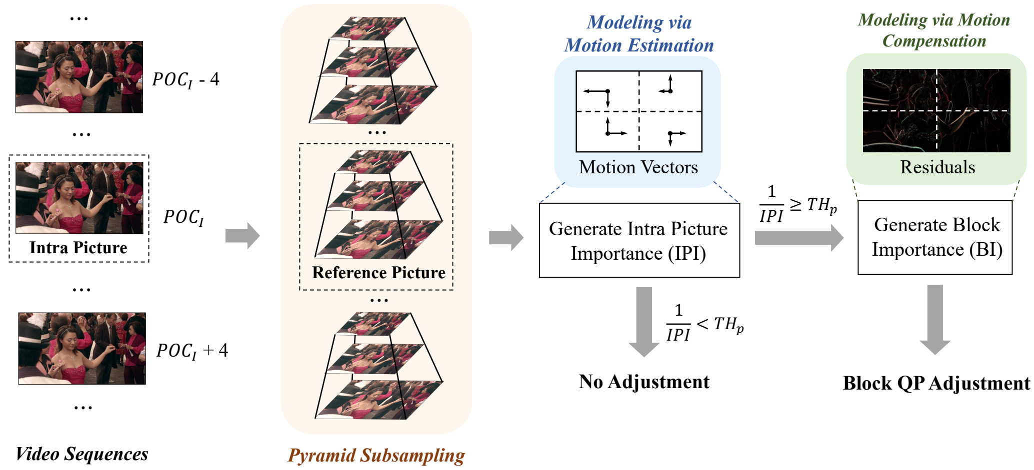

Fig. 1. The overall framework of the proposed adaptive block-level QP adjustment for intra pictures.

lower resolution pictures, 2) performing motion estimation on adjacent pictures using intra picture as a reference to obtain the intra picture importance (IPI), and 3) conducting motion compensation to calculate prediction residuals, and derive block importance (BI) for determining the final block QP adjustment. In what follows, the proposed IPI is described in Section II-A, which determines whether intra pictures can increase their QPs or not. Then, Section II-B presents our proposed BI, which serves as a guide for block-level QP adjustment.

A. Intra Picture Importance

Generally, when the pictures to be encoded highly resemble the reference pictures, the quality of the reference pictures can significantly affect the coding efficiency of the pictures to be encoded, and vice versa. To quantify this impact, we introduce the concept of intra picture importance (IPI) to measure how quality variations of the intra picture actually influence the subsequent pictures. Notably, a lower IPI indicates that the intra picture is less important to subsequent pictures. Accordingly, we can attempt to increase the QP of these intra pictures with lower IPI for the bit-rate fluctuation reduction. Consequently, accurate modeling of the IPI assumes critical significance in achieving optimal QP adjustments.

Based on the observations, we have noticed that as the motion between intra pictures and subsequent pictures increases, the correlation between them gradually weakens. In fact, decreasing the quality of intra pictures can actually enhance the coding efficiency in certain scenarios. This is because significant motion between pictures not only complicates the motion estimation but also leads to higher residuals after the motion compensation, which ultimately both hampers the coding efficiency. Additionally, inter pictures often have multiple reference pictures, among which reference pictures with larger motions tend to have the least impact due to the rate-distortion optimization (RDO) process. Therefore, we propose to model the IPI using the motion information between intra pictures and subsequent pictures to be encoded.

