Texture and Position Based Multiple Transform for Inter-Predicted Residue Coding

Liqiang Wang*, Benben Niu*, Yongbing Lin†, Quanhe Yu†, Jianhua Zheng†, Yun He*
* Tsinghua National Laboratory for Information Science and Technology,
Department of Electronic Engineering, Tsinghua University, Beijing, China
† HiSilicon Technologies Co., Ltd, Beijing, China
E-mail: wlq15@mails.tsinghua.edu.cn

***Abstract*—The next generation video coding standard beyond HEVC is in the study. Many efficient coding tools are introduced to the hybrid framework. Among the new coding tools, EMT focuses on transform module by multiple transform modes to improve the transform efficiency. However, EMT is implemented on residual blocks ignoring the details of the residual distribution. Hence, EMT has some drawbacks to be repaired. In this paper, we propose a method, named texture and position based multiple transform (TPBMT), to refine EMT in the framework. What's more, the efficiency of TPBMT is verified by sufficient experimental data. When implemented on JEM7.0, TPBMT attains average 1.26%, 1.66% and 1.62% for Y, U and V, respectively, up to 6.07%, 5.19% and 4.87%.**

I. Introduction

After the newly published High Efficiency Video Coding standard (HEVC) [1], the Joint Video Exploration Team (JVET) of ITU-T and ISO/IEC MEPG was established to explore the next video coding standard on Joint Exploration Model (JEM), which provides about 28.5% BD-rate saving compared with HEVC reference software (HM-16.6) in Random Access (RA) [2].

Block partitioning structure is the basis of hybrid framework. In HEVC [3], a coding tree unit (CTU) can be split into coding units (CU) by quadtree structure (QT). Further, a CU can be split to one or more prediction units (PU) for Motion Estimation (ME). Similar to the derivation method of PU, transform units (TU) are derived rooted from CU. However, more flexible quadtree plus binary tree block structure (QTBT), one of new coding tools in JEM, removes the concepts of CU, PU and TU. In the other words, CU, PU and TU are the same block. What's more, CU can be square or non-square.

As a basic tool of hybrid coding framework, transform module plays a vital role in compressing prediction residue. Due to the difference of intra and inter prediction in essence, the corresponding residual characteristics vary widely, and this paper mainly focuses on transform for inter residue. Based on HM-16.6, JEM introduces many new coding tools for many important modules, including transform module [4]. Three main points associated with inter transform are newly introduced. Firstly, the higher-frequency transform coefficients are forced to zeros, and only the lower-frequency transform coefficients are coded for saving BD-rate. Secondly, Enhanced Multiple core Transform (EMT) introduces a block-level flag to indicate whether using the classical DCT2 or not. If not, additional two bits are signaled to choose transform cores from pre-defined transform modes for horizontal and vertical transform, respectively [5]. Thirdly, Signal Dependent Transform (SDT) obtains a better coding efficiency by training Karhunen-Loèove transform (KLT) cores based on similar blocks [6]. Compared to general transform process, the process of deriving KLT cores and searching similar blocks in transform module are both additional processes for encoder and decoder, which brings extremely high complexity. Thus, SDT is closed by default [4].

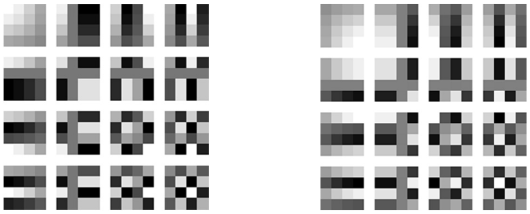

(a) DCT8 (b) DST7

Fig. 1. 2D transform bases for 4×4 DCT8 and DST7.

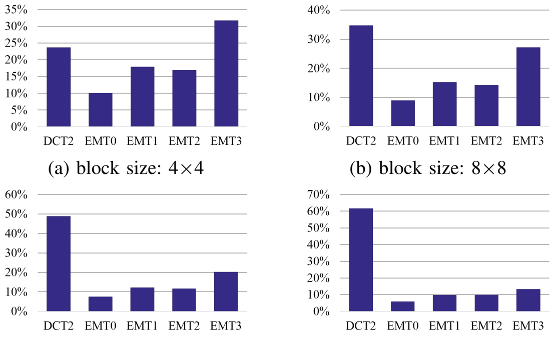
As far as the situation of call for proposals for the next video coding standard, the concept of more transform sets or transform modes than EMT is introduced by multiple proponents [7].

