Revisiting Local Texture Correlation for Rate-distortion Optimized Intra Coding

Meng Wang[1], Junru Li[2], Li Zhang[3], Hongbin Liu[4],

Jizheng Xu[3] and Shiqi Wang[1]

[1] Department of Computer Science, City University of Hong Kong, Hong Kong, China.

[2] Institute of Digital Media, Peking University, Beijing, China.

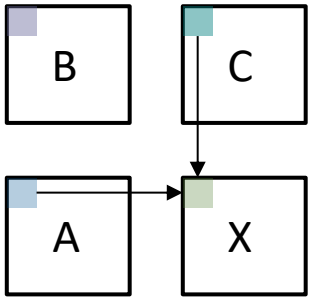
[3] Bytedance Inc., San Diego CA. 92122 USA.

[4]Bytedance (HK) Limited., Hong Kong, China.

Outline

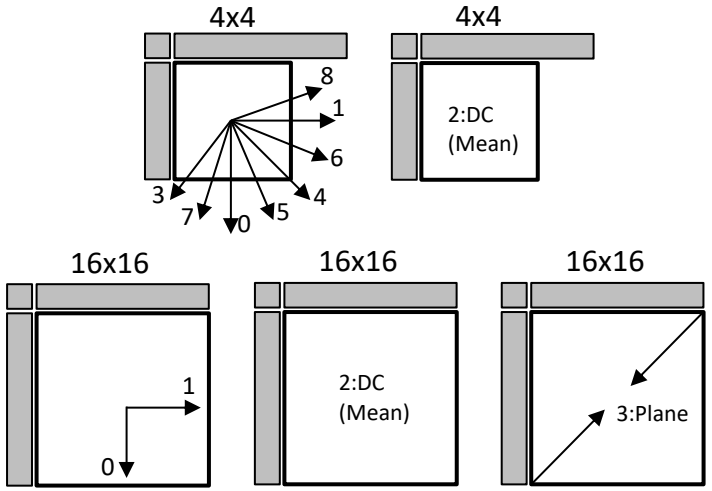
- Introduction
- Texture-Dependent Correlations and Local Texture Directions
- Full-RD List Optimization
- Experimental Results

Intra Prediction



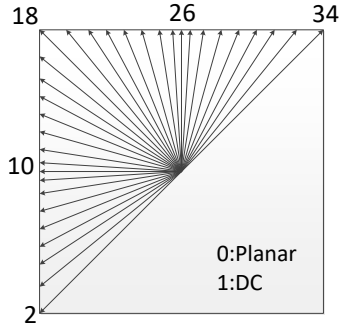
DC/AC Prediction

MPEG-2/MPEG-4



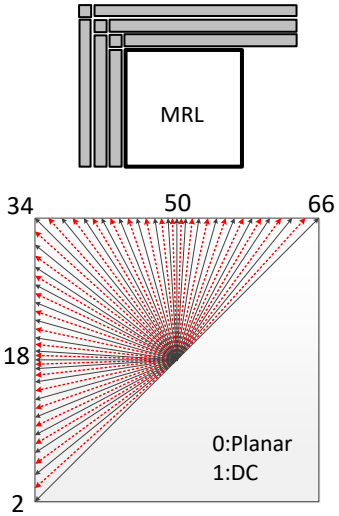
4x4 9 intra modes
16x16 4 intra modes
1 MPM

H.264



35 intra modes
(33 Angular+Planar+DC)

H.265



67 intra modes
(65 Angular+Planar+DC)

VVC/AVS3

Intra Prediction

- Rate-distortion Optimization is served as the mode selection criteria:

$$J^i = D^i + \lambda \cdot R^i$$

- The main-stream video codecs employ rough mode decision (RMD) method with Hadamard cost:

$$J^i_{HAD} = D^i_{HAD} + \lambda \cdot \hat{R}^i$$

- Consequently, *X* candidates with lower J^i_{HAD} are selected as full-RD candidates.

