Revisiting Local Texture Correlation for Rate-distortion Optimized Intra Coding

Meng Wang*, Junru Li†, Li Zhang®, Hongbin Liu‡, Jizheng Xu® and Shiqi Wang*
*Department of Computer Science, City University of Hong Kong, Hong Kong, China.
†Institute of Digital Media, Peking University, Beijing, China.
®Bytedance Inc., San Diego CA. 92122 USA.
‡Bytedance (HK) Limited., Hong Kong, China.

Abstract

In this paper, we focus on computationally modeling of the local texture correlations, in an effort to better explore the coding modes with higher priorities in the rate-distortion optimized intra coding. In particular, strong correlations and continuities of local texture with neighboring blocks have been revealed in our analysis, and empirical justifications provide us inspirations on the joint optimization of rate-distortion-complexity when angular modes become finer to adapt the local textures. We examine the philosophy with extensive experiments conducted for refining the intra full-RD list. The results show that better coding performance with on average 0.72% and 3.00% BD-Rate savings for the natural scene and screen content sequences can be achieved in AVS3 test model HPM-5.0 under all intra configuration, with negligible encoding and decoding time variations.

1. Introduction

Intra prediction, which takes advantage of the spatial homogeneity to eliminate the redundancies in a single frame, has been a long-standing problem in video coding. The recent development of intra prediction reveals that increasing the number of mode candidates is beneficial for promoting the prediction accuracy. The number of intra prediction modes has been largely augmented from H.264/AVC [1] (9 intra modes) to High Efficiency Video Coding (HEVC) [2] standard (35 intra modes). In Versatile Video Coding standard [3] and the third generation of audio video coding standard (AVS3), in total 67 intra modes are designed [4]. Such fine-grained prediction scheme is capable of effectively capturing the arbitrary texture directions in natural videos, as illustrated in Fig. 1.

Regarding the determination of the optimal intra mode, rate distortion optimization (RDO) [5] is employed to choose the one with the minimum rate-distortion (RD) cost. However, it is unlikely to traverse all the available intra modes with full RD procedure, due to the involvement of considerable encoding complexity. The mainstream video codecs wisely employ rough mode decision (RMD) method [6]. In this way, the candidate intra modes are evaluated in the first round with the optimization of the Hadamard distortion plus rate,

$$J_{HAD}^{(i)} = D_{HAD}^{(i)} + \lambda \cdot \hat{R^{(i)}}, \tag{1}$$

where i represents the mode index and $D_{HAD}^{(i)}$ is the Hadamard distortion. $\hat{R^{(i)}}$ denotes the estimated bits for coding mode index. Subsequently, X candidates with

2375-0359/20/$31.00 ©2020 IEEE
DOI 10.1109/DCC47342.2020.00035

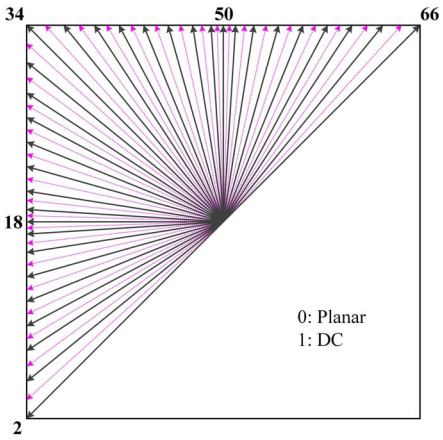

Figure 1: Illustration of the intra prediction mode in VVC [3].

smaller $J_{HAD}^{(i)}$ are selected as full-RD candidates, where X is predefined according to coding block sizes. As such, RDO is only conducted with the modes in the full-RD list. Moreover, in VTM and HM platform, the most probable modes (MPM) [2] are selectively merged in the full-RD list.