In [8], it is observed that there remains much texture information in inter residual, and the texture information is similar to the texture information in prediction image, so the eigenvector matrices, which are derived by applying Singular Value Decomposition (SVD) on the corresponding prediction block, are used as the transform cores for residual block.

Motivated by observation result of [9], which indicates the magnitude of inter residue is bigger closer to PU boundary, [5] chooses two transform cores (DCT8, DST7) from the discrete sinusoidal transform families to accommodate this distribution. The 2D transform bases of DCT8 and DST7 are showed in Fig. 1. It can be seen that DCT8 and and DST7 show gradually decreasing and increasing distribution along the transform direction, respectively. Thus, a combination of DCT8 and DST7 can accommodate the residual characteristic. However, EMT ignores the distribution, whose prediction error of four corners is bigger than inner area. In addition, EMT

TABLE I
THE SYNTAX DESCRIPTION OF EMT FOR INTER, AND THE SYMBOL '-' DENOTES THE SYNTAX FLAG WOULD NOT BE CODED.

EMT flag	**EMT index**	**Luminance**	
		HorT	**VerT**
0	-	DCT-II	DCT-II
1	0	DCT-VIII	DCT-VIII
	1	DST-VII	DCT-VIII
	2	DCT-VIII	DST-VII
	3	DST-VII	DST-VII



(a) block size: 4×4

(b) block size: 8×8

(c) block size: 16×16

(d) block size: 32×32

Fig. 2. The ratios of different transform modes derived at decoder side.

only aims at luminance component (Y), and it costs more bits to signal EMT flag and index leading to a larger loss in chrominance components (U and V). Based on the analysis above, TPBMT is proposed to refine EMT.

The remaining part of this paper is organized as follows. In the Section 2, a brief review of EMT and SVD are presented. The process and syntax description of the proposed method TPBMT are introduced in Section 3. Then, Section 4 shows the experimental results. Finally, Section 5 draws the finally conclusions of this paper.

II. Overview for EMT and SVD

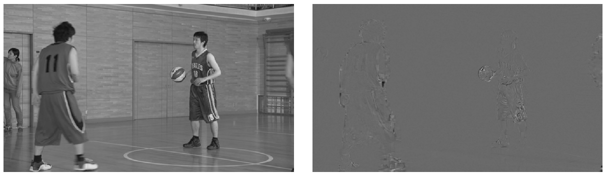
In this section, the process of EMT and SVD are briefly presented. Residual block is generally transformed by DCT2 in the past video coding standards. However, the characteristic of residual distribution is diverse, so DCT2 is difficult to well accommodate this residual characteristic [5]. Under the background, EMT is proposed, and it is proved to be efficient. Furthermore, SVD reduces the redundant information in residual by utilizing the similar texture information in the corresponding prediction block.

A. Overview for EMT

In general, a M×N residual block is transformed by horizontal transform (HorT) and vertical transform (VerT).

$$Y_{M\times N} = C_{M\times M} \times X_{M\times N} \times R_{N\times N} \tag{1}$$

where $X_{M\times N}$ denotes a $M \times N$ residual block. $C_{M\times M}$ and $R_{N\times N}$ denote the transform matrices of vertical and horizontal transform, respectively.



(a) Prediction image

(b) Residual image

Fig. 3. The similarity between residual image and prediction image, 2nd frame of *BasketballPass* (416×240), where QP is set as 0.

When the size of residual block is less than 64, whether EMT is used or not is signaled by one bit, named EMT flag. If EMT flag is zero, horizontal and vertical transforms are both implemented by classical DCT2 transform. If not, four transform modes would be tried to transform the residual block, and the index of best transform mode would be coded in bitstream. The operations above are only for luminance component, chrominance components would still use DCT2 as the transform mode.

The criterion for finding the best transform mode is rate-distortion optimization criterion (RDO) [10]. The value of RD cost is influenced by bits cost (B) and the sum of squared error (D) between the reconstructed block and the original block, then the RD cost (J) is calculated as:

$$J = D + \lambda \cdot B \tag{2}$$

where λ is the Lagrangian multiplier.

Table I shows the syntax of EMT. Encoder chooses the best transform mode by RDO, and the index of best transform mode would be coded in bitstream as Table I. Fig. 2 shows the ratios of different transform modes. We can see the ratios of four newly-introduced transform modes differ widely. As far as EMT is concerned, The ratio of index 3 is the largest, and the values of index 1 and 2 are nearly equivalent, but the ratio of index 1 is the smallest. The main reasons may be that the reference information is mainly derived from the left and above directions. Thus, this side information signaled method is not the best way.