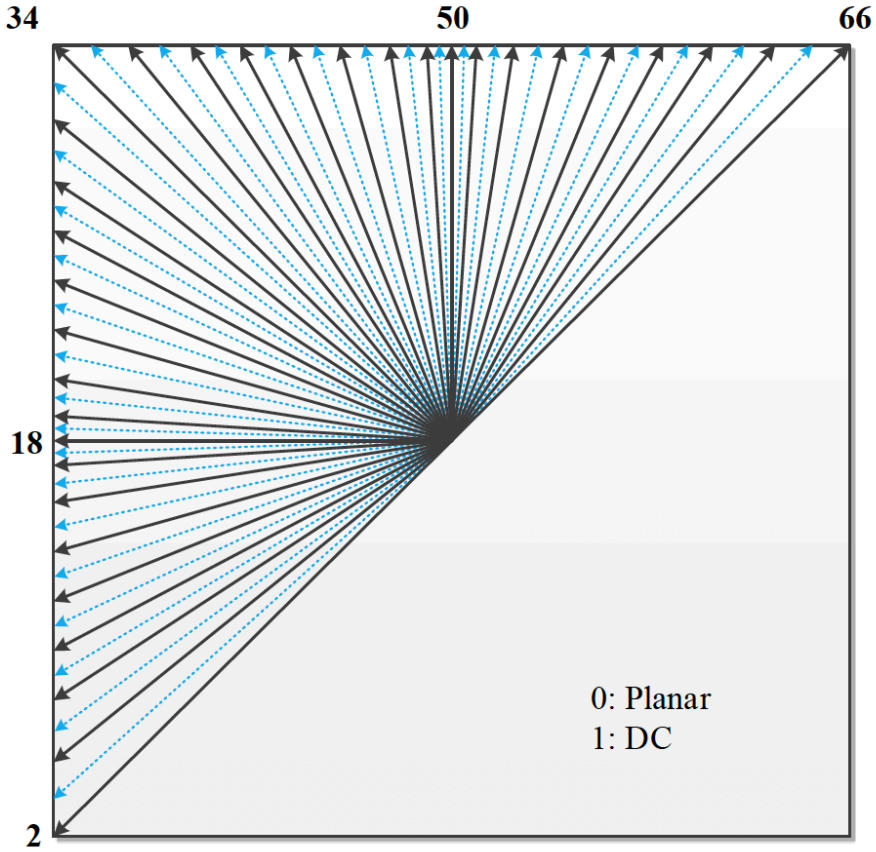


Fig 1. Intra prediction modes in VVC and AVS3

Hadamard-cost vs. RD-cost

- Hadamard-cost reveals positive correlation with the actual RD-cost.
- Merits of Hadamard-cost
  - Computation simplicity
  - Analytic tractability
- However, Hadamard cost cannot always effectively discriminate the optimal intra mode.