The RMD typically relies on the local Hadamard cost, while the theoretical and empirical analysis of texture-dependent correlations, which have been shown to be highly relevant with the operational rate distortion functions in intra coding [7], has been largely ignored in this context. In this paper, we first comprehensively probe the characteristics of natural and screen content sequences in terms of the local similarities and correlations. The continuities of local textures and the strong correlations of neighboring blocks act as useful guidance for determining the priorities of intra modes. Subsequently, the limitations of Hadamard cost are further analyzed. These analyses motivate us to look for an optimization strategy for the intra full-RD list to combat the existing limitations, and extensive experiments show that the philosophy in involving the texture-dependent correlations results in significant rate-distortion performance improvement in the AVS3 HPM-5.0 platform [8].

2. Texture-Dependent Correlations and Local Texture Directions

Image correlation model was studied in early in the 1970s [9] which assumed that a pixel is a random variable with zero mean and a constant variance following Gaussian distribution. The correlation between two pixels $I_{x,y}$ and $I_{x',y'}$ and can be described as follows,

$$\rho(I_{x,y}, I_{x',y'}) = e^{-\mu|x-x'|-v|y-y'|}, \tag{2}$$

where x and y denote the coordinates of pixels in the horizontal and vertical directions, respectively. Different μ and v lead to the correlations along different directions. Given the fact that video coding operates at block-level, Pearson Correlation Coefficient is adopted to evaluate the inter-block correlation. Assuming the current block

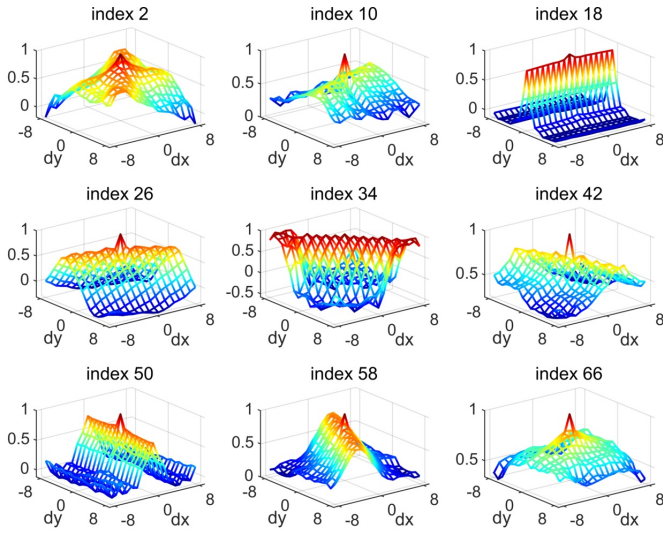

Figure 2: Illustration of the average $\hat{\rho}(B, \hat{B})$ in *BasketballDrive* under certain texture directions, and the x-axis and y-axis denote the position offset between two blocks.

is B and the neighboring block is $\hat{B}$, the correlation between B and $\hat{B}$ can be derived as follows,

$$\hat{\rho}(B,\hat{B}) = \frac{\sum(B_{x,y}\cdot\hat{B}_{x,y}) - \frac{\sum B_{x,y}\cdot\sum\hat{B}_{x,y}}{S}}{\sqrt{[\sum B_{x,y}^2 - \frac{(\sum B_{x,y})^2}{S}][\sum \hat{B}_{x,y}^2 - \frac{(\sum \hat{B}_{x,y})^2}{S}] + \epsilon}}, \tag{3}$$

where the block size is set as 8×8 such that the block area S is 64. (x, y) denotes the pixel coordinates in the current and neighboring blocks. In addition, the position offset $(\Delta X, \Delta Y)$ between B and $\hat{B}$ is limited within the range of [-8, +8]. ϵ is a smoothing factor which prevents the denominator from zero during the simulations.

Herein, the local textures [7] with respect to the 65 directional intra prediction modes are investigated. The projection and interpolation methods developed in VVC are re-implemented for exploring the relationship between intra prediction and spatial correlations. To be more specifically, the intra modes corresponding to different directions are traversed, and the conventional angular prediction is conducted for each 8×8 luma block, where neighboring samples (top 16, left 16 and top-left one) are involved for obtaining the prediction block. The prediction mode M_{opt} is decided according to the mean square error (MSE) between the original block and the predicted one.