B. Overview for SVD

As shown in Fig. 3, prediction image and residual image are derived from HEVC reference software HM10.0 [11]. We can find that residual image is similar to prediction image with regard to texture information, especially in complex motion region. Current motion estimation (ME) methods, which mainly adopt 2-dimensional estimation, are difficult to express the complex motion in reality.

For an inter prediction block (P), SVD puts a singular value decomposition on P, as in (3). Then, the singular matrixes U and V, namely the transform cores, can be obtained.

$$P = U\Sigma V^T \tag{3}$$

where Σ is a diagonal matrix. U and V are both orthogonal matrixes.

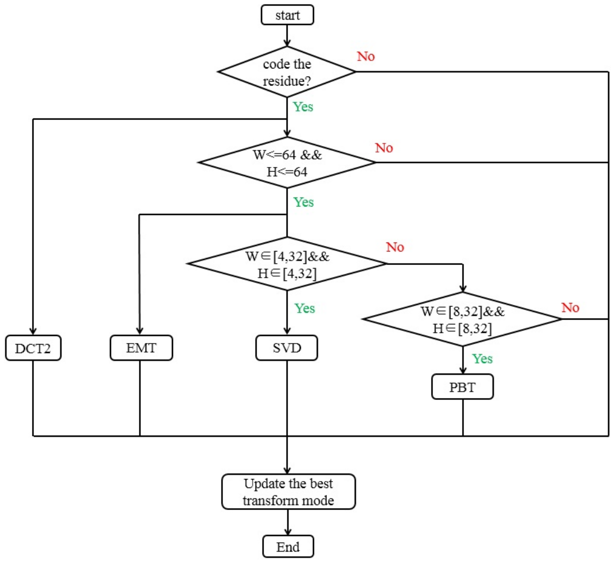


Fig. 4. The flowchart of TPBMT.

Then, for the inter residual block R, transform can be implemented as in (4).

$$Y = U^T R V \tag{4}$$

where Y is coefficient matrix.

Correspondingly, the inverse transform can be implemented as in (5).

$$R' = U Y' V^T \tag{5}$$

where Y' and R' denote reconstructed Y and reconstructed R, respectively.

In [8], traditional transform DCT2 is retained, because DCT2 and SVD both can not express the residual distribution independently. The optimal transform mode is decided by RDO, so 1-bit flag needs to be signaled. Signaling 1-bit flag for each TU is better to transform efficiently, but the cost of side information can not be neglected. Further, [8] has proved that the performance of signaling in CU level is more better than TU level. In the other words, whether all the TUs in one CU are transformed by SVD or not is decided by the same 1-bit flag. It is worth mentioning that only the inter square blocks are possible to be transformed by SVD.

III. The Proposed Method TPBMT

In this section, we will present the detail about the proposed method at encoder and decoder, respectively. Based on the characteristic of residual distribution, position based transform (PBT) splits the residual block to four sub-blocks, and fixes the transform mode by the position of sub-block. Put another way, the side information to indicate the transform mode can be saved by fixing the transform mode. In addition, SVD is extended to be applied for both square and non-square blocks.

TABLE II
The syntax description of transform, and the symbol '-' denotes the syntax flag would not be coded.

PBT flag	EMT flag	SVD flag	EMT index	Luminance	
				HorT	**VerT**
0	0	-	-	DCT-II	DCT-II
	1	0	0	DCT-VIII	DCT-VIII
			1	DST-VII	DCT-VIII
			2	DCT-VIII	DST-VII
			3	DST-VII	DST-VII
		1	-	SVD	SVD
1	-	-	-	PBT	PBT

A. Encoder Process

After inter prediction, the residual block is obtained. When the CU needs to transform the residual information, DCT2, EMT, SVD and PBT are all possible to be implemented for current residual block, and the best transform mode would be derived by RDO process. Meanwhile, four RD costs of different transform modes are calculated. Finally, the best transform mode, whose RD cost is the smallest, will be chosen, and the encoding result will also be saved. The RDO process of different transform modes is shown as Fig. 4. Furthermore, PBT only focuses on the inter residual block, whose width and height are both between 8 and 32, including square block and non-square block. Similarly, SVD only focuses on the inter residual block, whose width and height are both between 4 and 32, including square block and non-square block. Motivated by [12], SVD also performs same transform as luminance component for chrominance component, when the size of chrominance residual block is larger than 2.

