Mode Dependent Coding Tools for Video Coding

Siwei Ma, Shiqi Wang, Qin Yu, Junjun Si, and Wen Gao, *Fellow, IEEE*

***Abstract*—This paper provides an overview of the mode dependent coding tools in the development of video coding technology. In video coding, the prediction mode is closely related with the local image statistical characteristics. For instance, the angular intra mode contains the oriented structure information and the interpartition mode can reveal the local texture properties. Characterized by this fact, the prediction mode can be employed to facilitate the further encoding process, such as transform and coefficient coding. Recently, many mode dependent coding tools were proposed to provide more flexibility and thereby improve the compression performance. In general, these coding tools can be classified into three categories: mode dependent transform, residual reorder before transformation and coefficient scan after transformation. In this paper, these advanced coding algorithms are detailed and experiments are conducted to demonstrate their superior compression performance.**

***Index Terms*—Intra prediction, mode dependent transform, residual reorder, transform coefficient scan, video coding.**

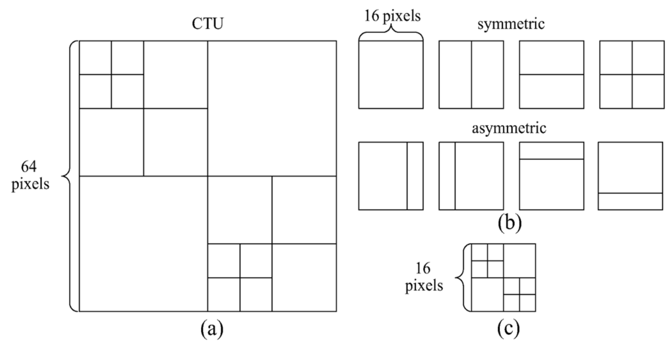


Fig. 1. Example of a 64 $\times$ 64 CTU, and its associated CU, PU and TU partitions. (a) a 64 $\times$ 64 CTU that is divided into CUs. (b) partition modes for PU specified in HEVC. (c) residual quadtree for TU partitions.

I. Introduction

THE series of video compression standards, such as MPEG-2 [1], H.263 [2], MPEG-4 Visual [3] and H.264/AVC [4], reflect the technological track of video compression in various applications. Recently, as the number of high definition (HD) videos generated every day is increasing dramatically, high efficiency compression of HD video is highly desired. The high efficiency video coding (HEVC) standard [5], [6], which is the joint video project by the Moving Picture Experts Group (MPEG) and Video Coding Expert Group (VCEG), has achieved significant improvement in coding efficiency compared to H.264/AVC with more than 50 percent bit rate saving in terms of perceptual quality [7].

In HEVC, many new coding tools are adopted to improve the compression performance. As shown in Fig. 1, an adaptive quadtree structure based on the coding tree unit (CTU) is employed, in which three new concepts termed as coding unit (CU), prediction unit (PU) and transform unit (TU) [8] are introduced to specify the basic processing unit of coding, prediction and transform. CTU is employed as the root of the coding tree and each leaf of the quadtree is called CU. Therefore, one CTU can be partitioned into multiple CUs and each CU is identified with one coding category: inter CU or intra CU. CU can be further split into one, two or four PUs to specify the prediction information. The residual block of each CU can be transformed with a quadtree structure, which is usually called residual quadtree (RQT), and the transform is performed on each leaf node of this quadtree.

Up to and including the H.264/AVC video coding standard, the prediction mode information has not been fully exploited in the following transform and coefficient scan. In H.264/AVC, a scaled integer transform, as an approximation to the Discrete Cosine Transform (DCT), is performed followed by the subsequent procedures including quantization and coefficient coding. This process doesn't fully consider the local image characteristics and cannot adapt to the prediction mode, which will provide valuable information to assist the subsequent transform coding. Being aware of the importance of the mode dependent coding, many research works have been conducted.

Since the local image statistical properties can be inferred from the intra coding modes, many efforts have been devoted to further promote the transform efficiency for intra coding. A noteworthy advance in this research field is the exploration of new directional transforms [9]–[13]. The motivation behind them lies in that for the sources with arbitrarily directed edges, the 2-D DCT may become inefficient for energy compaction. Inspired by the observation that the residual samples present different statistical properties for different intra prediction modes, mode-dependent directional transform (MDDT) scheme was thereby presented in [10]. Recently, to further improve the performance of MDDT, many directional transform methods were proposed, such as mode dependent sparse transform (MDST) [11], rate-distortion optimized transform (RDOT) [12] and direction adaptive residual transform (DART) [13]. To simplify the MDDT, orthogonal mode dependent directional transform (OMDDT) and mode dependent DCT

Manuscript received January 30, 2013; revised May 09, 2013; accepted July 16, 2013. Date of publication July 26, 2013; date of current version November 18, 2013. This work was supported in part by the Major State Basic Research Development Program of China (973 Program, 2009CB320900), the National Science Foundation (61121002 and 61103088), the National High-tech R&D Program of China (863 Program, SS2012AA010805), and the National Key Technology R&D Program (2011BAH08B01). The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Marek Domański.

The authors are with the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: swma@pku.edu.cn; sqwang@pku.edu.cn; qyu@pku.edu.cn; jjsi@pku.edu.cn; wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2013.2274951

(MDDCT) were also proposed in [14]. In the development of HEVC, MDDT was further investigated in terms of both coding efficiency and implementation complexity [15]–[18]. However, the directional transform schemes such as MDDT usually requires many transform matrices at a particular size, which may not be hardware friendly. To further reduce the number of the transform matrices, in [55], [56], the residuals between prediction and transform stages are spatially reordered so that the distribution statistics of reordered residual samples present less mode-dependent characteristic. Then the matrices for different intra modes can be merged together and thus the implementation complexity in MDDT is reduced.

Another motivation in the advances of HEVC is to derive the discrete sine transform (DST) [22]–[26] with performance close to the optimal Karhunen-Loeve Transform (KLT) [21]. In [22], [23], Han *et al.* derived that for intra prediction residuals with horizontal and vertical modes, DST is better than DCT under the assumption of a separable first order Gauss-Markov model. In [27], it is further illustrated that for the oblique modes, the combination of DCT-II and DST-VII is the optimal transform. Due to the superior coding performance and low complexity of DST, it was studied in HEVC project [28]–[34] and finally adopted [32].

A further advance to the mode dependent DCT/DST is the secondary transform scheme. In [39], the rotational transform (ROT) was proposed based on the observation that after the primary transform, the coefficients still preserve directional information and can be further processed by rotations. In [37], [38], another secondary transform scheme was proposed. This is the first attempt to apply secondary transform in both intra and inter coding, which employs both the PU and TU information to facilitate the subsequent coding process. In this scheme, the lower 4×4 or 8×8 frequency coefficients of 8×8 or larger block size DCT are extracted and further transformed with a trained matrix. Compared to larger size DCT/DST, lower complexity and higher coding efficiency are obtained. In HEVC these tools have also been extensively studied [40]–[46].

For inter coding, the transform shape is closely related to prediction shape, and for the non-square prediction, non-square transform has been extensively studied in the previous coding standards [47]–[49]. In the course of HEVC project, the non-square quadtree transform (NSQT) scheme has also been proposed to benefit the various inter prediction shapes [50], [51].

Recently, efforts have also been devoted to mode dependent entropy coding by optimizing the coefficient scan order [57]–[59]. In the literature, mode dependent coefficient scan (MDCS) schemes mainly concentrate on scanning the coefficients in the order of decreasing energy/variance, and algorithms that lead to the optimal order with the aid of the intra prediction modes have been extensively studied. In HEVC, the mode dependent scan algorithms have also been investigated for intra coding [60]–[67].