Subsequently, we investigate the block correlations between the current and neighboring blocks under certain texture direction M_{opt}. More specifically, $\hat{\rho}(B, \hat{B})$ can be derived given position offset $(\Delta X, \Delta Y)$, where $\hat{B}$ is obtained by exhaustively traversing. The averaged block correlation with different prediction mode is illustrated in Fig. 2 and Fig. 3 for natural scene and screen content videos. For better visualization, only nine intra modes with distinct directions are depicted. The ridges of correla-

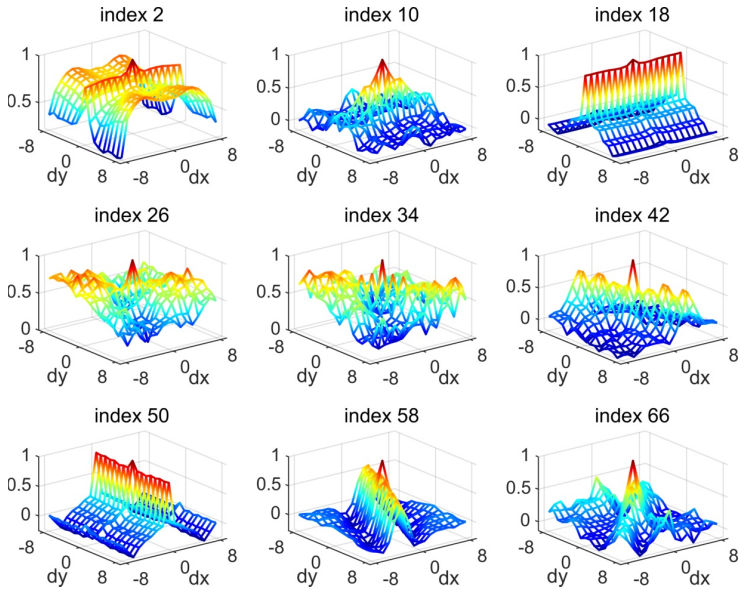

Figure 3: Illustration of the average $\hat{\rho}(B, \hat{B})$ in *ChineseEditing* under certain texture directions, and the x-axis and y-axis denote the position offset between two blocks.

tion mesh along the mode directions can be observed in Fig. 2 and Fig. 3, indicating that blocks locating along the texture directions share much stronger correlations. Moreover, local texture continuities can serve as the useful guidance for the following determination of the optimal intra prediction modes.

In addition, the local texture characteristics are studied from the perspective of block correlations. Given an interested center block B with certain texture direction $M_{cur}^{(i)}$, assuming that the mode index of k-th neighboring blocks is $M_{nei}^{(k)}$, the expectation of $M_{nei}^{(k)}$ is given by,

$$\mathbb{E}(M_{nei}^{(k)}) = \frac{\sum_{k=1}^{N} M_{nei}^{(k)}}{N}, N > 0, \tag{4}$$

where N is the number of neighboring blocks which should satisfy the constraint of the block correlation $\hat{\rho}$. The relationship between the local texture of current block $M_{cur}^{(i)}$ and the expectation of the textures regarding the neighboring blocks $\mathbb{E}(M_{nei}^{(k)})$ is depicted as Fig. 4. It is interesting to observe that with the rising of the lower bound of $\hat{\rho}$, $\mathbb{E}(M_{nei}^{(k)})$ and $M_{cur}^{(i)}$ become more consistent.

These observations reveal that the strong correlations occur along the texture directions. Moreover, higher correlation blocks tend to share the similar texture direction. As such, it is feasible to employ the block correlations and texture characteristics of the current block and neighboring block to optimize the intra mode subset with higher priorities. In particular, if the current block and neighbor block have high correlation, they may share similar texture direction. With the design philosophy of intra prediction in video coding where the neighboring left and above reconstructed blocks are available for reference, it is feasible to utilize the neighboring intra modes that are finally selected with full RDO to refine the full RD mode list of the current block.