When PBT is chosen, the residual block will be split to four small blocks. As shown in Fig. 5, a CTU, whose size is 64×64, is split by QTBT. The black solid line indicates the CU boundary, and the red dotted line indicates the split line by PBT. Besides, 0 sub-block to 3 sub-block are the four sub-blocks after splitting the CU, and the encoding order is along 0 to 3. Each RDO process is similar with general RDO process for each sub-block. In addition, two corresponding chrominance residual sub-blocks are also split like luminance residual sub-block, except the chrominance residual block whose size is less than 8. The transform modes of 0 to 3 sub-blocks are same with the transform modes of EMT index 0 to 3 showed in Table I, correspondingly.

At the encode side, 1-bit PBT flag need to be coded in bitstream, which decides the best transform mode. When PBT is checked for the best mode, PBT flag would be one. Otherwise, it would be zero and EMT flag is further signaled. More details about the syntax information are described in Table II.

B. Decoder Process

At the decode side, PBT flag will be parsed before inverse transform process. When PBT flag is one, PBT inverse transform process would be implemented. Otherwise, EMT flag needs to be further parsed to select DCT2 or not.

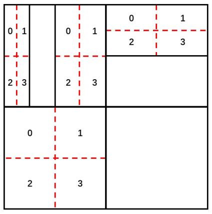
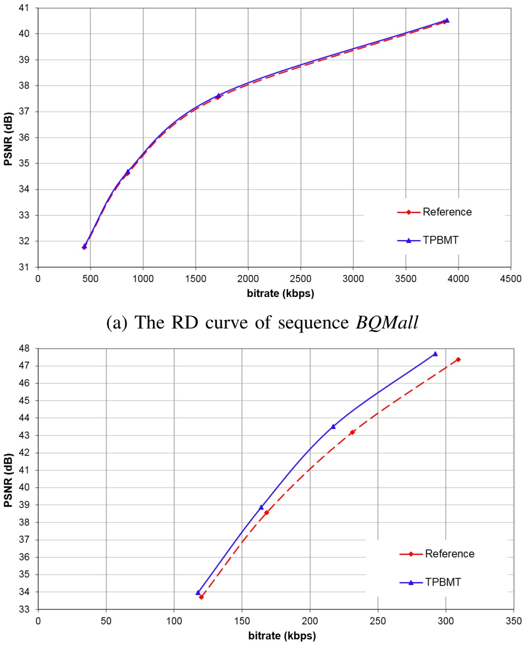


Fig. 5. A splitting example of PBT.

TABLE III
THE CLOSED CODING TOOLS OF THE ANCHOR FOR EXPERIMENT, THE ABBREVIATIONS CAN BE SEARCHED IN JEM7.0.

Closed coding tools list			
ATMVP	OBMC	IMV	FURC
IlluCompEnable	ALF	IntraFourTapFilter	IntraBoundaryFilter
LMChroma	BIO	DMVR	PDPC
NSST	AFFINE	AClip	BilateralFilter
EMT Intra	EMT Fast		



(a) The RD curve of sequence *BQMall*

(b) The RD curve of sequence *SlideEditing*

Fig. 6. RD curves of Luminance components in LB case.

When PBT is one, the coefficients parsed processes of sub-blocks are along 0 to 3 order, same with the order in Fig. 5. Then, the current residual block can be reconstructed by reconstructing four residual sub-blocks in general way.

IV. Experimental Results

To verify the performance of the proposed method, we implement the method on the latest reference software JEM7.0 [13] released by JVET and evaluate it under common test conditions (CTC) [14]. For verifying the efficiency quickly, some new coding tools having less connection with inter transform are closed, which are all listed in Table III.

TABLE IV
THE BD-RATE (%) RESULTS OF EMT+SVD AND THE PROPOSED METHOD TPBMT COMPARED TO JEM7.0 (LB) WITH SOME CODING TOOLS OFF (REFERENCE), AND THE MINUS SIGN DENOTES BD RATE SAVING.

