IMPLICIT-SELECTED TRANSFORM IN VIDEO CODING

[1]Yuhuai Zhang, [2]Kai Zhang, [2]Li Zhang, [2]Hongbin Liu, [2]Yue Wang,
[1]Shanshe Wang, [1]Siwei Ma, [1]Wen Gao

[1]Institute of Digital Media, Peking University, Beijing, China
[2]Bytedance Inc., San Diego, CA, USA
{yhzhangvcl, sswang, swma and wgao}@pku.edu.cn
{zhangkai.video, lizhang.idm, liuhongbin.01 and wangyue.v}@bytedance.com

ABSTRACT

AVS-3 is an emerging next-generation video coding standard, in which the transform-coding plays an important role. This paper presents a method of Implicit-Selected Transform (IST) to further improve the transform-coding in AVS-3. With IST, two different types of separable transform cores are introduced to perform transform-coding on residual signals of blocks with intra-prediction. At the encoder side, the two types of transform cores are tested and selected according to a Rate-Distortion Optimization (RDO) criterion. Instead of signaling the selected transform type explicitly, the information is hidden into the Parity of the Number of Non-zero Coefficients (PNNC) of a transformed block. At the decoder side, the selection is implicitly determined by checking PNNC. Experimental results show that the proposed method can achieve 0.64% and 0.35% BD-rate savings on average under All-intra and Random-access configurations, respectively, with negligible decoding time changes. IST has been adopted into AVS-3.

Index Terms— AVS3, transform-coding, Implicit-Selected Transform

1. INTRODUCTION

The third generation of the audio video coding standard (AVS-3) is a new generation video coding standard developed by the Audio Video Coding Standard Working Group of China, which formed the main part of IEEE 1857 WG [1]. The baseline profile of AVS-3 [2], which can achieve about 20% BD-rate saving compared to AVS-2 [3] and high efficient video coding (HEVC) [4], has been finalized in Mar. 2019. To further improve the coding performance, the AVS working group has started to develop a high profile of AVS-3 since then.

In the past year, a number of efficient coding algorithms have been adopted into AVS-3, including new partitioning methods, efficient intra/inter prediction, adaptive transform, advanced entropy coding, etc. These new coding tools strengthen the adaptability of coding structure and improve the quality of reconstruction frames.

Transform-coding and transformed coefficient signaling play an important role in video coding. Two new transform methods have been adopted in AVS-3, both of which are applied to the transform-coding of inter-coded blocks. Partition-Based Transform (PBT) [5] splits the residuals into four sub-blocks. Each sub-block utilizes fixed horizontal and vertical transform cores. It is beneficial to transform a residual block in which there are obvious boundaries. Another transform named Sub-Block Transform (SBT) [6]. SBT splits the residuals blocks into two subblocks vertically or horizontally, and one of the split sub-block is zero out, the other sub-block contained non-zero coefficients utilize fixed transform core, based on the split type and position in the whole block. Besides transform-coding, a novel coefficient coding method named Scan Region based Coefficient Coding (SRCC) [7] has also been adopted into AVS-3. With SRCC, the horizontal ordinate of the most right non-zero coefficient and the vertical ordinate of the bottom non-zero coefficient are signaled to restrict the scan region of transform block, which can efficiently reduce the size of significance map. Moreover, SRCC applies a zig-zag scanning pattern to scan the significance map, coefficient levels and sign data inside the non-zero region.

Although previous works contribute a lot on the transform and coefficient coding in AVS-3, there are still two shortcomings of the transform-coding in AVS-3. First, the previous works mostly focus on the transform-coding of inter-coded blocks, ignoring the intra-coded blocks. However, the residual signals of intra-coded blocks usually possess much higher energy than those of inter-coded blocks. Second, the transforms in AVS-3 are fixed, lacking the flexibility to the diversity of image contents.

To address the two problems, we propose a method of Implicit-Selected Transform (IST) to further improve the transform-coding in AVS-3. With IST, two different types of separable transform cores are introduced to perform transform-coding on residual signals of blocks with intra-prediction. At the encoder side, the two types of transform

978-1-7281-1485-9/20/$31.00 © 2020 IEEE

cores are tested and selected according to a Rate-Distortion Optimization (RDO) criterion. Instead of signaling the selected transform type explicitly, the information is hidden into the Parity of the Number of Non-zero Coefficients (PNNC) of a transformed block. At the decoder side, the selection is implicitly determined by checking PNNC. Experimental results show that the proposed method can achieve 0.64% and 0.35% BD-rate savings on average under All-intra and Random-access configurations, respectively, with negligible decoding time changes. IST has been adopted into AVS-3.