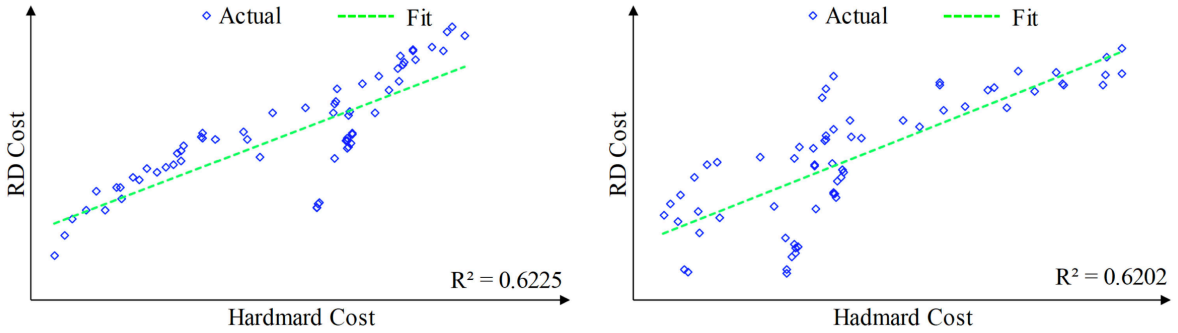


Fig 2. Two blocks from natural scene video

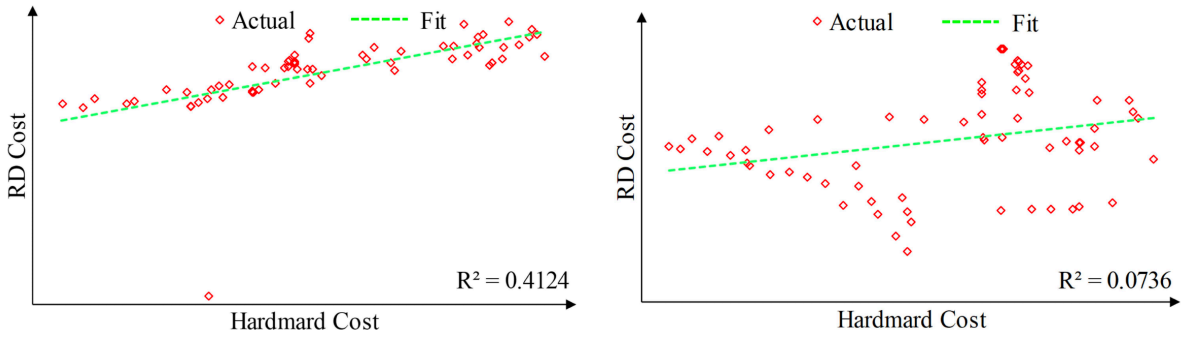


Fig 3. Two blocks from screen content video

Hadamard-cost vs. RD-cost

- *X* modes with lower 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.
- The hit ratio *P* of mode set *M* is studied with respect to different block sizes *S* and the capacity of the mode set.

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

- Full RD searching is applied with 67 intra modes for each block, with an effort to dig out the ultra optimal mode.

Hit Ratios of Set *M*

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

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

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

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

Hit Ratios of individual mode in set *M*

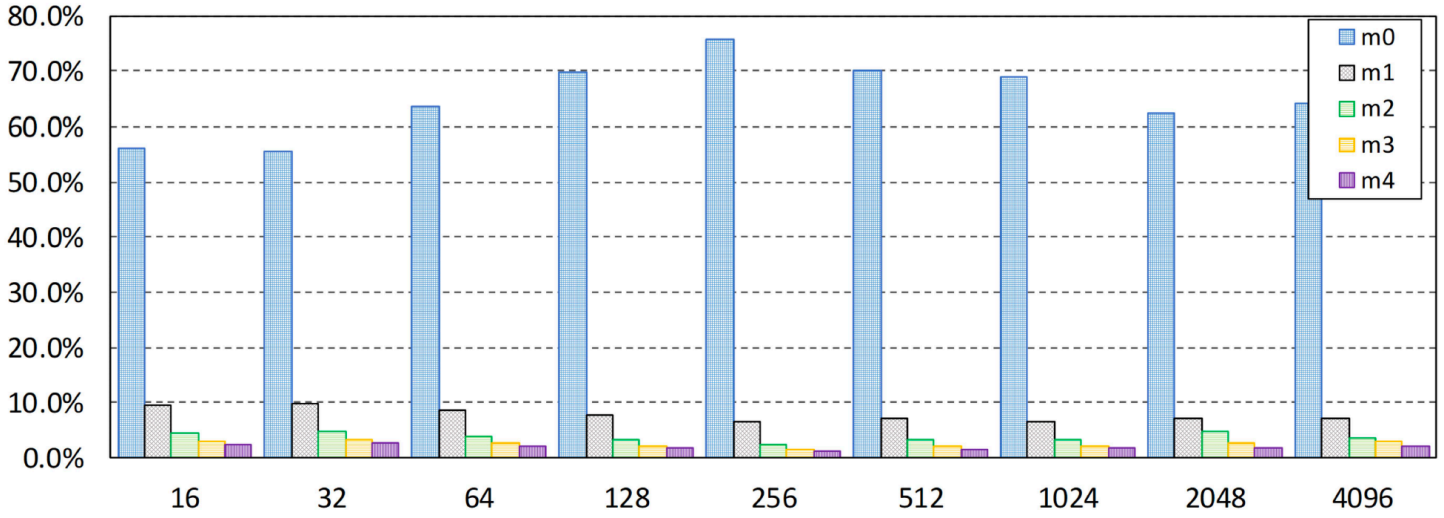


Fig. 4: Illustration of the mode hit ratios on natural scene videos with different *S* and *X* values.