In this paper, we present an overview of the mode dependent coding tools, including the design philosophy as well as the algorithm details. The rest of the paper is organized as follows. In Section II, the mode dependent transform algorithms are discussed. In Section III, the mode dependent residual reordering scheme is introduced and its connections with MDDT are detailed. In Section IV, we introduce the mode dependent coefficient scan. Performance analyses of these coding schemes are provided in Section V, and Section VI concludes the paper.

$$\begin{bmatrix} 4.47 & 4.43 & 4.44 & 4.59 \\ 5.87 & 5.65 & 5.75 & 6.09 \\ 6.83 & 6.63 & 6.82 & 7.16 \\ 7.77 & 7.64 & 7.82 & 8.14 \end{bmatrix}, \begin{bmatrix} 4.35 & 5.92 & 6.72 & 7.51 \\ 4.31 & 5.83 & 6.65 & 7.75 \\ 4.46 & 6.06 & 6.93 & 7.89 \\ 4.47 & 6.17 & 7.06 & 8.13 \end{bmatrix}, \begin{bmatrix} 5.68 & 5.79 & 5.99 & 6.38 \\ 5.99 & 5.85 & 5.97 & 6.38 \\ 6.18 & 6.10 & 6.22 & 6.72 \\ 6.55 & 6.48 & 6.52 & 7.20 \end{bmatrix}$$

(a) Mode 0 (b) Mode 1 (c) Mode 2

$$\begin{bmatrix} 5.05 & 4.72 & 4.69 & 4.61 \\ 6.55 & 6.17 & 6.00 & 5.95 \\ 6.71 & 6.77 & 6.63 & 6.59 \\ 7.33 & 7.33 & 7.34 & 7.57 \end{bmatrix}, \begin{bmatrix} 4.45 & 5.08 & 5.19 & 5.59 \\ 4.90 & 6.19 & 6.46 & 7.12 \\ 4.99 & 6.59 & 7.18 & 7.83 \\ 5.29 & 6.93 & 7.73 & 8.48 \end{bmatrix}, \begin{bmatrix} 4.07 & 4.48 & 4.69 & 4.97 \\ 5.03 & 5.85 & 6.33 & 6.98 \\ 5.52 & 6.70 & 7.24 & 8.11 \\ 5.86 & 7.26 & 8.07 & 8.86 \end{bmatrix}$$

(d) Mode 3 (e) Mode 4 (f) Mode 5

$$\begin{bmatrix} 4.21 & 5.29 & 5.57 & 5.91 \\ 4.45 & 6.01 & 6.51 & 6.96 \\ 4.61 & 6.22 & 7.10 & 7.81 \\ 5.03 & 7.02 & 8.29 & 9.04 \end{bmatrix}, \begin{bmatrix} 4.86 & 4.64 & 4.55 & 4.48 \\ 6.53 & 6.39 & 6.00 & 6.10 \\ 6.94 & 6.82 & 6.42 & 6.91 \\ 7.50 & 7.19 & 7.19 & 7.48 \end{bmatrix}, \begin{bmatrix} 5.06 & 6.64 & 7.01 & 7.55 \\ 4.56 & 5.96 & 6.52 & 6.98 \\ 4.35 & 5.62 & 6.38 & 7.23 \\ 4.66 & 6.33 & 7.23 & 7.92 \end{bmatrix}$$

(g) Mode 6 (h) Mode 7 (i) Mode 8

Fig. 2. Normalized distributions of absolute residual magnitudes in 4×4 intra prediction modes.

II. Mode Dependent Transform

In this section, we will introduce the development of mode dependent transform schemes in video coding. We start with the training based directional transform; including MDDT, which is the fundamental work for mode dependent intra coding; MDST, which applies the L_0-norm regularization to robustly estimate the vertical and horizontal transform; and RDOT, which is an extension of MDDT by providing more transform matrixes for selection. Secondly, we discuss the DART and DST, which simplify the training based directional transform but achieve high coding efficiency. Thirdly, we introduce the secondary transform schemes, which further improve the coding performance significantly. Finally, the idea of the inter partition mode dependent non-square quadtree transform is discussed.

A. Training Based Directional Transform

1) Mode Dependent Directional Transform: The conventional 2-D DCT is implemented separately by performing 1-D DCT twice, horizontally and vertically. Naturally, this separable 2-D transform is capable to capture the signal correlation along either the horizontal or vertical direction. However, the conventional 2-D DCT would be sub-optimal when there exist directional edges [9]. In intra coding, for different intra modes, the energy of residual samples is distributed differently within the region of a single residual block. The normalized absolute residual magnitudes in 4×4 intra modes are shown in Fig. 2. It is observed that the sample is distributed differently and there is significant directional information within the residual block. This illustrates that the DCT may no longer approximate the optimum transform KLT for directional intra prediction residual. Moreover, it is also observed that residual samples present distinguishable distribution characteristics for different intra prediction modes. Therefore, mode-dependent directional transform [10] was proposed by assigning each intramode a distinct transform matrix,

$$F = C_i \cdot Y \cdot R_i \tag{1}$$

where Y indicates the residual block. C_i and R_i are the vertical and horizontal transform functions for intra mode mode i, respectively. F denotes the resulting transform coefficient matrix. The vertical and horizontal transform matrices are trained off-line with the collected residual blocks actually generated by prediction mode i. However, storing multiple transform matrices may increase implementation complexities to both the encoder and decoder. Moreover, as the DST scheme has achieved comparable coding performance to MDDT and the implementation complexity of DST is much lower, the MDDT scheme has not been adopted in HEVC.

2) Mode Dependent Sparse Transform: In MDDT, KLT-based learning process is employed, which is prone to outliers and may result in arbitrarily principal component directions. To address this issue, a new algorithm of mode-dependent transform for intra coding was proposed in [11]. The proposed algorithm, termed as mode dependent sparse transform, employs the L_0-norm regularization as a more robust way to learn the 2-D transforms. The sparsity based cost function is defined as follows,

$$\min_{C_i,R_i}\left(\sum_{j\in S_i}\min_{F_i^j}\left\|Y_i^j - C_i^T F_i^j R_i^T\right\|_2^2 + \lambda\left\|F_i^j\right\|_0\right)$$
$$s.t.\ C_i^T C_i = I,\quad R_i^T R_i = I. \tag{2}$$

where $\|\bullet\|_0$ denotes the L_0-norm counting the number of nonzero coefficients in the vector. S_i represents the training set over which the cost function is minimized and I represent the identity matrix. To learn the optimal vertical and horizontal transforms for each intra prediction mode i, the transform matrices are firstly initialized as the traditional 2-D DCT. Then an iterative method is proposed by alternating between the calculating of transform coefficients and an update process of the transform matrices to better fit the data.

The transform coefficients with MDST may exhibit different energy distributions compared to the 2-D DCT, and this may affect the subsequent entropy coding process. Therefore, the columns of the vertical and horizontal transform are reordered depending on the energy of the coefficient values in the training set. Compared with the KLT-based training method, MDST can achieve better performance when outliers exist in the training data.

3) Rate-Distortion-Optimized Transform: In MDDT, motivated by the observation that each intra mode has distinct directional information statistically, one transform matrix is employed for each intra mode. Although this requires no additional signaling information or rate distortion search, within each intra mode, one transform matrix may not be able to accommodate the potential variability of the residual characteristics because of the diversity of image content.

Several intra- and inter-predicted residual blocks are presented in Fig. 3. All the 4 × 4 blocks in Fig. 3(a) are actual residual blocks obtained from vertical intra prediction. It is observed that the left block presents a strong vertical edge. However, the other two blocks present irregular textures and contain both vertical and horizontal edges. It is the same case for the horizontal inter-predicted residual blocks shown in

$$\begin{bmatrix} 26 & 11 & -44 & -84 \\ 22 & 19 & -45 & -85 \\ 29 & 6 & -43 & -84 \\ 40 & 26 & -43 & -86 \end{bmatrix}\begin{bmatrix} 1 & 1 & -34 & -22 \\ 6 & -2 & -78 & -5 \\ 14 & -29 & -107 & 52 \\ 19 & -53 & -115 & 91 \end{bmatrix}\begin{bmatrix} 16 & 45 & 3 & -10 \\ 40 & 59 & 1 & -12 \\ 27 & 27 & -2 & -8 \\ 10 & -22 & -22 & -1 \end{bmatrix}$$

(a) Vertical intra-predicted residual blocks

$$\begin{bmatrix} 17 & 19 & 19 & 19 \\ 5 & 5 & 7 & 11 \\ -34 & -40 & -34 & -40 \\ -32 & -28 & -30 & -30 \end{bmatrix}\begin{bmatrix} 2 & 30 & 35 & 33 \\ 2 & 0 & -5 & 7 \\ 12 & 18 & 21 & 29 \\ 16 & 36 & 50 & 53 \end{bmatrix}\begin{bmatrix} -7 & -3 & 23 & 0 \\ 11 & 9 & 21 & 10 \\ 11 & 13 & 30 & 33 \\ -45 & -34 & -43 & -30 \end{bmatrix}$$

(b) Horizontal intra-predicted residual blocks

$$\begin{bmatrix} 20 & 19 & 22 & 25 \\ 17 & 20 & 21 & 24 \\ 13 & 16 & 19 & 21 \\ 18 & 18 & 18 & 19 \end{bmatrix}\begin{bmatrix} -32 & -22 & -5 & 3 \\ -33 & -28 & -4 & 2 \\ -16 & -25 & -2 & 1 \\ -4 & -13 & 3 & 3 \end{bmatrix}\begin{bmatrix} -34 & -17 & -1 & 2 \\ -24 & -10 & 4 & -1 \\ -5 & -6 & 4 & -3 \\ 3 & 0 & 7 & 0 \end{bmatrix}$$

(c) Inter-predicted residual blocks

Fig. 3. Actual predicted residual blocks with the same prediction mode. (a) vertical intra prediction, (b) horizontal intra prediction, (c) inter prediction.