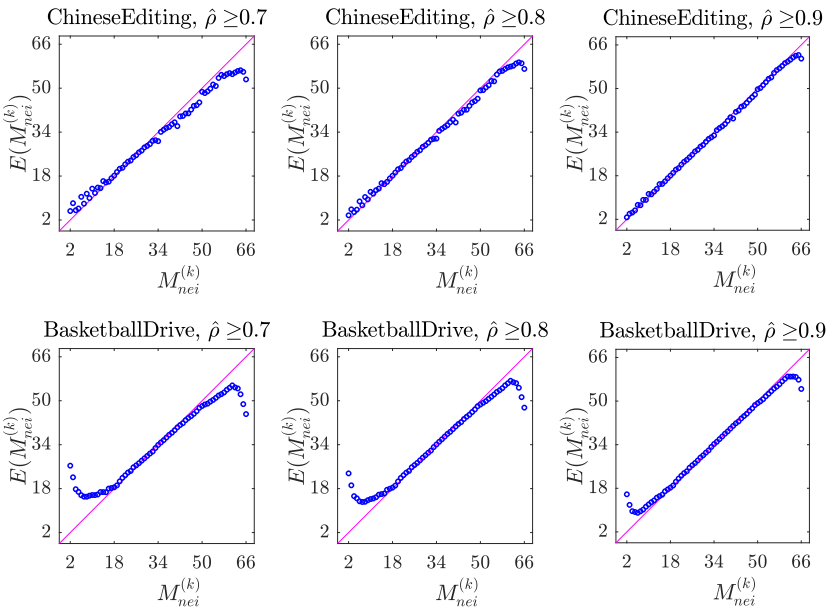

Figure 4: Illustration of the relationship between the mode index of the current block and the average mode index of the neighboring blocks under the certain constraint of $\hat{\rho}$.

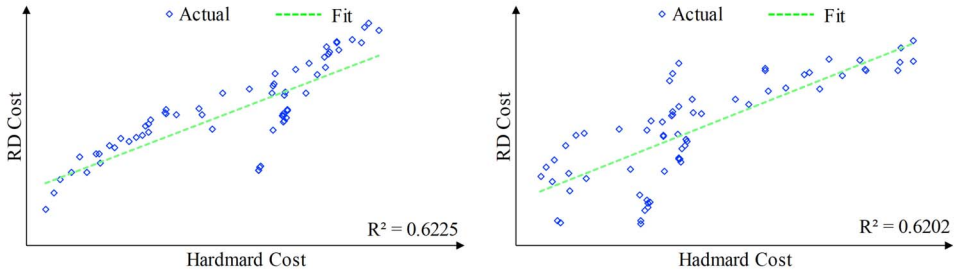

Figure 5: Illustration of the correlation between RD cost and Hadamard cost with different intra modes on two natural scene blocks.

3. Full RD List Optimization

Generally speaking, Hadamard cost reveals positive correlation with the actual RD-cost. As such, the computation simplicity and analytic tractability make it frequently employed by the video encoder. However, Hadamard cost cannot always effectively discriminate the optimal intra modes, especially when the total number of them has been doubled in the emerging video coding standard. To demonstrate this, we show the scatters of the actual RD cost and the Hadamard cost with respect to 67 intra modes in Fig. 5 and Fig. 6. Blocks with size of 16×16 are respectively selected from natural scene and screen content sequences, and the corresponding Hadamard cost and RD cost of each intra modes are recorded. It can be observed that there exhibits a positive correlation between the Hadamard cost and RD cost. However, the low correlations also suggest that there is still room to further optimize the full-RD list.