The remainder of this paper is organized as follows: transform coding tools are reviewed in section II. And the proposed transform coding method is presented in Section III followed by experimental results in Section IV. Lastly, Section V concludes this paper.

2. REVIEW OF TRANSFORM CODING

2.1. KLT Theory and Application

Theoretically, Karhunen-Loève Transform (KLT) [8] is the optimal transform method, which aims to derive an optimal orthonormal core for the sample vectors.

However, the transform core is not stationary, and it should be adaptive to residual blocks, thus DCT-II is used to approximate to KLT, which is a separable transform method. Separable transform owns low complexity for a 2D block since each row and column are transformed separately. For a block with size N×N, the multiplication matrix size is N×N as well, and 2N times transform are needed including N times horizontal transform and N times vertical transform. For the other transform method, non-separable transform, it needs to reshape a 2D N×N block into a 1D $1\times N^2$ vector, thus the final core size is $N^2\times N^2$, therefore, it can separate different components of residual blocks more successfully at the cost of high computation complexity. By comparison, the separable transform can balance the performance and complexity better.

2.2. Development of Separable Transform

One typical work of separable transform is directional transform [9]. After intra prediction, the significantly directional texture information still remains in residuals, especially when the prediction mode is directional. Based on the directional edge feature, a separable directional transform is proposed to design the transform basis to de-correlate the directional residual signals.

In [10], the resulting optimal transform of H.264/AVC [11] is shown to be close to a sine transform. Prediction and boundary informations are utilized to switch the transform between a sine-like transform and DCT.

Multiple-Model KLT (MM-KLT) [12] and Rate-Distortion Optimization Transform (RDOT) [13] are proposed to introduce multiple transform bases based on both the directional edge feature and the diverse characteristics of residual signals. The optimal transform core among several core candidates is selected according to the RDO criterion. To indicate the best transform core, several Transform Unit (TU) level indices are signaled, which brings overhead bits.

To reduce R-D comparison calculations, content dependent directional transform (CDDT) [14] is proposed. The content features are considered, and feature matching is performed to select the best transform basis at the encoder side. Each transform basis candidate is trained by singular value decomposition (SVD) off-line.

Multiple Transform Set (MTS) is a transform technique adopted in Versatile Video Coding (VVC) [15], which is developed from enhanced multiple transform (EMT) [16]. According to statistics of inter and intra predicted block residual, DCT-VIII and DST-VII are utilized in MTS. As in [12] [13] [14], the selection of transforms is signaled.

A commutative mode-dependent transform [17] is proposed to make a trade-off between the coding efficiency and the coding complexity. Only DCT-II and DST-VII are utilized. According to the observation that the residuals of adjacent intra modes have similar characteristics, a refined commutative mode-dependent method is utilized to select the optimal transform type.

The formula of DST-VII is given as Table.1, and the intuitive basis images of 8×8 size DST-VII is shown as Fig.1.

Table 1: Increased Transform Basis

Transform	**Basis function** $T_i(j), i, j = 0, 1, ..., N-1$
DST-VII	$T_i(j) = \sqrt{\frac{2}{N+1}} \cdot sin(\frac{\pi \cdot (2i+1) \cdot (j+1)}{2N+1})$

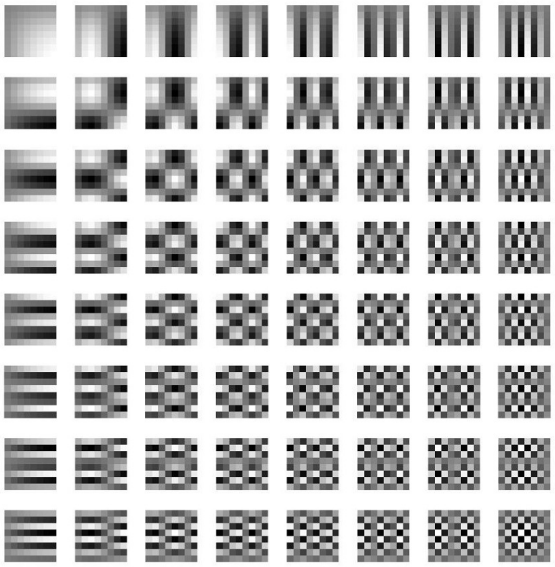


Fig. 1: Basis images of separable DST-VII.