Fig. 3(b). For inter coding, the residual blocks also present distinctive characteristics, as shown in Fig. 3(c). This implies that multiple transform basis functions within the same coding mode can possibly further improve the compression performance. Inspired by this, a rate-distortion-optimized transform was proposed in our work [12], where the prediction residuals are transformed with different basis functions, and the best one is selected in terms of the rate-distortion (R-D) performance.

In RDOT, for each intra direction, there are distinct K pairs of vertical and horizontal transform candidates trained off-line, which produce totally K different transform paths for each intra mode. The encoder tries all the candidates and selects the optimal path with minimum R-D cost value. The transform indices are explicitly signaled in the bitstream. Compared with MDDT, RDOT further refines the transform by imposing both mode- and data-dependency, and thus better energy compaction can be achieved in the transform domain.

Though RDOT can extend the MDDT by providing more transform matrices, for each transform path, the encoder has to perform transform, quantization, entropy coding, de-quantization, inverse transform and reconstruction, which imposes high computational burden to the encoder. In light of this limitation, several fast RDOT schemes were also employed to collaboratively accelerate the encoding process [12]. One approach is to employ coding results of the DCT to skip the unnecessary RDOT trails. Specifically, DCT is implemented prior to RDOT, and if the rate-distortion (R-D) cost with DCT is lower than a threshold, the RDOT will be performed. This indicates that the RDOT will only be performed when DCT can also achieve a good coding performance, as the optimal DCT- and RDOT-based coding modes are highly correlated. The other approach is to apply the luminance coding speedup (LCS) technique in RDOT. In LCS, the luminance coding results for DC_PRED chroma mode are restored for the remaining modes.

B. Direction-Adaptive Residual Transform

The introduced MDDT, MDST and RDOT are mainly featured by the mode-dependent design of directional transform to improve the intra coding performance. However, they also introduce higher complexity as a result of the following issues:

1) Additional memory requirement for storing the transform matrices.
2) Offline training process is required.
3) More computational complexities might be introduced.

Unlike the training based approaches, inspired by the separable directional 2-D DCT in image coding, a new structure of DART was proposed in [13]. DART is consisted of a primary and a secondary transform. In the primary transform stage of DART, different 1-D transforms are employed along each oriented path in each direction. Then only the DC coefficients produced by the first stage are processed with DCT in the secondary transform stage. In some cases, short DCT paths are observed, and this may limit the performance of DCT. To address this issue, path folding is performed by combining pixels from neighboring scans. Compared with the existing KLT-based mode dependent directional transforms, this method is more flexible as no training is required to perform. The DART scheme has been implemented into the Key Technical Area (KTA) software and better coding performance is observed compared to MDDT.

C. Discrete Sine Transform

To achieve similar coding performance with lower complexity, Han *et al.* [22], [23] introduced a DST scheme for mode dependent intra coding. It is shown that under the first-order Gauss-Markov assumption, the DST is derived to closely match the actual KLT for intra-predicted residuals of horizontal and vertical modes. Assuming the first-order Gauss-Markov model for the image pixels,

$$x_k = \rho x_{k-1} + e_k \tag{3}$$

where ρ is the correlation coefficient between the pixels and e_k is a white-noise process with zero mean, $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]$ denote the random vector to be encoded. It is shown that the when the boundary value is available, autocorrelation matrix of the prediction residual y can be approximated to be

$$R_{yy} = (1-\rho^2)Q^{-1}Q^{-T} \tag{4}$$

where

$$P_1 = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & 0 & \cdots \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1+\rho^2-\rho \end{pmatrix} \approx Q^T Q \tag{5}$$

In this case, the KLT is explicitly calculated as a sinusoidal transform [19],

$$[T_S]_{i,j} = \frac{2}{\sqrt{2N+1}} \sin\frac{(2i-1)j\pi}{2N+1} \tag{6}$$

For the detailed mathematical derivation of DST to approximate the optimal KLT transform, the readers are referred to [22]–[27].

In HEVC, DST has also been extensively studied [28]–[34], including the coding efficiency as well as the implementation complexity issues, and the DCT/DST transform scheme for 4×4 intra luma blocks as described in [32] was finally adopted. In [34], the DST scheme was further simplified in HEVC by always employing DST in 4×4 luma intra coding. During the HEVC development, applying larger size DCT/DST scheme to larger size square blocks was also studied. However, given the high complexity of 8×8 to 32×32 DST [31] and limited coding gain, only luma 4×4 intra coding is performed with DST.

To employ the 4×4 DST in video coding, several low complexity integer DST schemes have been proposed. In [23], a low complexity integer approximation to the DST was introduced, and it is observed that the coding performance of integer DST is also very close to that of KLT. In [25], low complexity orthogonal 4-point DST was proposed where only a small number of bit-shifts and adds are needed. In HEVC standardization process, the complexity of DST has also been studied [31]–[33], such as bit precision, encoding and decoding time, etc. Finally, the DST scheme in [32] with fast implementation was adopted in HEVC. In [36], fast algorithms for DST-VII and DST-VI were presented, and the complexity of the proposed length 4 DST-VII factorization is close to the complexity of "Loeffler, Ligtenberg, and Moschytz" (LLM) factorization of DCT-II. Recently, the factorizations for fast joint computation of DST and DCT at different sizes were further derived in [35].

D. Secondary Transform

The secondary transform is applied after the primary transform such as DCT in order to get better energy compaction. The rotational based secondary transform ROT [39] was proposed and studied in HEVC [40], [41]. The rationale behind it lies in that allowing partial exchange of energy between columns/rows of transform coefficients matrix can achieve better energy compaction. However, in ROT the prediction mode information has not been fully exploited, and thereby choice from the transform dictionary is determined by R-D search and signaled in the bitstream.

For blocks of size 8×8 and larger, a different category of secondary transform was proposed for both intra and inter coding [37], [38], [42]–[46]. The secondary transform can be applied to the lower frequency (4×4 or 8×8) DCT coefficients at larger size of primary transform blocks. Therefore, much lower complexity is introduced compared to performing larger size DST. For intra coding, this secondary transform considers the directional information to avoid the search complexities and signaling bits. For inter coding, motivated by the observation that the energy of the prediction error is larger near PU boundary than in the middle of PU, TU partitions and positions are employed to adaptively decide the secondary transform process. The transform matrix is trained off-line and implemented in HEVC. Extensively experimental results show that with this secondary transform, significant coding gains can be achieved with negligible complexity overhead.

E. Non-Square Quadtree Transform

To solve the non-stationary problem in video coding, block partitioning technique has become a successful method in the video coding. To provide more flexibility to the block partition pattern for prediction, HEVC defines two kinds of prediction partitions for intra PUs and eight for inter PUs. Suppose the CU size to be $2\mathrm{N} \times 2\mathrm{N}$, in intra prediction, $\mathrm{PART\_2N} \times 2\mathrm{N}$ and $\mathrm{PART\_N} \times \mathrm{N}$ PU partitions are defined; and in inter prediction, two square partitions ($\mathrm{PART\_2N} \times 2\mathrm{N}$ and $\mathrm{PART\_N} \times \mathrm{N}$), two rectangular partitions ($\mathrm{PART\_2N} \times \mathrm{N}$ and $\mathrm{PART\_N} \times 2\mathrm{N}$) and four asymmetric motion partitions ($\mathrm{PART\_2N} \times \mathrm{nU}$, $\mathrm{PART\_2N} \times \mathrm{nD}$, $\mathrm{PART\_nL} \times 2\mathrm{N}$ and $\mathrm{PART\_nR} \times 2\mathrm{N}$) are supported. The asymmetric motion partitions (AMP) improve the coding efficiency for the irregular object boundaries, which cannot be successfully represented by symmetric partitions.