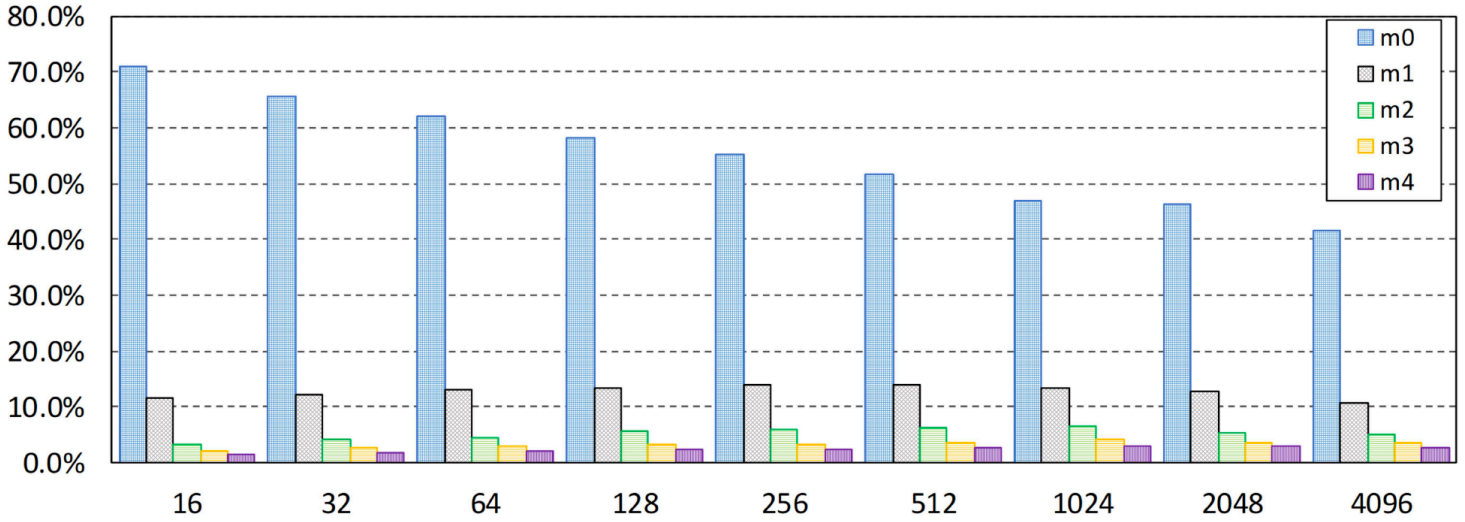
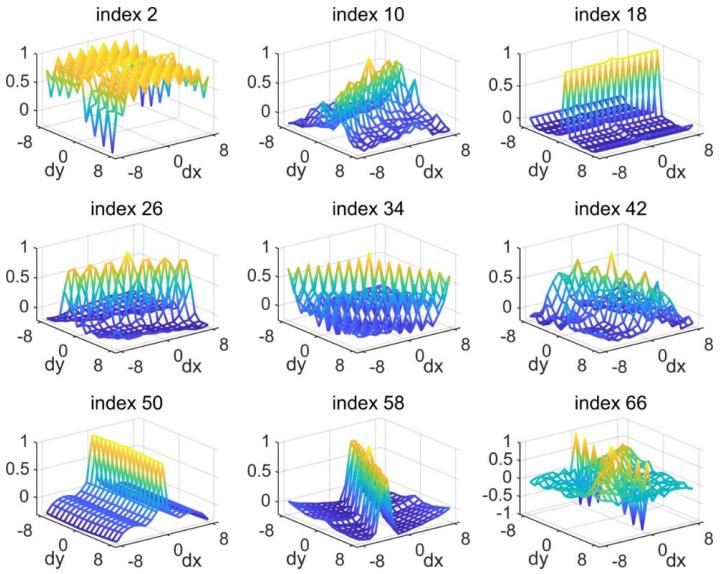


Fig. 5: Illustration of the mode hit ratios on screen content videos with different *S* and *X* values.