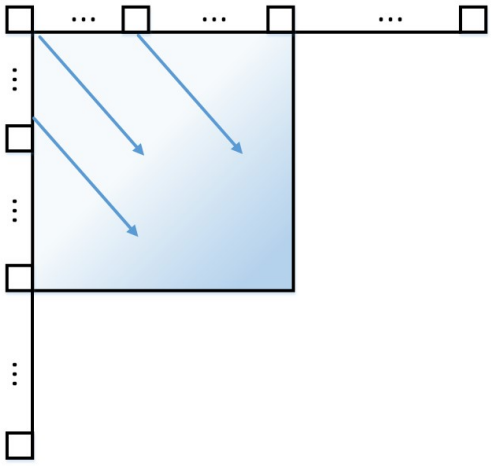
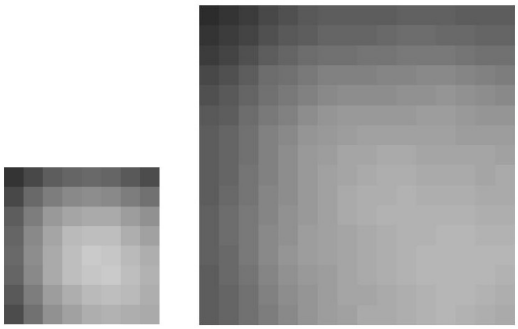


Fig. 2: The prediction accuracy of intra prediction modes.



(a) 8×8 size (b) 16×16 size

Fig. 3: The average intra residuals of different size.

3. IMPLICIT SELECTED TRANSFORM

In this section, we detail the method of IST, including the motivations, design principles and the work flow.

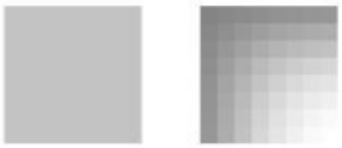
3.1. Motivations

Generally, the top-left part of the current block can be predicted well in intra prediction modes. Because, the reference pixels region is on the top and left of the current block, and the spatial correlation is limited. As Fig.2 shown, a typical intra directional mode, whose direction is showed by arrows. Along the direction of the arrow, the prediction accuracy gradually decreases. Statistics data is collected from residuals of 8×8 and 16×16 intra-coded blocks, and the average data is visualized in Fig.3.

Obviously, the further away from the left and the top boundary, the larger the residual data is. Since the most characteristics of intra block residuals is that the distortion is relative to the position in one residual block, DCT-II is not an optimal transform. A secondary transform is used in AVS-3 to alleviate the shortcomings of DCT-II. However, the benefit of the secondary transform is limited because it can only be applied after the primary transform. It is necessary to improve the efficiency of the primary transform.

How to indicate the choosing transform type is critical to the coding performance. In [12] [13] [14] and [18], the transform type is directly signaled with extra bits. Based on them, prediction mode dependency is a general hint to select the transform type, such as in [9] [17].

To avoid the overhead bits, we propose an information hiding method to select the optimal transform without signaling the indication of the selected transform.



(a) DCT-II (b) DST-VII

Fig. 4: The first basis image of separable DCT-II and separable DST-VII.

Table 2: Transform Types indication

Transform Type	Horizontal Type	Vertical Type
I	DCT-II	DCT-II
II	DST-VII	DST-VII

3.2. Proposed Method

According to the previous studies [10] [16] [18], DST-VII is an efficient transform core to de-correlate the relevance of residuals of intra-coded blocks. Therefore, the proposed method introduces DST-VII as an additional transform candidate for horizontal and vertical transform, as shown in Table.1.

And the intuitive first basis image of 8×8 size DCT-II and 8×8 size DST-VII are shown in Fig.4. It is easy to find the difference between the first basis images of two figures. The first basis image of DCT-II, which is called direct current (DC) basis as usual, extracts the average value of one block. But, the first basis image of DST-VII shows much more energy bias on the left-bottom residuals of the block.