Although the AMP modes can improve the accuracy of motion compensation to some extent, the conventional square quadtree transform will cross the prediction boundaries, which may have negative impact on the final performance. Moreover, as horizontal and vertical frequencies often take dominated power of the directional energy in an image [53], vertical partition modes will be applied to the coding block with vertical texture and vice versa. This results in significant oriented structure information in residual signal, which can be employed to optimize the transform structure [54].

These motivations provide useful guidelines in designing the non-square quadtree transform (NSQT) scheme in HEVC [50]–[52]. The NSQT is in combination with the AMP scheme to get better coding efficiency. In view of the relationship between residual signal properties and partition modes, non-square transforms are applied at subdivision levels from the root of NSQT. The shape of non-square transform is correlated with the partition modes. For PU with size of $2\mathrm{N} \times \mathrm{N}$, $2\mathrm{N} \times \mathrm{nD}$ or $2\mathrm{N} \times \mathrm{nU}$, the $2\mathrm{N} \times 2\mathrm{N}$ transform block will be split into four $2\mathrm{N} \times 0.5\mathrm{N}$ transform blocks, while for prediction units with size of $\mathrm{N} \times 2\mathrm{N}$, $\mathrm{nR} \times 2\mathrm{N}$ and $\mathrm{nL} \times 2\mathrm{N}$, $0.5\mathrm{N} \times 2\mathrm{N}$ shaped transforms are performed on the prediction residual. The scan strategy of NSQT is also diagonal by partitioning the transform block into multiple 4×4 blocks, and diagonal scan is performed both within and between these blocks.

III. Mode Dependent Residual Reorder

Although the mode dependent transform such as MDDT further refine the transform for intra prediction residuals by applying non-orthogonal column and row transform matrices, the complexity is dramatically increased in terms of the number of employed transform functions. In this section, we introduce our work on the mode dependent residual reordering (MDRR) scheme [55], [56], which further simplify the MDDT scheme by reducing the number of transform matrices.

In MDRR, between the prediction and transform stages, a certain kind of reordering is implemented on the residual samples for each mode in spatial. The reordering is manipulated in a way that the distribution statistics of reordered residual samples present less mode-dependent characteristics. For example, the normalized distributions of absolute residual magnitudes in vertical and horizontal prediction are shown in Fig. 4(a) and (b). They exhibit different distribution statistics, but after reordering the horizontal prediction residuals in Fig. 4(c), the reordered residuals have similar distribution with vertical predicted residuals. Fig. 5 illustrates the residual reordering scheme, which

$$\underset{(a)}{\begin{bmatrix} 4.47 & 4.43 & 4.44 & 4.59 \\ 5.87 & 5.65 & 5.75 & 6.09 \\ 6.83 & 6.63 & 6.82 & 7.16 \\ 7.77 & 7.64 & 7.82 & 8.14 \end{bmatrix}}, \underset{(b)}{\begin{bmatrix} 4.35 & 5.92 & 6.72 & 7.51 \\ 4.31 & 5.83 & 6.65 & 7.75 \\ 4.46 & 6.06 & 6.93 & 7.89 \\ 4.47 & 6.17 & 7.06 & 8.13 \end{bmatrix}}, \underset{(c)}{\begin{bmatrix} 4.35 & 4.31 & 4.46 & 4.47 \\ 5.92 & 5.83 & 6.06 & 6.17 \\ 6.72 & 6.65 & 6.93 & 7.06 \\ 7.51 & 7.75 & 7.89 & 8.13 \end{bmatrix}}$$

Fig. 4. Normalized distributions of absolute residual magnitudes for the illustration of residual reordering scheme. (a) vertical prediction (b) horizontal prediction, (c) reordered residual of horizontal prediction.

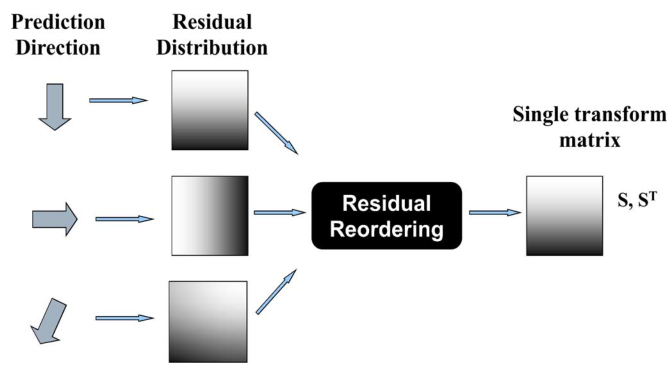


Fig. 5. An example of MDRR scheme for different prediction directions.

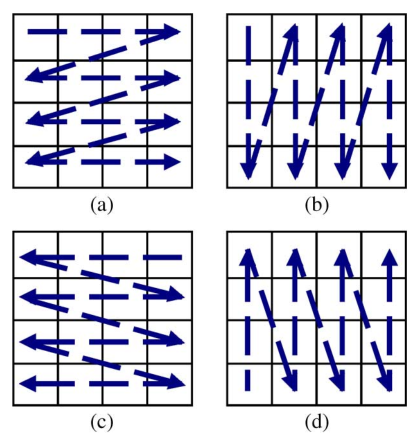


Fig. 6. Residual reordering methods for 4×4 blocks.

demonstrates that different prediction mode will produce similar residual statistics after reordering. Therefore, the intra prediction modes can be assigned into two groups (directional, non-directional), and only one transform matrix is employed for each group.

As different direction modes carry different statistics, considering the intra prediction directions in HEVC, four residual reordering mechanisms are proposed in MDRR, which is illustrated in Fig. 6. The mapping between intra prediction mode and reordering method is illustrated in Fig. 7. It is shown that 33 angular prediction directions are grouped into four categories and each category corresponds to one reordering order. Therefore, only one transform matrix is assigned for the 33 directional intra prediction modes, and one for DC/Planar mode. This scheme significantly reduces the number of transform matrices and implementation complexity of MDDT.

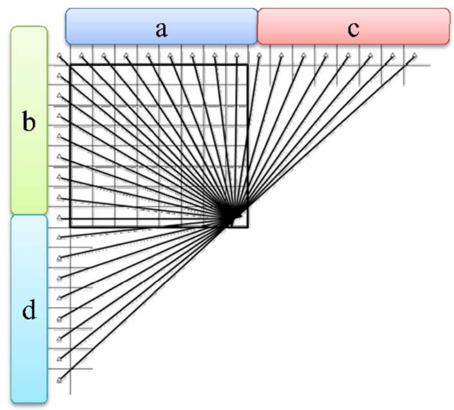


Fig. 7. Mapping from 33 angular prediction directions to residual reordering method.

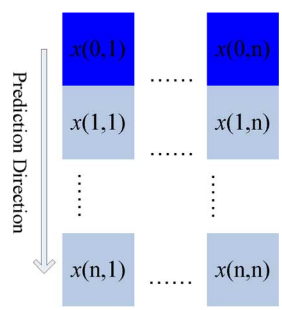


Fig. 8. Reference and source pixels in an $n \times n$ vertical predicted block.

TABLE I
VARIANCE VALUE FOR EACH TRANSFORM COEFFICIENT IN 8×8 BLOCK

8.087	6.312	4.876	3.601	2.744	2.213	1.901	1.737
2.068	1.614	1.247	0.921	0.702	0.566	0.486	0.444
0.712	0.555	0.429	0.317	0.241	0.195	0.167	0.153
0.341	0.266	0.206	0.152	0.116	0.093	0.080	0.073
0.215	0.168	0.130	0.096	0.073	0.059	0.051	0.046
0.156	0.122	0.094	0.069	0.053	0.043	0.037	0.033
0.127	0.099	0.076	0.056	0.043	0.035	0.030	0.027
0.112	0.088	0.068	0.050	0.038	0.031	0.026	0.024

IV. MODE DEPENDENT TRANSFORM COEFFICIENT SCAN

In the development of video coding standards, fixed scan patterns are investigated, such as the zigzag and diagonal scans. These scan methods arrange the 2-D transform coefficients into 1-D array according to its frequency in the descending order. However, since these scan patterns are not adaptive to the value of coefficients; there is still much room for improvement by employing adaptive coefficient scan methods. In this section, we will start with a theoretical analysis by deriving the variance of transform coefficients according to the image correlation model. Then the representative mode-dependent coefficient scan (MDCS) schemes in the course of HEVC development are introduced.