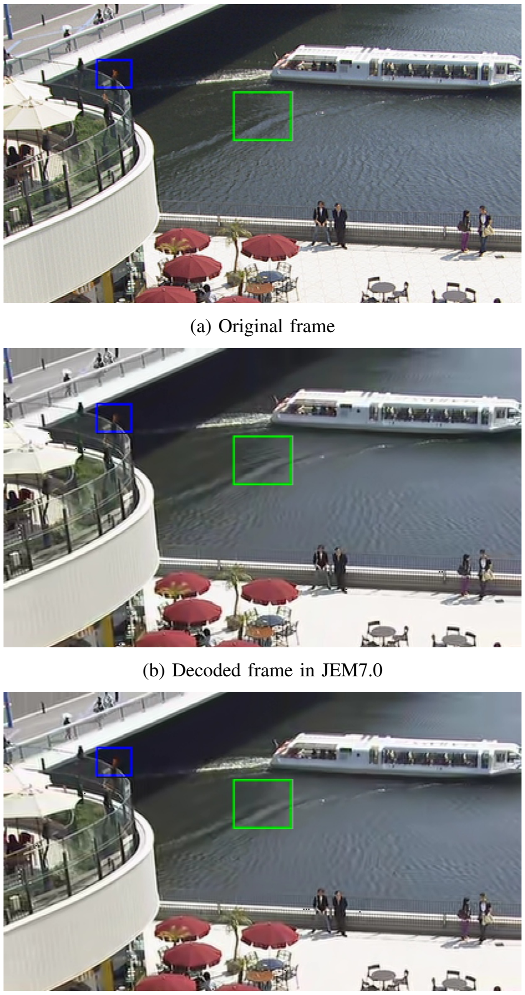
Test Sequence	EMT+SVD			TPBMT		
	Y(%)	U(%)	V(%)	Y(%)	U(%)	V(%)
Kimono	+0.07	−1.89	−2.40	−0.24	−2.44	−2.57
ParkScene	−0.34	−1.90	−0.89	−0.84	−3.24	−2.28
Cactus	−0.41	−1.63	−1.47	−0.54	−1.89	−1.51
BasketballDrive	−0.51	−0.73	−0.75	−0.67	−1.10	−1.02
BQTerrace	−0.76	−0.22	+1.21	−0.92	−2.31	−1.38
average	**−0.39**	**−1.27**	**−0.86**	**−0.64**	**−2.20**	**−1.75**
BasketballDrill	−0.34	−1.57	−1.39	−0.56	−1.33	−1.56
BQMall	−1.15	−0.32	−0.42	−1.35	−0.39	−1.17
PartyScene	−0.75	−0.56	−0.17	−0.96	−1.25	−1.30
RaceHorses	−0.24	−1.37	−0.84	−0.45	−2.28	−1.82
average	**−0.62**	**−0.96**	**−0.70**	**−0.83**	**−1.31**	**−1.46**
BasketballPass	−0.63	−0.22	−0.37	−0.81	−0.06	−0.68
BQSquare	−0.66	+1.57	+1.17	−0.87	−3.29	−4.27
BlowingBubbles	−0.57	−0.64	−1.12	−0.76	−1.29	−1.58
RaceHorses	−0.07	−1.20	−0.49	−0.26	−1.33	−1.23
average	**−0.48**	**−0.12**	**−0.20**	**−0.68**	**−1.49**	**−1.94**
FourPeople	−0.65	−0.11	−0.92	−0.70	−0.62	−1.00
Johnny	−1.10	+0.09	+0.62	−1.15	−1.18	−0.99
KristenAndSara	−0.31	−0.50	−0.54	−0.52	−0.17	−0.99
average	**−0.69**	**−0.17**	**−0.28**	**−0.79**	**−0.66**	**−0.99**
BasketballDrillText	−1.23	−1.80	−1.41	−1.26	−2.11	−1.60
ChinaSpeed	−1.49	+0.28	+0.01	−1.63	−0.33	+0.04
SlideEditing	−5.92	−4.74	−4.62	−6.07	−5.19	−4.87
SlideShow	−4.61	−1.26	−2.08	−4.57	−1.31	−0.59
average	**−3.31**	**−1.88**	**−2.02**	**−3.38**	**−2.23**	**−1.75**
Overall average	**−1.08**	**−0.94**	**−0.84**	**−1.26**	**−1.66**	**−1.62**
Enc. Time	116%			123%		
Dec. Time	103%			103%		

The objective performance is evaluated by BD-rate [15], and the final experimental results of low delay B (LB) case are provided in Table IV. EMT+SVD denotes that PBT is not implemented. Compared with EMT+SVD, the BD-rate reductions are achieved for all test sequences. Especially for chrominance, up to 4.86% and 5.44% gains can be seen for sequence *BQSquare*. From the Table IV, we can see the average gains of the proposed method TPBMT are 1.26%, 1.66% and 1.62% for Y, U and V, respectively. By assigning specific transform mode for four sub-blocks, it obtains better performance for luminance component, compared with EMT. The gains of Y, U and V can reach up to 6.07%, 5.19% and 4.87%, respectively, for screen content sequence *SlideEditing*. Fig. 6 shows the PSNR-Rate curves of *BQMall* and *SlideEditing* between reference and TPEMT, and it can be seen that TPBMT achieve BD-rate saving through full QP range for these two sequences. For the reasons of side information reduction led by PBT, considerable gains are also observed on chrominance components.