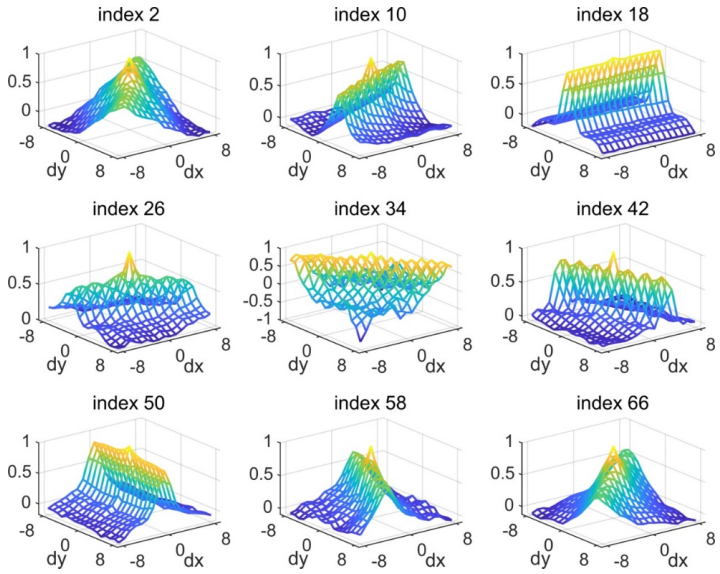
Texture-dependent Correlations

- Pearson Correlation Coefficient
  - Block size: 8 x 8
  - Offset: [-8, +8]

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



Natural scene video： *Cactus*



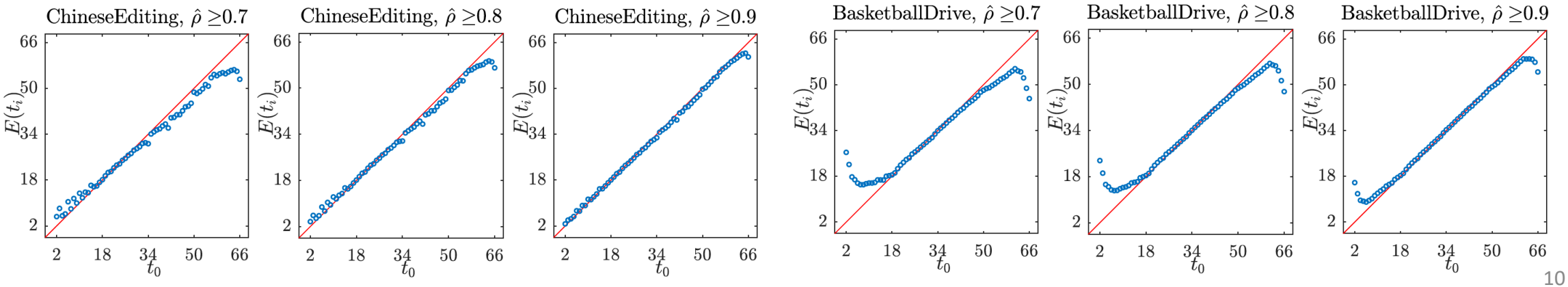
Screen content video： *Console*

Local Texture Directions

- Expectation of the neighboring blocks

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

  - N is the total number of neighboring blocks that satisfy the lower bound of the constraint regarding block correlations $\hat{\rho}$.
  - t_i is the mode index of i-th neighboring block.



Reconstruction of the Mode Set *M*

- Reconstruct the set *M* with neighboring modes-->*M**

Table 3: Illustration of the 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%**

Table 4: Illustration of the 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%**

Experimental Results

- Test Platform:
  - AVS3 test model HPM-5.0.
- QPs are set with 27, 32, 37, 45 conforming to common test conditions.
- Configuration : AI and RA.

Experimental Results

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%	

Experimental Results

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%	

Conclusion

- Provide comprehensive analyses on the local texture direction and block correlation characteristics.
- Based on these, the optimization of the rate-distortion optimized intra coding with certain complexity constraint is made possible.
- Reconstruct the full RD mode list for AVS3.
- 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.

香港城市大學
City University of Hong Kong


Thank You
Q & A