This further motivates us to investigate the correspondence between optimal

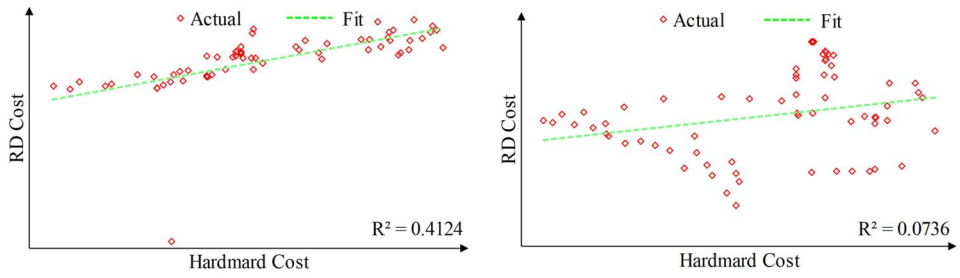

Figure 6: Illustration of the correlation between RD cost and Hadamard cost with different intra modes on two screen content blocks.

Table 1: Illustration of the original and improved mode hit ratios on natural scene videos with different S and X values

S	$\mathcal{M}$					$\mathcal{M}^\star$
	$X=1$	$X=2$	$X=3$	$X=4$	$X=5$	
16	70.9%	82.6%	85.8%	87.9%	89.4%	93.4%
32	65.5%	77.8%	81.8%	84.5%	86.4%	90.5%
64	62.2%	75.4%	80.0%	83.0%	85.0%	89.2%
128	58.3%	71.8%	77.4%	80.7%	83.0%	86.9%
256	55.2%	69.2%	75.1%	78.5%	81.0%	85.8%
512	51.8%	65.9%	72.2%	75.9%	78.7%	82.5%
1024	46.9%	60.4%	66.9%	71.0%	74.0%	78.2%
2048	46.3%	59.1%	64.5%	68.2%	71.3%	74.1%
4096	41.7%	52.5%	57.7%	61.4%	64.2%	67.9%
Average	**59.6%**	**73.0%**	**78.1%**	**81.2%**	**83.4%**	**87.6%**

modes and derived modes from Hadamard based rate-distortion optimization. We use AVS3 reference platform as an example. In particular, X modes with smaller Hadamard cost compose the full-RD set $\mathcal{M} = \{m_0, ..., m_i, ..., m_{X-1}\}$, where the maximum of X is set to 5 in AVS3 encoder. We comprehensively study the hit ratio P of set $\mathcal{M}$ with respect to different block sizes S and the capacity of set $\mathcal{M}$. More specifically, full RD searching is applied with 67 intra modes for each block, with an effort to dig out the ultra optimal mode. For the case that the best mode m_{opt} is included in $\mathcal{M}$, it can be defined as a hit event. Otherwise, it is a miss-hit event, indicating that the Hadamard based cost cannot well predict the actual intra mode. The hit ratio P can be described as,

$$P = \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}[m_{opt}^{(k)} \in \mathcal{M}^{(k)}], \tag{5}$$

where N is the total block number with certain block size. $m_{opt}^{(k)}$ and $\mathcal{M}^{(k)}$ denote the optimal intra mode and full RD set for k-th block, respectively. Statistical results are reported in Table 1 and Table 2. The average hit ratio on natural scene videos is 83.4% when set $\mathcal{M}$ contains five Hardmard-cost-picked modes. Besides, $\mathcal{M}$ performs

Table 2: Illustration of the original and improved mode hit ratios on screen content videos with different S and X values

S	$\mathcal{M}$					$\mathcal{M}^\star$
	$X=1$	$X=2$	$X=3$	$X=4$	$X=5$	
16	56.1%	65.7%	70.1%	73.2%	75.7%	82.0%
32	55.5%	65.4%	70.2%	73.4%	76.0%	79.4%
64	63.4%	72.2%	76.1%	78.8%	80.8%	82.3%
128	69.7%	77.6%	80.8%	83.0%	84.7%	85.3%
256	75.7%	82.3%	84.8%	86.5%	87.8%	87.9%
512	70.0%	77.1%	80.4%	82.6%	84.2%	83.3%
1024	69.0%	75.5%	78.8%	81.0%	82.8%	81.3%
2048	62.2%	69.4%	74.0%	76.8%	78.6%	76.9%
4096	64.1%	71.3%	74.9%	77.8%	79.9%	78.4%
Average	**60.2%**	**69.3%**	**73.5%**	**76.4%**	**78.7%**	**81.6%**