To avoid signaling the indication of the selection, we propose a coefficients-dependent transform-type selection method, without imposing any overhead bits. PNNC is utilized to determine the transform type, which is shown in Table.2, and it is straight-forward to derive the parity by setting a counter to count the non-zero coefficients at the de-

Table 3: Transform Types indication

PNNC	Transform Type
even	I
odd	II

coder. The specific indication is shown in Table.3.

However, PNNC does not always correspond to the selected transform type at the encoder side after quantization. To address this problem, a parity adjustment method is embedded into the quantization and coefficients coding process at the encoder. Based on the quantization results, setting a non-zero coefficient to zero or vice-versa can reverse the parity. Theatrically, the optimal setting solution can be found by traversing all possible combinations, with an extremely high complexity at the level of $O(2^N)$. As a feasible solution, a simplified strategy is designed as shown in Fig.5:

- Step 1: Checking the number of non-zero coefficients, whether the PNNC indicates the correct transform type.
- Step 2: If the number is bigger than one, find the last non-zero coefficient and set it zero.
- Step 3: If the number is equal to one, setting the first zero coefficient to 1 or -1, according to the sign of original value before quantization.

Since the last non-zero coefficient usually holds the minimum value of all coefficients in a block, the impact on the quantization results is negligible. Besides, some bit savings can be achieved by setting the last non-zero coefficient zero. After the adjustment at the encoder, the PNNC can correctly indicate the selected transform type.

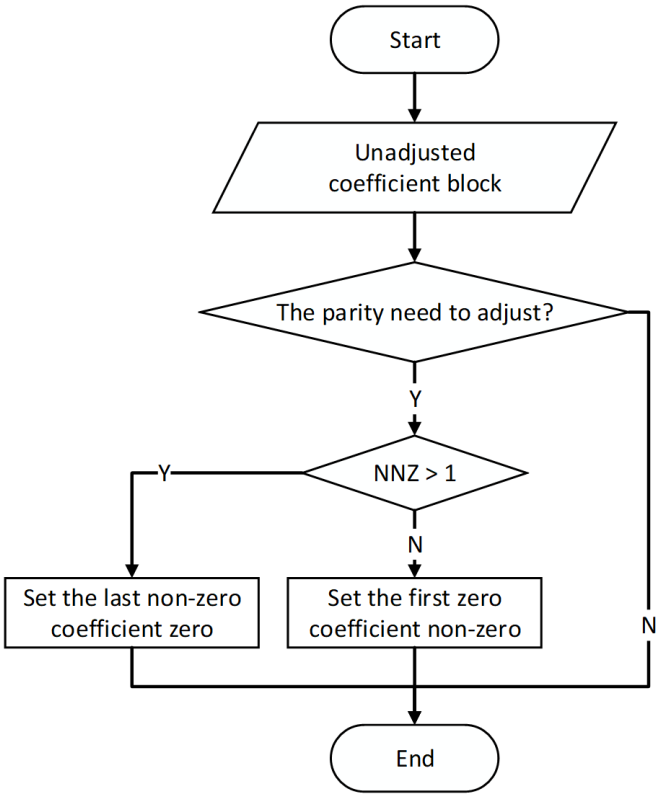


Fig. 5: Flow-chart of the parity adjustment process.

4. EXPERIMENTAL RESULTS

To verify the efficiency of our method fairly, we implemented the proposed method in HPM-5.0 and tested it under the Common Test Condition (CTC) [19]. To show a base performance of IST, the coefficients restriction technique for big blocks and the harmonization approach between IST and SRCC are both turned off.

The performance of the proposed method is shown in Table 4 and Table 5. The experimental results verify that the proposed method is efficient to improve coding performance. The proposed method can achieve 0.64% and 0.35% BD-rate savings on average under All-intra and Random-access configurations, respectively, with negligible decoding time changes. On sequences with rich contents and movements included, such as Tango2, MarketPlace, and City, the coding gains are even higher. The encoding time under the AI configurations is increased because of the selection process of the two transforms. But the encoding time changes are negligible under the RA configurations.

Fig.6 shows the percentages of the two transforms selected at the encoder. The probability of choosing DST-VII is around 35% on average. Especially, for high QP tests, the probability of choosing DST-VII can achieve exceed 40%. Although there is a secondary transform process which is applied after DCT-II but not after DST-VII, DST-VII can still be chosen by about one-third blocks. The statistics demonstrate that DST-VII is an efficient transform type for intra-coded blocks.