A. Observations of Transform Coefficient Variance

We assume an image correlation model as follows,

$$E\left\{x(p_0,q_0)\cdot x(p_1,q_1)\right\} = \sigma_s^2 \cdot \rho_x^{|p_0-p_1|} \cdot \rho_y^{|q_0-q_1|} \tag{7}$$

where σ_s^2 is the variance of signal samples; $x(p_0, q_0)$ and $x(p_1, q_1)$ are two samples with coordinates of (p_0, q_0) and (p_1, q_1); ρ_x and ρ_y are correlation coefficients for horizontal and vertical directions, respectively. To simplify the derivation, we assume that each image pixel is a random variable with zero mean and unit variance [28]. We take an $n \times n$ block X predicted along vertical direction as an example, as illustrated in Fig. 8. If we define the transform coefficient of X as a 1-D vector: $\overrightarrow{f} = [f_{11}, f_{11} \ldots f_{nn}]^T$, then in the Appendix, we show that the variance of the element in $\overrightarrow{f}$ can be calculated as

$$E\left\{\vec{f}_i^2\right\} = E\left\{\sum_{k=1}^{n(n+1)} \sum_{l=1}^{n(n+1)} A_{i,k}A_{i,l}\vec{p_k}\vec{p_l}\right\} \tag{8}$$

where A is a constant matrix related to the transform and $\overrightarrow{p}$ is the composed of the original and predicted samples: $\overrightarrow{p} = [x_{0,1} \ldots x_{0,n}, x_{11} \ldots x_{nn}]^T$. Putting (7) into (8), it is derived that the transform residual variance is dependent on the correlation coefficients of horizontal and vertical directions, which is closely related to the image oriented structure and thereby the intra prediction mode. In general, blocks with high ρ_y values tend to select vertical prediction as the optimal prediction direction. We give an example by setting $\rho_x = 0.4$, $\rho_y = 0.9$ and $\mathrm{n} = 8$, and the calculated variance value for each transform coefficient is shown in Table I. It is observed that in this case, the first row takes up more than 50% of the total energy. The coefficient variances of more test cases can be derived in the same way with (8). Given the variances of each transform coefficient, the corresponding optimal scanning order can be retrieved by sorting the positions in a descending order of the variance values. Based on this principle, it is observed that the horizontal scan is approximated to be the optimal scan pattern. The variance of the transform coefficients in horizontal intra prediction can be derived in a similar way, and it can also be derived that vertical scan is generally more appropriate for the horizontal intra prediction.

B. Mode Dependent Coefficient Scan Algorithms

In recent years, the MDCS scheme has been intensively studied. In our work [57], the research on MDCS for intra coding blocks was presented. For a specific intra prediction mode, the proposed scan method approximately orders the coefficients in decreasing order of the variance with the assumption of the Laplace distribution. Another adaptive coefficient scan scheme for intra coding was further proposed in [58], where six scanning orders were developed based on the energy of the transform coefficients. In [59], the authors sorted the scan table according to the proportion of significant coefficients to achieve intra mode dependent coefficient coding. In [10], the MDCS scheme was developed collaboratively with MDDT. This scheme is more flexible by incorporating the scan order updating mechanism; that is, after coding each macro-block, the statistics of nonzero coefficients at each position are gathered and the coefficient scan order can be updated.

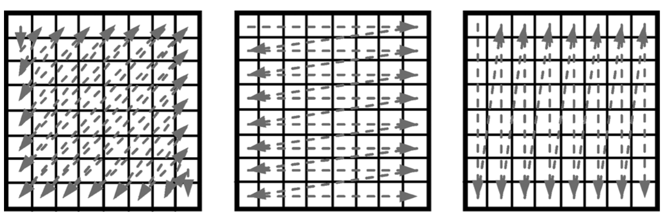
In the course of the development of HEVC standard, MDCS has also been thoroughly studied [60]–[67]. In KTA, the coefficient scan order is trained off-line for each prediction mode

TABLE II
CODING PERFORMANCE OF MODE DEPENDENT TRANSFORM AND RESIDUAL REORDER TOOLS (AI MAIN CONFIGURATION) [%]

Sequence	MDDT			MDRR			DST (HEVC)			DST [33]			RDOT		
	Y	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*
Class A	-1.0	-0.3	-0.2	-0.9	-0.1	-0.1	-1.2	-1.1	-1.1	-1.0	-0.9	-0.9	-0.6	0.9	0.9
Class B	-0.8	0.0	-0.1	-0.7	0.0	0.1	-0.7	-0.6	-0.7	-0.7	-0.5	-0.5	-0.9	1.1	1.1
Class C	-1.0	-0.3	-0.3	-1.0	-0.1	-0.1	-0.8	-0.9	-0.9	-0.7	-0.7	-0.7	-1.6	2.0	2.0
Class D	-1.1	-0.4	-0.4	-1.0	-0.2	-0.2	-0.8	-0.9	-0.9	-0.5	-0.7	-0.6	-2.1	1.7	1.8
Class E	-1.6	-0.7	-0.6	-1.5	-0.5	-0.5	-1.7	-1.7	-1.6	-1.6	-1.4	-1.2	-0.9	1.0	1.1
Average	**-1.1**	**-0.3**	**-0.3**	**-1.0**	**-0.2**	**-0.1**	**-1.0**	**-1.0**	**-1.0**	**-0.8**	**-0.8**	**-0.7**	**-1.3**	**1.4**	**1.4**

TABLE III
CODING PERFORMANCE OF MODE DEPENDENT TRANSFORM AND RESIDUAL REORDER TOOLS (RA MAIN CONFIGURATION) [%]

Sequence	MDDT			MDRR			DST (HEVC)			DST [33]			RDOT		
	Y	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*	*Y*	*U*	*V*
Class A	-0.3	1.5	1.7	-0.3	1.6	1.6	-0.5	-0.2	-0.2	-0.5	-0.3	-0.2	-0.1	1.7	1.8
Class B	-0.2	1.7	1.8	-0.2	1.8	2.0	-0.3	-0.2	-0.1	-0.3	-0.1	0.0	-0.5	2.1	2.1
Class C	-0.5	2.0	2.0	-0.5	2.1	2.0	-0.3	-0.4	-0.3	-0.2	-0.3	-0.1	-1.1	2.1	2.2
Class D	-0.5	2.2	2.4	-0.4	1.9	2.7	-0.2	-0.2	-0.5	-0.1	-0.2	-0.2	-1.1	2.5	2.5
Class E															
Average	**-0.4**	**1.9**	**2.0**	**-0.4**	**1.9**	**2.1**	**-0.3**	**-0.2**	**-0.3**	**-0.2**	**-0.2**	**-0.1**	**-0.8**	**2.2**	**2.2**



(a) (b) (c)

Fig. 9. Three scanning patterns in HEVC: (a) diagonal; (b) horizontal; (c) vertical.

to get the scan table. However, this requires additional memory for storing the coefficient scan orders. To solve this problem, in [60] the authors proposed to use the same scan table for contiguous directions. In [61] the number of scan patterns is further reduced to three, namely zigzag, horizontal and vertical. Afterwards, zigzag is substituted by diagonal scan for its convenience in developing parallel coding [20]. The three scan patterns finally adopted in HEVC are shown in Fig. 9. In HEVC, the horizontal and vertical scans are only applied in the intra case for 4×4 and 8×8 TUs according to a look up table, which maps the intra prediction mode to one of the scan patterns. Specifically, vertical scan is used for directions close to the horizontal prediction; while horizontal scan is used for directions close to vertical; and diagonal is used for other directions. To further improve the compression performance, an extended mode-dependent coefficient scan method is proposed to HEVC [64], [65]. Based on the MDCS in HEVC, this scheme makes the extension that a modified MDCS is applied to larger block sizes including 16×16 and 32×32.

V. PERFORMANCE ANALYSIS

In this section, we present and analyze the coding performance of the mode dependent coding tools, which demonstrate their superior performance. The test platform is the HEVC reference software HM10.0 [68] with All Intra (AI) main and Random Access (RA) main settings.