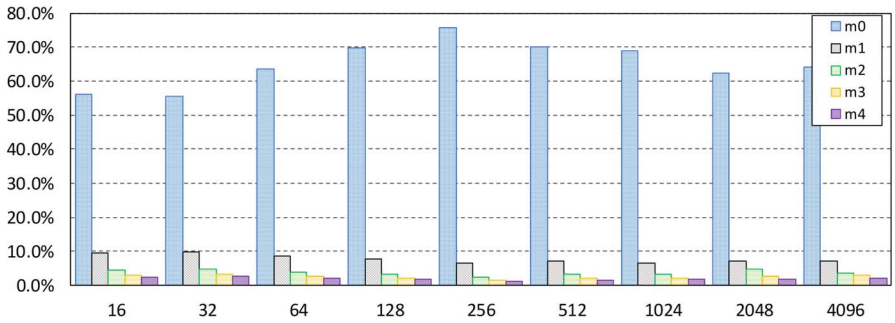

Figure 7: Illustration of the hit ratio with respect to different modes in set $\mathcal{M}$ on natural scene videos.

better on smaller blocks, especially on 4×4 blocks, where the associated hit ratio reaches up to 90%. Moreover, it can be noticed that the hit ratio increases with X, while the increasing trend is smoother when X is larger than 3. It is interesting to observe that the average hit ratio on natural scene videos far exceeds the screen content videos, indicating that the performance of set $\mathcal{M}$ could be influenced by video content.

Subsequently, we focus on the hit ratios regarding the individual mode m_i in $\mathcal{M}$ and temporarily ignore the mis-hitting cases of $\mathcal{M}$. It should be noted that the order of modes in $\mathcal{M}$ is determined according to the associated Hadamard cost, which is given by,

$$J_{HAD}^{(m_i)} < J_{HAD}^{(m_{i+1})}. \tag{6}$$

The hit ratio distribution of m_i is demonstrated in Fig. 7 and Fig. 8. It can be noticed that the hit ratios of last two modes are constantly low, with less than 3% among all block sizes, which indicates that m_3 and m_4 play a relatively minor role in the intra mode selection.

Based on the previous analyses regarding the local texture and correlation, it

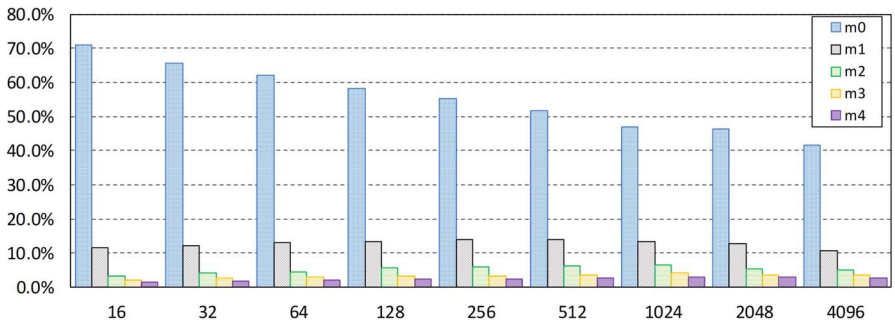

Figure 8: Illustration of the hit ratio with respect to different modes in set $\mathcal{M}$ on screen content videos.