5. CONCLUSION

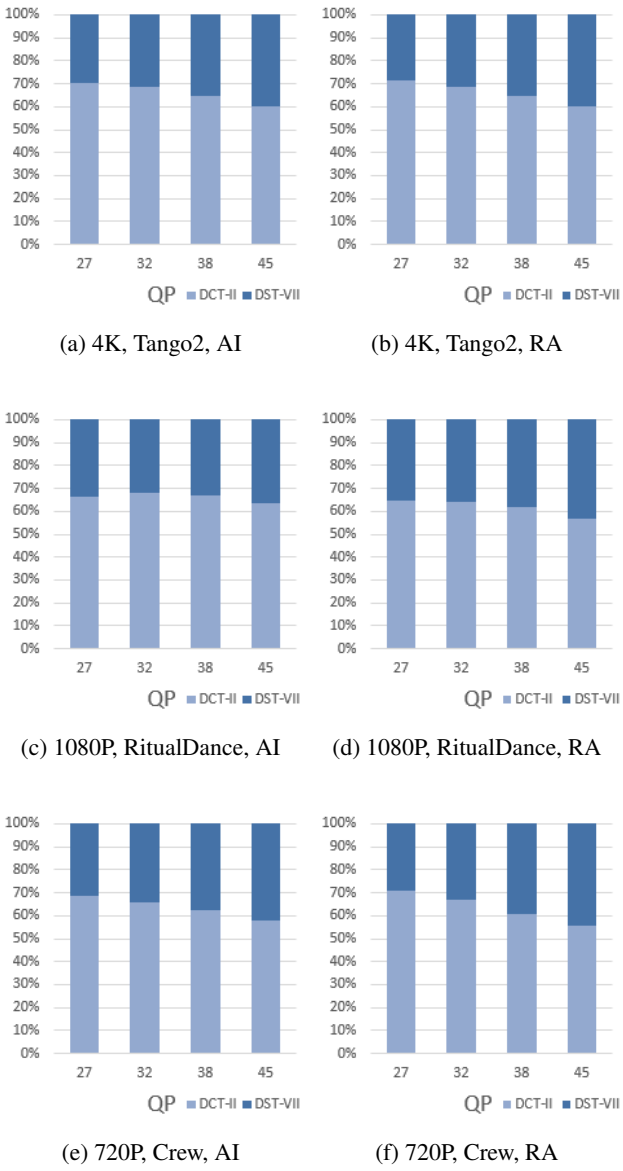
In this paper, we propose a novel IST method. DCT-II and DST-VII transform are involved as the transform candidates. Especially, the parity of non-zero coefficients is utilized to indicate the transform type, so there are no overheads bits. The experimental results show that the proposed IST method can achieve 0.64% and 0.35% BD-Rate savings on average for AI and RA configurations, respectively. IST has been adopted into AVS-3.

Table 4: Coding Performance under AI configurations

resolution	sequence	BD-Rate		
		Y	*U*	*V*
4K	Tango2	-0.63%	-1.77%	-1.54%
	Campfire	-0.60%	-0.63%	-0.68%
	Parkrunning3	-0.43%	-1.80%	-1.76%
	DayLightRoad2	-0.46%	-1.21%	-1.28%
1080P	BasketballDrive	-0.29%	-1.01%	-0.97%
	Cactus	-0.75%	-1.54%	-1.36%
	MarketPlace	-1.34%	-2.12%	-1.42%
	RitualDance	-0.74%	-1.87%	-1.64%
720P	City	-0.78%	-1.45%	-1.27%
	Crew	-0.39%	-1.29%	-0.91%
	Vidyo1	-0.62%	-1.45%	-1.88%
	Vidyo3	-0.68%	-0.66%	-1.28%
4K (3840×2160)		-0.53%	-1.35%	-1.31%
1080P (1920×1080)		-0.78%	-1.63%	-1.35%
720P (1280×720)		-0.62%	-1.21%	-1.34%
Overall		-0.64%	-1.40%	-1.33%
Enc time		121%		
Dec time		101%		