TABLE IV
HEVC MDCS CODING PERFORMANCE IN TERMS OF BD-RATE [%] (ANCHOR: MERELY DIAGONAL SCAN, PLATFORM: HM10.0)

Sequence	***AI-main***			***RA-main***		
	Y	*U*	*V*	*Y*	*U*	*V*
Class A	-0.7	-0.9	-0.8	-0.4	-0.2	-0.3
Class B	-0.5	-0.7	-0.9	-0.3	-0.3	-0.4
Class C	-1.0	-1.5	-1.5	-0.4	-0.6	-0.6
Class D	-1.0	-1.6	-1.7	-0.3	-0.6	-0.9
Class E	-1.4	-0.5	-0.6			
Average	**-0.9**	**-1.1**	**-1.1**	**-0.3**	**-0.4**	**-0.6**

TABLE V
EXTENDED MDCS [65] CODING PERFORMANCE IN TERMS OF BD-RATE [%] (ANCHOR: HM10.0)

Sequence	***AI-main***			***RA-main***		
	Y	*U*	*V*	*Y*	*U*	*V*
Class A	-0.4	-0.9	-0.9	-0.1	-0.3	-0.4
Class B	-1.3	-2.8	-2.7	-0.4	-1.1	-1.1
Class C	-0.6	-1.1	-1.0	-0.2	-0.3	-0.3
Class D	-0.5	-1.1	-0.9	-0.1	-0.4	-0.2
Class E	-2.6	-3.8	-3.7			
Average	**-1.1**	**-2.0**	**-1.9**	**-0.3**	**-0.6**	**-0.6**

A. Performance of Mode Dependent Transform and Residual Reorder

In this subsection, the coding efficiency of mode dependent transform and residual reordering tools are investigated, including MDDT, RDOT, DST and MDRR. It is noted that the coding performance of both the DST scheme in HEVC [32], [34] and the DST scheme in [33] are studied. For a fair comparison, all the tools are implemented on 4×4 luma transform block, and the anchor is produced by disabling the DST scheme in HM10.0 software.

The coding performance of these coding tools of AI and RA configurations are shown in Table II and III, respectively. For the AI main setting, we can observe that MDDT for 4×4 can achieve 1.1% bit rate reduction, and the DST scheme in HEVC can achieve comparable coding performance with

TABLE VI
ENCODING AND DECODING COMPUTATIONAL COMPLEXITY EVALUATION OF MODE DEPENDENT CODING TOOLS [%]

Sequence	MDDT		MDRR		DST (HEVC)		DST [33]		RDOT		MDCS (HEVC)		MDCS [65]	
	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}	ΔT_{enc}	ΔT_{dec}
Class A	1.65	0.83	0.39	1.67	0.37	1.04	0.77	1.41	75.13	1.74	0.80	1.05	1.23	2.10
Class B	1.83	1.06	0.48	2.00	0.49	2.32	0.19	2.04	79.48	-0.26	0.24	0.79	0.71	0.78
Class C	2.03	0.18	0.94	1.10	0.16	1.55	0.16	1.00	74.54	2.98	1.25	0.31	0.95	0.59
Class D	1.79	1.43	0.98	1.08	0.10	0.17	0.09	1.42	77.45	4.29	0.47	0.51	1.50	1.07
Class E	1.16	2.91	-0.09	3.37	0.63	1.59	0.06	0.51	72.48	1.37	0.55	0.09	1.79	1.05
Average	**1.70**	**1.21**	**0.56**	**1.78**	**0.34**	**1.37**	**0.26**	**1.35**	**76.17**	**1.94**	**0.64**	**0.59**	**1.18**	**1.10**

MDDT. However, the DST has much lower implementation complexity. In MDRR, the number of transform matrices is also significantly reduced, and we can also observe that MDRR introduces 0.1% coding performance loss compared with MDDT. Since RDOT employs much more transform candidates for selection, 1.3% bit rate reduction is achieved compared with the anchor. For RA configuration, as much less intra CUs are employed for a sequence, the performance gains are from -0.2% to -0.8% for these coding tools.

B. Performance of Mode Dependent Coefficient Scan

Performance of the MDCS scheme in HEVC is presented in Table IV, where the anchor is generated by disabling the HEVC MDCS scheme in [61]. As Table IV shows, the mode dependent scan scheme in HEVC can bring 0.9% BD-rate reduction on average for AI-main configuration. The coding performance improvement for RA-main setting is -0.3% as MDCS only applies in intra case. On top of the MDCS in HEVC, the coding performance of the extended MDCS scheme [65] in Table V shows that, by applying MDCS for large block size of 16×16 and 32×32, 1.1% and 0.3% rate reductions are observed for AI-main and RA-main configurations, respectively.

C. Complexity Evaluation and Analysis

In this subsection, we demonstrate the encoding and decoding complexity for these mode dependent coding tools. For the coding complexity evaluation, we calculate ΔT as follows,

$$\Delta T = \frac{T_{pro} - T_{anc}}{T_{anc}} \times 100\% \tag{9}$$

where T_{pro} and T_{anc} denote the total coding time with the evaluated coding tool and anchor, respectively. Table VI shows the computational overhead for both encoding and decoding. The coding time is obtained with Intel Xeon @2.80 GHz Core processor and 24 GB random access memory.

From Table VI, it is observed that the encoder computational overheads for MDDT, MDRR, DST and MDCS are about 0.3%$\sim$1.7%. As the utilization of these coding tools is implied in the coding mode, little additional computational task except the transform itself is required in the encoder. For MDDT and RDOT, full matrix multiplication is employed, and the complexity of RDOT is in proportion to the number of transform matrices for selection in each mode. Therefore, the full RD search in RDOT will impose high computational burden to the encoder, which leads to 76.2% encoding time increasing. All the coding tools won't have significant impact on the decoding complexity. Moreover, the decoding complexity of MDRR is increased compared to MDDT, as in the decoder the residual reorder is also required.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we review a number of advances in mode dependent video coding technology, including mode dependent transform, residual reordering, and transform coefficient scan. These coding tools are finely tuned with the prediction mode information. In this paper, these advanced coding tools are implemented in HEVC reference software HM10.0, and experimental results show that mode dependent transform and scan will each bring about 0.3$\sim$1.1% bit rate reduction in terms of different coding configurations. For mode dependent residual reordering, 0.1% coding performance loss is observed, while the implementation complexity can be significantly reduced.

In general, the natural images exhibit non-stationary properties, and to fully exploit image statistical properties, the video coding technology should adapt to the local image content to achieve advanced coding performance. The prediction modes provide important image context information at the beginning of video coding date flow, and all the subsequent process can benefit from this prediction mode indication. Moreover, applying the mode information to video coding will economize on the computational resources consumed in the image characteristics exploration, and also save some additional signaling information. These new features provide mode dependent coding with a competitive compression performance. In the future, the mode dependent technology in other modules will be further investigated, such as mode dependent in-loop filtering. We believe that the mode dependent video coding technology described in this paper or similar technology developed from this ground could play an important role in future video standardization.

APPENDIX

As shown in Fig. 8, for an $n \times n$ block X predicted along vertical direction with the prediction samples, the residual block Y is calculated as

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,n} \end{bmatrix}$$
$$= \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,n} \end{bmatrix} - \begin{bmatrix} x_{0,1} & x_{0,2} & \cdots & x_{0,n} \\ x_{0,1} & x_{0,2} & \cdots & x_{0,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0,1} & x_{0,2} & \cdots & x_{0,n} \end{bmatrix} \tag{10}$$

which can be alternatively written as

$$\begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,n} \\ y_{2,1} \\ \vdots \\ y_{n,n} \end{bmatrix} = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ \vdots \\ x_{0,n} \\ x_{1,1} \\ x_{1,2} \\ \vdots \\ x_{1,n} \\ x_{2,1} \\ \vdots \\ x_{n,n} \end{bmatrix}$$

$$= P \cdot \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ \vdots \\ x_{0,n} \\ x_{1,1} \\ x_{1,2} \\ \vdots \\ x_{1,n} \\ x_{2,1} \\ \vdots \\ x_{n,n} \end{bmatrix} \tag{11}$$

To perform transform on Y, the following arithmetic is implemented,

$$F = C \cdot Y \cdot R \tag{12}$$

where C and R are the $n \times n$ column and row transform matrix and are defined as follows

$$C = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \cdots & c_{n,n} \end{bmatrix}, \quad R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,1} & r_{n,2} & \cdots & r_{n,n} \end{bmatrix} \tag{13}$$