To obtain the motion information between intra pictures and subsequent pictures to be encoded accurately and efficiently, a hierarchical motion estimation scheme, similar to [16], is utilized to obtain the motion vector. Specifically, the intra pictures and adjacent pictures are all downsampled twice to generate pictures with two resolutions. Motion searches are initially performed at the lowest resolution. Then, the best motion vector (MV) obtained in last step is scaled and used as a starting point for a larger resolution. Additionally, the sub-pixel motion estimation is also performed at native resolution. It is worth noting that only intra pictures are used as references to conduct motion estimation on adjacent pictures. Additionally, taking the time complexity into account, only four pictures before and after the current intra picture are considered in our scheme. Finally, motion vectors are used to model the IPI as follows:

$$IPI = \frac{8}{\sum_{k \in \boldsymbol{K}} MV_{xy,poc_I+k}}, \tag{1}$$

$$\boldsymbol{K} = \{-4, -3, -2, -1,\ 1,\ 2,\ 3,\ 4\}, \tag{2}$$

$$MV_{xy,i} = \sum |MV_{x,i}| + \sum |MV_{y,i}|, \tag{3}$$

where $MV_{x,i}$ and $MV_{y,i}$ are motion vectors of horizontal and vertical direction in ith picture respectively, and poc_I is the picture order count (POC) of intra picture.

From equation (1), we can see that when adjacent pictures have a large motion relative to intra pictures, the impact of intra pictures on subsequent pictures becomes negligible, that is, the IPI is low. In this case, it is feasible to appropriately increase the QP of intra pictures for reducing bit-rate fluctuations while maintaining the coding performance. Therefore, we define an adjustment start threshold TH_p to determine whether the QP of intra picture needs to be adjusted. Moreover, our observations have revealed a positive correlation between TH_p and picture resolution, defined as follows:

TABLE I
PERFORMANCE OF PROPOSED METHOD

Class	Seq	Adjustable	BD-Rate %	ΔD_{pv} %	$\Delta\sigma_{bits}$ %	$\Delta\sigma_{bitrate}$ %
A1	Tango2	✓	-0.185	7.589	3.378	7.180
	FoodMarket4	✓	0.313	4.004	3.250	2.261
	Campfire	✓	1.976	19.901	14.354	8.519
A2	CatRobot1	×	/	/	/	/
	DaylightRoad2	✓	1.189	10.218	7.359	10.189
	ParkRunning3	✓	-0.407	14.861	5.802	12.404
B	MarketPlace	✓	0.176	0.0	0.705	0.195
	RitualDance	✓	-0.156	3.390	3.035	1.757
	Cactus	×	/	/	/	/
	BasketballDrive	✓	-0.050	0.088	2.391	2.087
	BQTerrace	✓	0.277	0.010	0.841	0.467
C	RaceHorsesC	✓	-0.019	10.145	6.384	6.458
	BQMall	✓	-0.049	0.009	1.916	1.542
	PartyScene	×	/	/	/	/
	BasketballDrill	✓	0.294	3.433	1.913	2.785
D	RaceHorses	✓	-0.184	8.406	6.138	8.422
	BQSquare	×	/	/	/	/
	BlowingBubbles	×	/	/	/	/
	BasketballPass	✓	0.135	0.000	1.262	0.592
Average on adjustable			0.236	5.861	4.195	4.633
Average on 4K resolution			0.577	11.315	6.828	8.111

$$TH_p = \alpha_{THp} \cdot (W \times H)^{\beta_{THp}}, \tag{4}$$

where W and H represents the width and height of videos respectively, which denotes the video resolutions. α_{THp} and β_{THp} are empirical parameters.

B. Block Importance

The IPI mentioned above globally determines whether the QP of the intra picture can be adjusted at the picture level. If an intra picture is eligible for QP adjustment, we then perform block-level QP adjustments. It is worth mentioning that the video coding usually conducts a block-by-block prediction. Even within the same picture, different blocks are always referenced with different frequencies. Furthermore, blocks that are not often used as references usually have less impact on subsequent pictures to be encoded. Therefore, by appropriately adjusting the QP for this type of block, we can not only reduce the bit-rate fluctuations but also minimize the impact on coding efficiency. Based on these analysis, we propose the block importance (BI) to measure the impact of the quality of blocks in intra-picture on coding efficiency. Similar to modeling the IPI, we only use a total of eight pictures (four consecutive pictures before and after) for motion compensation and calculate the residuals to model BI. The residuals can be obtained after the motion estimation introduced in Section II-A and subsequent motion compensation. Then, the BI is calculated as shown below:

$$BI = \frac{8}{\sum_i^{\boldsymbol{K}} BSSD_i}, \tag{5}$$

$$BSSD_i = \frac{\sum_{j \in \boldsymbol{S}} \sum |R_{i,j}|}{N_{\boldsymbol{S}}}, \tag{6}$$

where $R_{i,j}$ represents the residuals of the jth block in the ith picture that selects the current block as a reference. $\boldsymbol{S}$ represents block index set in the ith picture and $N_{\boldsymbol{S}}$ represents the number of blocks in $\boldsymbol{S}$. It can be seen from equation (5) and (6) that the smaller the BI, the lower the importance of the current block. Increasing its QP will not affect the encoding efficiency of subsequent picture to be encoded.

C. Adaptive Block-Level QP Adjustment

According to the IPI and the BI described in Section II-A and Section II-B respectively, we propose an adaptive block-level QP adjustment scheme and employ a two-stage procedure as shown in Fig. 1. Firstly, we utilize the IPI to decide whether the QP of the intra picture needs to be increased, which is formulated as follows:

$$QP_{adjust} = \mathcal{F}(IPI, BI) = \begin{cases} 0, & \dfrac{1}{IPI} < TH_p \\ \mathcal{G}(BI), & \dfrac{1}{IPI} \geq TH_p \end{cases} \tag{7}$$

where $\mathcal{F}(\cdot)$ represents the function that calculates the adjusted block-level QP, denoted as QP_{adjust}. When the $\frac{1}{IPI}$ is smaller than the picturel-level threshold TH_p, QP_{adjust} is set to 0, indicating there is no QP adjustment on current intra picture. Otherwise QP_{adjust} is then determined by $\mathcal{G}(BI)$, which is defined according to:

$$\mathcal{G}(BI) = \begin{cases} 0, & \dfrac{1}{BI} < TH_{b0} \\ \alpha_b \cdot \dfrac{1}{BI} + \beta_b, & TH_{b0} \leq \dfrac{1}{BI} \leq TH_{b1} \\ QP_{max}, & \dfrac{1}{BI} > TH_{b1} \end{cases} \tag{8}$$

$$\alpha_b = \frac{QP_{max}}{TH_{b1} - TH_{b0}}, \tag{9}$$

$$\beta_b = \frac{TH_{b0} \cdot QP_{max}}{TH_{b0} - TH_{b1}}, \tag{10}$$

where $\mathcal{G}(BI)$ is a piece-wise function that determined by two block-level thresholds, TH_{b0} and TH_{b1} (set to 2000 and 40000 respectively). When $\frac{1}{BI}$ is smaller than TH_{b0}, there is no QP adjustment. When $\frac{1}{BI}$ is lager than TH_{b1}, the QP adjustment maintains a constant value denoted as QP_{max}. In other cases, QP_{adjust} is positively related to the $\frac{1}{BI}$, and α_b and β_b are determined by TH_{b0}, TH_{b1} and QP_{max}.

Additionally, we also found that higher resolution pictures usually tend to have less content covered by the same-sized picture block. Therefore, for higher resolution videos, the QP_{max} can be set to larger than lower resolution ones. Based on our observations, QP_{max} exhibits a positive correlation with picture resolution, as shown bellows:

$$QP_{max} = \alpha_{QP} \cdot (W \times H)^{\beta_{QP}} + \gamma_{QP} \tag{11}$$

where α_{QP}, β_{QP} and γ_{QP} are empirical parameters.