Table 5: Coding Performance under RA configurations

resolution	sequence	BD-Rate		
		Y	*U*	*V*
4K	Tango2	-0.39%	-0.92%	-0.62%
	Campfire	-0.15%	-0.51%	-0.52%
	Parkrunning3	-0.34%	-0.79%	-0.68%
	DayLightRoad2	-0.17%	-0.23%	-0.44%
1080P	BasketballDrive	-0.40%	-0.03%	-0.66%
	Cactus	-0.57%	-0.58%	-1.37%
	MarketPlace	-0.54%	-0.63%	-0.37%
	RitualDance	-0.15%	-0.41%	0.12%
720P	City	-0.65%	-1.67%	-1.28%
	Crew	-0.31%	0.65%	0.34%
	Vidyo1	-0.34%	-0.64%	0.10%
	Vidyo3	-0.20%	0.20%	-0.33%
4K (3840×2160)		-0.27%	-0.61%	-0.56%
1080P (1920×1080)		-0.41%	-0.41%	-0.57%
720P (1280×720)		-0.37%	-0.36%	-0.29%
Overall		-0.35%	-0.46%	-0.48%
Enc time		101%		
Dec time		100%		



(a) 4K, Tango2, AI (b) 4K, Tango2, RA

(c) 1080P, RitualDance, AI (d) 1080P, RitualDance, RA

(e) 720P, Crew, AI (f) 720P, Crew, RA

Fig. 6: The probability of switching DST-VII transform.

6. ACKNOWLEDGEMENT

This work was done while the author was a research intern in ByteDance Inc.

7. REFERENCES

[1] Siwei Ma, Shiqi Wang, and Gao Wen, "Overview of ieee 1857 video coding standard," in *2013 IEEE International Conference on Image Processing*, 2014.

[2] Fan Liang, Siwei Ma, and Wu Feng, "Overview of avs video standard," in *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, 2004.

[3] Shanshe Wang, Falei Luo, and Siwei Ma, "Overview of the second generation avs video coding standard (avs2)," *ZTE COMMUNICATIONS*, vol. 14, no. 1, pp. 7–15, 2016.

[4] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas 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, 2012.

[5] Liqiang Wang, Benben Niu, Yongbing Lin, Quanhe Yu, Jianhua Zheng, and Yun He, "Texture and position based multiple transform for inter-predicted residue coding," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2018, pp. 31–35.

[6] Yin Zhao and Haitao Yang, "Ce1.2: Sub block transform, m4876," AVS3-P2, Sep. 2019.

[7] Fan Wang, Xiao Ouyang, and ZhuoYi Lv, "Scan region-based coefficient coding, m4763," AVS3-P2, Jun. 2019.

[8] Nasir Ahmed, T. Natarajan, and Kamisetty R. Rao, "Discrete cosine transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[9] Ye Yan and Marta Karczewicz, "Improved h.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008.

[10] Jingning Han, Ankur Saxena, and Kenneth Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010.

[11] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay 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, 2003.

[12] Kai Zhang, Shawmin Lei, and Wen Gao, "Enhanced intra prediction and transform for video coding," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 3381–3384.

[13] Xin Zhao, Li Zhang, Siwei Ma, and Wen Gao, "Video coding with rate-distortion optimized transform," *IEEE transactions on Circuits and Systems for Video Technology*, vol. 22, no. 1, pp. 138–151, 2011.

[14] Long Xu and King Ngi Ngan, "Video content dependent directional transform for high performance video coding," in *2012 IEEE International Conference on Multimedia Expo Workshops (ICMEW 2012)*, 2012.

[15] JR Ohm and GJ Sullivan, "Versatile video coding–towards the next generation of video compression," in *Picture Coding Symposium 2018*, 2018.

[16] Xin Zhao, Jianle Chen, Marta Karczewicz, Li Zhang, Xiang Li, and Wei-Jung Chien, "Enhanced multiple transform for video coding," in *2016 Data Compression Conference (DCC)*. IEEE, 2016, pp. 73–82.

[17] Min Mao, Ce Zhu, Yuyang Liu, Yongbing Lin, and Jianhua Zheng, "Commutative mode-dependent transform for video intra coding," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2018.

[18] Xin Zhao, Jianle Chen, Marta Karczewicz, Amir Said, and Vadim Seregin, "Joint separable and non-separable transforms for next-generation video coding," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, pp. 1–1.

[19] Jie Chen and Kui Fan, "Avs3-p2 common test condition v6.0, n2727," AVS3-P2, Mar. 2019.