is reasonable to consult the intra modes of neighboring blocks when constructing the set $\mathcal{M}$, if the current block and neighboring block possess high correlations. Considering the fact that strong correlations always exist between neighboring blocks, it is proposed to directly include the neighboring intra modes into set $\mathcal{M}$ without further computing $\hat{\rho}$. However, appending the modes of neighboring blocks to $\mathcal{M}$ will escalate RD rounds, which results in the dramatic increase of coding complexity. $\mathcal{M}$ can be adjusted by weighting the priority of neighboring modes and the last two modes in AVS3. In particular, if the neighboring mode is an angular mode, it will be assigned to higher priority than the last two modes. Furthermore, the less-likely modes will be removed and the capacity of $\mathcal{M}$ is maintained to five. The refined set $\mathcal{M}$ is denoted as $\mathcal{M}^{\star}$. Statistical experiment is conducted with the set $\mathcal{M}^{\star}$, and the results are shown in the last column in Table 1 and Table 2. It can be noticed that the hit ratio increases by 3% to 4%, indicating that the encoder earns more opportunities to access and select the ultra optimal intra modes. Therefore, this also well explains the full-RD list construction with MPMs in HM and VTM, as the modes of neighbors are considered when constructing MPM list.

4. Experimental Results

The proposed joint optimized full-RD list is verified on AVS3 test model HPM-5.0 [8] conforming to common test conditions (CTC) [10] under All Intra (AI) and Random Access (RA) configurations. Quantization parameters are set as 27, 32, 37 and 45. The recommended test sequences regarding natural scene and screen content videos are all involved in the experiment. In addition, BD-Rate [11] is adopted to evaluate the coding performance and negative BD-Rate corresponds to the performance improving. Simulation results are shown in Table 3 and Table 4. The refined full-RD list brings on average 0.72% and 0.35% BD-Rate gains for natural scene videos under AI and RA configurations, respectively. Moreover, on average 3.00% and 2.56% BD-Rate savings can be achieved on screen content videos under AI and RA configurations. Under the complexity constraint, the intra RDO rounds is maintained to be close to the anchor, such that the variations of encoding and decoding time are negligible. The optimized intra full-RD list construction strategy has been adopted by AVS3 software.

Table 3: Performance of the proposed method on HPM-5.0 with natural scene videos under AI and RA configurations.

Seq		AI			RA		
		Y	U	V	Y	U	V
720p	City	-0.80%	-0.11%	-0.53%	-0.47%	1.84%	-0.09%
	Crew	-0.62%	-0.08%	-0.18%	-0.34%	0.38%	0.32%
	Vidyo1	-0.77%	-0.25%	-0.63%	-0.37%	-0.09%	0.22%
	Vidyo3	-0.51%	-0.55%	-0.24%	-0.18%	-0.08%	-1.14%
1080p	BasketballDrive	-0.83%	-0.37%	-0.54%	-0.34%	-0.20%	-0.59%
	Cactus	-0.66%	-0.30%	-0.21%	-0.47%	-0.29%	0.05%
	MarketPlace	-0.80%	-0.49%	-0.13%	-0.08%	-0.16%	0.77%
	RitualDance	-0.72%	-0.20%	-0.31%	-0.44%	-0.04%	0.06%
4k	Tango2	-0.92%	-0.18%	-0.83%	-0.34%	-0.18%	-0.94%
	Campfire	-0.54%	-0.33%	-0.63%	-0.58%	-0.68%	-0.52%
	ParkRunning3	-0.34%	-0.45%	-0.34%	-0.16%	-0.26%	-0.19%
	DaylightRoad2	-1.12%	-0.29%	-0.53%	-0.43%	-0.26%	-0.40%
720p		-0.67%	-0.25%	-0.40%	-0.34%	0.52%	-0.17%
1080p		-0.75%	-0.34%	-0.30%	-0.33%	-0.17%	0.07%
4k		-0.73%	-0.31%	-0.59%	-0.38%	-0.34%	-0.51%
Average		**-0.72%**	**-0.30%**	**-0.43%**	**-0.35%**	**0.00%**	**-0.20%**
Enc Time			97%			103%	
Dec Time			100%			100%	

5. Conclusions

In this paper, we provide comprehensive analyses on the local texture direction and block correlation characteristics, grounded on which the optimization of rate-distortion optimized intra coding with certain complexity constraint is made possible. This paper provides a scientifically-sound way to analyze the modes that should be within the final full RD list, and experimental results show that on average 0.72% and 3.00% BD-Rate savings can be achieved for natural scene and screen content videos with negligible changing of encoding and decoding time.