TABLE II

RESULTS OF BIT-RATE FLUCTUATION AND CODING EFFICIENCY ON *ParkRunning3* AND *Tango2*

Sequence	QP	BD-Rate	D_{pv}		ΔD_{pv}	σ_{bits}		$\Delta\sigma_{bits}$	$\sigma_{bitrate}$		$\Delta\sigma_{bitrate}$
		%	Anchor	Ours	%	Anchor	Ours	%	Anchor	Ours	%
ParkRunning3	22	-0.41	9411736	7803512	17.09	1724743.43	1628920.66	5.56	6257.87	5701.61	8.89
	27		5604368	4722528	15.73	995702.65	938772.42	5.72	3434.92	2992.09	12.89
	32		3180928	2684232	15.61	520377.40	483552.86	7.08	1976.46	1653.12	16.36
	37		1619736	1441432	11.01	268942.64	255881.63	4.86	1046.01	925.99	11.47
Tango2	22	-0.19	2235216	1920112	14.10	319784.23	303329.16	5.15	2113.65	1897.87	10.21
	27		1034672	975104	5.76	159567.86	155021.23	2.85	803.16	737.07	8.23
	32		622200	588240	5.46	91936.45	89378.07	2.78	427.59	392.81	8.13
	37		378064	358992	5.04	54506.62	53016.15	2.73	239.24	220.26	7.93

III. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed method, we implement it and conduct experiments based on VVC Test Model VTM-22.0 [17], under the random access (RA) configurations (GOP size = 16) of the common test conditions (CTCs) [18]. For the evaluation of coding efficiency, the commonly-used Bjøntegaard delta rate (BD-Rate) is used. In our experiments, the BD-Rate is obtained based on Y-PSNR for calculation. In terms of the measurement of the video bit-rate fluctuation, the bit peak-to-valley difference D_{pv} and the reduction of bit peak-to-valley difference ΔD_{pv} used in [14] are employed. Moreover, we have designed two metrics to further evaluate. The first one is the reduction of the standard deviation of video bits $\Delta\sigma_{bits}$, which is formulated as follows:

$$\Delta\sigma_{bits} = \frac{(\sigma_{bits}^{anc} - \sigma_{bits}^{pro})}{\sigma_{bits}^{anc}} \times 100\%, \tag{12}$$

where σ_{bits}^{anc} and σ_{bits}^{anc} are the standard deviations of the video bits across all pictures in the video coded by the anchor and proposed method, respectively. The other one is the reduction of the standard deviation of the instantaneous bit-rate $\Delta\sigma_{bitrate}$, defined as follows,

$$\Delta\sigma_{bitrate} = \frac{(\sigma_{bitrate}^{anc} - \sigma_{bitrate}^{pro})}{\sigma_{bitrate}^{anc}} \times 100\%, \tag{13}$$

where $\sigma_{bitrate}^{anc}$ and $\sigma_{bitrate}^{pro}$ represent the standard deviation of all the instantaneous bit-rate $bitrate$ in the video coded by anchor and proposed method, respectively, and $bitrate$ is defined as follows,

$$bitrate = \frac{\sum_{i=1}^{N_t} Bit_i}{N_t} \times FPS, \tag{14}$$

where N_t is the number of pictures within a time period of t. Bit_i represents bit consumption of ith picture. FPS is video frame-rate. In this paper, $bitrate$ is calculated every 0.5 seconds. Larger value of D_{pv} indicates larger fluctuation and σ_{bits} and $\sigma_{bitrate}$ are similar. Larger values of ΔD_{pv} indicate more reduced fluctuations and $\Delta\sigma_{bits}$ and $\Delta\sigma_{bitrate}$ are similar.

Table I shows the performance results of our method in comparison with Anchor, where the "*Adjustable*" indicating whether the current video has at least one intra picture that QPs can be adjusted. The results reveal that our proposed method effectively reduces bit-rate fluctuations without impacting coding efficiency. Notably, better results can be achieved in 4K resolution. As high-resolution videos typically have higher bit-rates, their bit-rate fluctuations can have a greater impact on video transmission. Therefore, reducing fluctuations of high resolution videos is more valuable for practical applications.