The subjective quality is a key evaluation parameter for a coding tool. The result of subjective quality comparison is show in Fig. 7. The blue and green directional contours are two areas for comparison. In Fig. 7(b), the area in green directional contour shows block artifacts, and the reconstructed quality of the area in blue directional contour is poor than Fig. 7(a) and Fig. 7(c).

V. Conclusion

Due to the characteristics of inter frame coding, there are plentiful residual blocks, whose prediction error of four



(a) Original frame

(b) Decoded frame in JEM7.0

(c) Decoded frame in JEM7.0 with TPBMT

Fig. 7. The result of subjective quality comparison, 530th frame of BQSquare_416×240 with QP32 in LB case.

corners is bigger than inner area. In the one hand, based on the similarity of texture between prediction block and residual block, SVD can reduce the redundant information in residue by utilizing the corresponding prediction blcok. In the other hand, based on position in a block, the better transform core can be fixed in order to improve transform efficieny and decrease side information cost. Combined SVD and PBT, a more efficient multiple transform method TPBMT can be obtained. Experimental results show that TPBMT is more efficient than EMT, and the subjective quality is also improved. What's more, the complexity of the proposed method at the decoder is perfectly tolerable, which is beneficial for fast decoding.

Acknowledgment

This work is partially supported by the National project of 863 2015AA015903, Key project from the Tsinghua National Laboratory for Information Science and Technology, and the Beijing National Research Center for Information Science and Technology.

References

[1] "High Efficiency Video Coding," in *Rec. ITU-T H.265 and ISO/IEC 23008-2*, Dec. 2016.

[2] Xiang Li and Karsten Suehring, "Report of AHG3 on JEM software development," in *JVET-I0003*, Gwangju, KR, Jan. 2018.

[3] Sullivan, Gary J., and et al, "Overview of the high efficiency video coding (HEVC) standard," in *IEEE Transactions on circuits and systems for video technology*, 22.12 (2012), pp. 1649-1668.

[4] Jianle Chen, Elena Alshina, Gary J. Sullivan, JensRainer Ohm, and Jill Boyce, "Algorithm Description of Joint Exploration Test Model 7 (JEM 7)," in *JVET-G1001*, Torino, IT, July 2017.

[5] Zhao X, Chen J, Karczewicz M, and et al, "Enhanced Multiple Transform for Video Coding," in *Data Compression Conference (DCC)*, 2016.

[6] Lan C, Xu J, Zeng W, and et al, "Variable Block-Sized Signal Dependent Transform for Video Coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[7] A. Said, and X. Zhao, "Description of Core Experiment 6: Transforms and transform signalling," in *JVET-J1026*, San Diego, Apr. 2018.

[8] X. Cao, and Y. He, "Singular vector decomposition based adaptive transform for motion compensation residuals," in *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 4127-4131, 2014.

[9] J. An, X. Zhao, X. Guo, and S. Lei, "Non-CE7: Boundary-Dependent Transform for Inter-Predicted Residue," in *JCTVC-G281*, Geneva, CH., November 2011.

[10] Gary J. Sullivan and Thomas Wiegand, "Rate-distortion optimization for video compression," in *IEEE signal processing magazine*, 1998, pp. 74-90.

[11] HM-10.0 [Online]. "https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags," 2013.

[12] T. Tsukuba, O. Nakagami, M. Ikeda, and T. Suzuki. "On Adaptive Multiple Core Transform for Chroma," *JVET-E0036*, Jan., 2017.

[13] HM-16.6-JEM-7.0, "https://jvet.hhi.fraunhofer.de/trac/vvc/browser/jem#tags," 2017.

[14] K. Suehring and X. Li., "JVET common test conditions and software reference configurations," in *JVET-G1010*, Torino, IT, July 2017.

[15] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," in *ITU-T SC16/Q6*, VCEGM33, Austin, USA, Apr. 2001.