Acknowledgment

This work was supported in part by Hong Kong ITF UICP under Grant 9440203, in part by the Hong Kong RGC Early Career Scheme under Grant 9048122 (CityU 21211018), and in part by the City University of Hong Kong Applied Research Grant under Grant 9667192. This work was done while the author was a research intern in ByteDance Inc.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

Table 4: Performance of the proposed method on HPM-5.0 with screen content videos under AI and RA configurations.

Sequence		AI			RA		
		Y	U	V	Y	U	V
TGM	FlyingGraphics	-3.68%	-2.80%	-2.66%	-0.67%	-0.57%	-0.17%
	Desktop	-4.05%	-3.60%	-3.47%	-2.28%	-2.93%	-2.65%
	Console	-3.39%	-2.48%	-2.50%	-1.09%	-1.10%	-0.89%
	ChineseEditing	-1.75%	-1.36%	-1.43%	-1.65%	-1.45%	-1.22%
	EnglishEditing	-1.80%	-1.23%	-1.20%	-2.04%	-1.37%	-1.41%
	Spreadheet	-2.94%	-2.40%	-2.76%	-3.19%	-2.99%	-3.20%
	BitstreamAnalyzer	-4.12%	-3.16%	-2.61%	-7.88%	-5.03%	-4.89%
	CircuitLayoutP	-1.59%	-1.04%	-1.00%	-1.87%	-0.80%	-0.88%
	Program	-2.80%	-2.32%	-2.65%	-2.05%	-1.73%	-2.14%
	Web_en	-2.94%	-2.22%	-2.37%	-1.84%	-1.89%	-1.99%
	Word_excel	-3.54%	-3.24%	-3.28%	-3.54%	-3.18%	-3.34%
MC	Program_vidyo	-3.41%	-2.48%	-2.57%	-2.60%	-2.20%	-2.52%
	TGM	-2.96%	-2.35%	-2.36%	-2.55%	-2.10%	-2.07%
	MC	-3.41%	-2.48%	-2.57%	-2.60%	-2.20%	-2.52%
Average		**-3.00%**	**-2.36%**	**-2.38%**	**-2.56%**	**-2.10%**	**-2.11%**
Enc Time			98%			99%	
Dec Time			100%			100%	

[2] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[3] J. Chen B. Bross and S. Liu, "Versatile video coding (draft 3)," *JVET L1001*, 2018.

[4] N. Choi, Y. Piao, K. Choi, and C. Kim, "CE 3.3 related: Intra 67 modes coding with 3 MPM," *Joint Video Exploration Team (JVET), doc. JVET-K0529*, Jul. 2018.

[5] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, Nov 1998.

[6] L. Zhao, L. Zhang, S. Ma, and D. Zhao, "Fast mode decision algorithm for intra prediction in HEVC," in *2011 Visual Communications and Image Processing (VCIP)*, Nov 2011, pp. 1–4.

[7] J. Hu and J. D. Gibson, "New rate distortion bounds for natural videos based on a texture-dependent correlation model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 8, pp. 1081–1094, Aug 2009.

[8] "AVS3 software repository," `/Public/codec/video_codec/HPM/HPM-5.0.zip`.

[9] Ali Habibi and Paul Wintz, "Image coding by linear transformation and block quantization," *IEEE Transactions on Communication Technology*, vol. 19, no. 1, pp. 50–62, 1971.

[10] J. Chen and K. Fan, "AVS3-P2 common test conditions v8.0," *Audio Video Coding Standard Workgroup of China, doc.AVS N2727*, Sep. 2019.

[11] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T SG 16 Q.6 VCEG-M33*, 2001.