To further demonstrate the performance of our method, Table II shows results of bit-rate fluctuation and coding efficiency for sequences *ParkRunning3* and *Tango2*. The results further substantiate the effectiveness of our method. The reductions of the peak-to-valley difference of bits are more than 14.10% when QP is 22. Additionally, there has been a small improvement in the coding performance. The reason for improvement in reducing fluctuation and performance is that two sequences exhibit significant motion between their pictures .

IV. CONCLUSION

In this paper, we propose an adaptive block-level QP adjustment method aimed at reducing the bit-rate fluctuation. Specifically, IPI based on the motion information is introduced to represent the impact of the quality of intra pictures on subsequent pictures. According to IPI, it can be determined whether to increase the QP of intra pictures. For intra pictures that can increase QP, we further introduce BI to indicate the impact of the quality of picture blocks on subsequent pictures. BI is modeled by using the residuals from motion compensation and determines the optimal adjustment values for block QP. Experimental results illustrate that our method significantly reduces the video bit-rate fluctuations without impacting the coding efficiency. Moreover, our method demonstrates better performance in high-resolution videos, which is valuable for practical applications because this type of videos typically have higher bit-rates and bit-rate fluctuations.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) document JVT-P014*, Berlin, Jul. 2005.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2006, pp. 1929–1932.

[4] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[5] B. Benjamin, Y. K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J. R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, 2021.

[6] M. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "An adaptive QP offset determination method for HEVC," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2016-August, 2016, pp. 4220–4224.

[7] T. Zhao, Z. Wang, and C. Chen, "Adaptive Quantization Parameter Cascading in HEVC Hierarchical Coding," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2997–3009, 2016.

[8] Y. Gong, K. Yang, Y. Liu, K. P. Lim, N. Ling, and H. Wu, "Quantization Parameter Cascading for Surveillance Video Coding Considering All Inter Reference Frames," *IEEE Trans. Image Process.*, vol. 30, pp. 5692–5707, 2021.

[9] T. Tang, L. Li, and D. Wang, "A Robust Hierarchical QP Setting for Screen Content Coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 2018-April, 2018, pp. 1777–1781.

[10] F. Raufmehr, M. R. Salehi, and E. Abiri, "A frame-level MLP-based bit-rate controller for real-time video transmission using VVC standard," *J. Real-Time Image Process.*, vol. 18, pp. 751–763, 2021.

[11] C. Liu, H. Sun, J. Katto, X. Zeng, and Y. Fan, "QA-Filter: A QP-adaptive convolutional neural network filter for video coding," *IEEE Trans. Image Process.*, vol. 31, pp. 3032–3045, 2022.

[12] M. H. Hyun, B. Lee, and M. Kim, "A VVC Intra Rate Control with Small Bit Fluctuations Using a Lagrange Multiplier Adjustment," *IEEE Trans. Multimedia*, 2024.

[13] J. Liao, L. Li, D. Liu, and H. Li, "Reinforcement Learning-based Frame-level Bit Allocation for VVC," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.* IEEE, 2023, pp. 1–5.

[14] H. Ren, S. Wang, S. Ma, and W. Gao, "An Adaptive Intra-frame Quantization Parameter Derivation Model Jointing with Inter-frame Analysis," in *Proc. Data Compress. Conf.*, vol. 2023-March, 2023, pp. 101–109.

[15] J. Zhang, C. Jia, M. Lei, S. Wang, S. Ma, and W. Gao, "Recent development of AVS video coding standard: AVS3," in *Proc. Picture Coding Symp.*, Ningbo, China, 2019, pp. 1–5.

[16] P. Wennersten, "Encoder-only GOP-based temporal filter," in *Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 document JVET-O0549*, Jul. 2019.

[17] "VVC software VTM-22.0," [Online], https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-22.0/ Accessed February 4, 2024.

[18] B. Frank, "VTM common test conditions and software reference configurations for SDR video," in *Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 document JVET-T2010*, Oct. 2020.