Similar to (11), (12) can be alternatively written as

$$\begin{bmatrix} f_{1,1} \\ f_{1,2} \\ \vdots \\ f_{1,n} \\ f_{2,1} \\ \vdots \\ f_{n,n} \end{bmatrix} = \begin{bmatrix} c_{1,1}r_{1,1} & c_{1,1}r_{2,1} \cdots c_{1,1}r_{n,1} & c_{1,2}r_{1,1} \cdots c_{1,n}r_{n,1} \\ c_{1,1}r_{1,2} & c_{1,1}r_{2,2} \cdots c_{1,1}r_{n,2} & c_{1,2}r_{1,2} \cdots c_{1,n}r_{n,2} \\ \vdots & \vdots \quad \ddots \quad \vdots & \vdots \quad \vdots \quad \vdots \\ c_{1,1}r_{1,n} & c_{1,1}r_{2,n} \cdots c_{1,1}r_{n,n} & c_{1,2}r_{1,n} \cdots c_{1,n}r_{n,n} \\ c_{2,1}r_{1,1} & c_{2,1}r_{2,1} \cdots c_{2,1}r_{n,1} & c_{2,2}r_{1,1} \cdots c_{2,n}r_{n,1} \\ \vdots & \vdots \quad \ddots \quad \vdots & \vdots \quad \ddots \quad \vdots \\ c_{n,1}r_{1,n} & c_{n,1}r_{2,n} \cdots c_{n,1}r_{n,n} & c_{n,2}r_{1,n} \cdots c_{n,n}r_{n,n} \end{bmatrix} \times \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,n} \\ y_{2,1} \\ \vdots \\ y_{n,n} \end{bmatrix} \tag{14}$$

Assume B to be the Kronecker product of C and R^T. Let $A = B \cdot P$, $\overrightarrow{f} = [f_{11}, f_{11} \ldots f_{nn}]^T$ and $\overrightarrow{p} = [x_{0,1} \ldots x_{0,n}, x_{11} \ldots x_{nn}]^T$., then we have

$$\vec{f} = A \cdot \vec{p} \tag{15}$$

Therefore, the variance of the element in $\overrightarrow{f}$ can be calculated as

$$\begin{aligned} E\left\{\vec{f}_i^2\right\} &= E\left\{\left(\sum_{j=1}^{n(n+1)} A_{i,j} \cdot \vec{p_j}\right)^2\right\} \\ &= E\left\{\sum_{k=1}^{n(n+1)} \sum_{l=1}^{n(n+1)} A_{i,k}A_{i,l}\vec{p_k}\vec{p_l}\right\} \end{aligned} \tag{16}$$

Acknowledgment

The authors thank the experts of ITU-T VCEG, ISO/IEC MPEG, and the ITU-T/ISO/IEC Joint Collaborative Team on Video Coding for their contributions on the mode dependent coding tools. The authors also would like to thank the anonymous reviewers for their valuable comments that significantly helped us in improving the presentation of the paper. The authors are grateful to Dr. Li Zhang and Dr. Xin Zhao for their valuable suggestions.

References

[1] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: video*, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 video), ITU-T and ISO/IEC JTC 1, Nov. 1994, MPEG-2 video.

[2] *Video Coding for Low Bit Rate Communication*, ITU-T Rec. H.263, ITU-T, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000, H.263.

[3] *Coding of Audio-Visual Objects—Part 2: Visual*, ISO/IEC 14492-2 (MPEG-4 Visual), ISO/IEC JTC 1, Version 1: Apr. 1999, Version 2: Feb. 2000, Version 3: May 2004, MPEG-4 visual.

[4] *Advanced Video Coding for Generic Audio-Visual Services*, ITU T Rec. H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC JTC 1, May 2003.

[5] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 10 (for FDIS & consent)," in *Proc. JCTVC-L1003*, Jan. 2013.

[6] 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, Dec. 2012.

[7] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[8] I. K. Kim, J. Min, T. Lee, W. J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1697–1706, Dec. 2012.

[9] J. Xu, B. Zeng, and F. Wu, "An overview of directional transforms in image coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2010, pp. 3036–3039.

[10] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 2116–2119.

[11] O. G. Sezer, R. A. Cohen, and A. Vetro, "Robust learning of 2-D Separable transforms for next-generation video coding," in *Proc. Data Compression Conf.*, 2011, pp. 63–72.

[12] X. Zhao, L. Zhang, S. Ma, and W. Gao, "Video coding with rate-distortion optimized transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 1, pp. 138–151, Jan. 2012.

[13] R. A. Cohen, S. Klomp, A. Vetro, and H. Sun, "Direction-adaptive transforms for coding prediction residuals," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2010, pp. 185–188.

[14] M. Budagavi and M. Zhou, "Orthogonal MDDT and mode dependent DCT," , ITU-T Q.6/SG16 VCEG, VCEG-AM20_r1. Kyoto, Japan, Jan. 2010.

[15] R. Cohen, C. Yeo, R. Joshi, W. Ding, S. Ma, A. Tanizawa, H. Yang, X. Zhao, and L. Zhang, "Tool experiment 7: MDDT simplification," in *Proc. JCTVC-B307*, Geneva, CH, Jul. 21–28, 2010.

[16] H. Yang, J. Zhou, and H. Yu, "Simplified MDDT (SMDDT) for intra prediction residual," in *Proc. JCTVC-B039*, Geneva, Switzerland, Jul. 21–28, 2010.

[17] R. Cohen, A. Vetro, and H. Sun, "Alternative performance measurement of MDDT and ROT in TMuC," in *Proc. JCTVC-C036*, Guangzhou, China, Oct. 7–15, 2010.

[18] R. Cohen, C. Yeo, and R. Joshi, "TE7: Summary report for MDDT simplification," in *Proc. JCTVC-C310*, Guangzhou, China, Oct. 7–15, 2010.

[19] W. C. Yueh, "Eigenvalues of several tridiagonal matrices," *Appl. Math. E-Notes*, vol. 5, pp. 66–74, Apr. 2005.

[20] V. Sze and M. Budagavi, "CE11: Parallelization of HHI_TRANSFORM_CODING," in *Proc. JCTVC-F129, 6th Meeting*, Torino, Italy, Jul. 14–22, 2011.

[21] M. Loève, "Probability theory, II," in *Graduate Texts in Mathematics*, 4th ed. New York, NY, USA: Springer-Verlag, 1978.

[22] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 726–729.

[23] J. Han, A. Saxena, V. Melkote, and K. Rose, "Jointly optimized spatial prediction and block transform for video and image coding," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1874–1884, Apr. 2012.

[24] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, "Mode-dependent fast separable KLT for block-based intra coding," in *Proc. ISCAS*, 2011, pp. 621–624.

[25] C. Yeo, Y. H. Tan, and Z. Li, "Low-complexity mode-dependent KLT for block-based intra coding," in *Proc. IEEE Int. Conf. Image Process.*, 2011, pp. 3685–3688.

[26] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, "Mode-dependent transforms for coding directional intra prediction residuals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 545–554, Apr. 2012.

[27] A. Saxena and F. C. A. Fernandes, "Mode dependent DCT/DST for intra prediction in block-based image/video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2011, pp. 1685–1688.

[28] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, "Mode-dependent fast separable KLT for block-based intra coding," in *Proc. JCTVC-B024*, Geneva, Switzerland, Jul. 2010.

[29] C. Yeo, Y. Tan, Z. Li, and S. Rahardja, "Choice of transforms in MDDT for unified intra prediction," in *Proc. JCTVC-C039*, Guangzhou, China, Oct. 7–15, 2010.

[30] A. Saxena and F. C. A. Fernandes, "Jointly optimal intra pre-diction and adaptive primary transform," in *Proc. JCTVC-C108*, Oct. 2010.

[31] A. Saxena and F. C. A. Fernandes, "Mode-dependent DCT/DST for intra-prediction in video coding," in *Proc. JCTVC-D033*, Jan. 2011.

[32] A. Saxena and F. C. A. Fernandes, "CE7: Mode-dependent DCT/DST without $4 * 4$ full matrix multiplication for intra prediction," in *Proc. JCTVC-E125*, Geneva, Switzerland, Mar. 16–23, 2011.

[33] X. Zhao, L. Zhang, S. Ma, and W. Gao, "Simplified multiplier less 4×4 DST for intra prediction residue," in *Proc. JCTVC-D286*, Guangzhou, China, Oct. 7–15, 2011.

[34] K. Ugur and A. Saxena, "Summary report of core experiment on intra transform mode dependency simplifications," in *Proc. JCTVC-J0021*, Jul. 2012.

[35] A. Saxena, F. C. A. Fernandes, and Y. A. Reznik, "Fast transforms for Intra-prediction-based image and video coding," in *Proc. Data Compression Conf.*, 2013, pp. 13–22.

[36] R. K. Chivukula and Y. A. Reznik, "Fast computing of discrete cosine and sine transforms of types vi and vii," in *Applications of Digital Image Processing XXXIV*, A. G. Tescher, Ed. Philadelphia, PA, USA: SPIE, 2011, vol. 8135, Proc. SPIE, p. 813, 505-1-14.

[37] A. Saxena and F. C. A. Fernandes, "On secondary transforms for intra prediction residual," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 1201–1204.

[38] A. Saxena and F. C. A. Fernandes, "On secondary transforms for prediction residual," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2012, pp. 2489–2492.

[39] E. Alshina, A. Alshin, and F. C. A. Fernandes, "Rotational transform for image and video compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2011, pp. 3689–3692.

[40] F. C. A. Fernandes, "Low complexity rotational transform," in *Proc. JCTVC-C096*, Guangzhou, China, Oct. 7–15, 2010.

[41] E. Alshina, A. Alshin, W.-J. Han, R. Joshi, and M. Coban, "TE12.4: Experimental results of MDDT and ROT by Samsung and Qualcomm," in *Proc. JCTVC-C202*, Guangzhou, China, Oct. 7–15, 2010.

[42] A. Saxena, Y. Shibahara, F. C. A. Fernandes, and T. Nishi, "CE 7: On secondary transforms for intra prediction residual," in *Proc. JCT-G108*, Geneva, Switzerland, Nov. 21–30, 2011.

[43] A. Saxena, Y. Shibahara, F. C. A. Fernandes, and T. Nishi, "CE 7: On secondary transforms for intra prediction residual," in *Proc. JCT-H0125*, San José, CA, USA, Feb. 2012.

[44] A. Saxena, Y. Shibahara, F. C. A. Fernandes, and T. Nishi, "CE 7: On secondary transforms for inter prediction residual," in *Proc. JCT-H0126*, San José, CA, USA, Feb. 2012.

[45] Y. Shibahara and T. Nishi, "Mode dependent 2-step transform for intra coding," , ITU-T & ISO/IEC JCTVC-F224, Jul. 2011.

[46] A. Saxena and F. C. A. Fernandes, "On secondary transforms for intra prediction residual," , ITU-T & ISO/IEC JCTVC-F554, Jul. 2011.

[47] I. H. Dinstein, K. Rose, and A. Heiman, "Variable block-size transform image coder," *IEEE Trans. Commun.*, vol. 38, no. 11, pp. 2073–2078, Nov. 1990.

[48] M. Wien and A. Dahlhoff, "Adaptive block transforms," in *Proc. VCEG-M62, 13th Meeting*, Austin, TX, USA, Apr. 2–4, 2001.

[49] C. Zhang, K. Ugur, J. Lainema, A. Hallapuro, and M. Gabbouj, "Video coding using spatially varying transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 2, pp. 127–140, Feb. 2011.

[50] Y. Yuan, X. Zheng, X. Peng, J. Xu, I.-K. Kim, L. Liu, Y. Wang, X. Cao, C. Lai, J. Zheng, Y. He, and H. Yu, "CE2: Non-square quadtree transform for symmetric and asymmetric motion partition," in *Proc. JCTVC-F412, 6th Meeting*, Torino, Italy, Jul. 14–22, 2011.

[51] Y. Yuan, X. Zheng, X. Peng, J. Xu, Y. Wang, X. Cao, C. Lai, J. Zheng, Y. He, and H. Yu, "CE2: Non-square quadtree transform for symmetric motion partitions," in *Proc. JCTVC-F410, 6th Meeting*, Torino, Italy, Jul. 14–22, 2011.

[52] Y. Yuan, I. Kim, X. Zheng, L. Liu, X. Cao, S. Lee, M. Cheon, T. Lee, Y. He, and J. Park, "Quadtree based non-square block structure for inter frame coding in HEVC," *IEEE Trans. Circuits System Video Technol.*, vol. 22, no. 12, pp. 1701–1719, Dec. 2012.

[53] Q. Yao, Z. Wang, H. Bian, and M. Xu, "Fundamentals in image coding," in . Hangzhou, China: Zhejiang Univ. Press, 1992.

[54] F. Kamisli and J.-S. Lim, "Transforms for the motion compensation residual," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 19-24, 2009, pp. 789–792.

[55] X. Zhao, L. Zhang, S. Ma, and W. Gao, "Mode-dependent residual reordering for Intra prediction residual," in *Proc. JCTVC-B102*, Geneva, Switzerland, Jul. 21–28, 2010.

[56] X. Zhao, L. Zhang, S. Ma, and W. Gao, "TE7: Results for mode-dependent residual reordering for intra prediction residual," in *Proc. JCTVC-C089*, Guangzhou, China, Oct. 7–15, 2010.

[57] X. Fan, Y. Lu, and W. Gao, "A novel coefficient scanning scheme for directional spatial prediction-based image compression," in *Proc. Int. Conf. Multimedia Expo (ICME'03)*, pp. II-557–II-560.

[58] B.-D. Choi, J.-H. Kim, and S.-J. Ko, "Adaptive coefficient scanning based on the intra prediction mode," *ETRI J.*, vol. 29, no. 5, pp. 694–696, Oct. 2007.

[59] Y. Lee, K. Han, D. Sim, and J. Seo, "Adaptive scanning for H264/AVC intra coding," *ETRI J.*, vol. 28, no. 5, Oct. 2006.

[60] H. Yang, J. Zhou, and H. Yu, "TE7: Symmetry-based simplification of MDDT," in *Proc. JCTVC-C102*, Guangzhou, China, Oct. 2010.

[61] Y. Zheng, M. Coban, J. Sole, R. Joshi, and M. Karczewicz, "CE11: Mode dependent coefficient scanning," in *Proc. JCTVC-D393, 4th Meeting*, Daegu, Korea, Jan. 20–28, 2011.

[62] V. Sze, K. Panusopone, J. Chen, T. Nguyen, and M. Coban, "Description of Core Experiment (CE11): Coefficient scanning and coding," in *Proc. JCTVC-D611, 4th Meeting*, Daegu, Korea, Jan. 20–28, 2011.

[63] J. Song, M. Yang, H. Yang, and H. Yu, "CE11: Adaptive coefficients scanning for inter-frame coding," in *Proc. JCTVC-E296*, Geneva, Switzerland, Mar. 16–23, 2011.

[64] X. Zhao, X. Guo, M. Guo, S. Lei, S. Ma, and W. Gao, "Extended mode-dependent coefficient scanning," in *Proc. JCTVC-F124, 6th Meeting*, Torino, Italy, Jul. 14–22, 2011.

[65] X. Zhao, X. Guo, M. Guo, S. Lei, S. Ma, and W. Gao, "CE11: Extended mode dependent coefficient scanning," in *Proc. JCTVC-G284, 7th Meeting*, Geneva, Switzerland, Nov. 21–30, 2011.

[66] J. Song, M. Yang, H. Yang, and H. Yu, "Mode dependent coefficient scan for inter blocks," in *Proc. JCTVC-D501, 6th Meeting*, Torino, Italy, Jul. 14–22, 2011.

[67] K. Sharman, J. Gamei, N. Saunders, and P. Silcock, "Hybrid horizontal/vertical with diagonal mode dependent coefficient scan," in *Proc. JCTVC-G491, 7th Meeting*, Geneva, Switzerland, Nov. 21–30, 2011.

[68] K. McCann, B. Bross, W.-J. Han, I.-K. Kim, K. Sugimoto, and G. J. Sullivan, "High efficiency video coding (HEVC) test model 10 (HM 10) encoder description," in *Proc. JCTVC-L1002*, Geneva, Switzerland, Jan. 14–23, 2013.

Siwei Ma received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. From 2005 to 2007, he held a post-doctorate position with the University of Southern California, Los Angeles. Then, he joined the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, where he is currently an Associate Professor. He has published over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

Shiqi Wang received the B.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently working toward the Ph.D. degree in computer science at Peking University, Beijing, China. He was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada from 2010 to 2011. From Apr. 2011 to Aug. 2011, he was with Microsoft Research Asia, Beijing, as an Intern. His current research interests include video compression, image and video quality assessment and multi-view video coding.

Qin Yu received the B.S. degree in electronic engineering from Wuhan University, Hubei, China, in 2011. Currently, she is working toward the M.S. degree with the Institute of Digital Media, Peking University, Beijing, China. Her research interest is in the area of video compression.

Junjun Si received the B.S. degree in computer science from Beijing Jiao Tong University, Beijing, China, in 2011. Currently, she is working toward the M.S. degree with the Institute of Digital Media, Peking University, Beijing, China. Her research interest includes video coding technology and video quality assessment.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Japan, in 1991. He is a professor of computer science at Peking University, China. Before joining Peking University, he was a professor of computer science at Harbin Institute of Technology from 1991 to 1995, and a professor at the Institute of Computing Technology of Chinese Academy of Sciences. He has published extensively including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. Dr. Gao served or serves on the editorial board for several journals, such as IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, EURASIP Journal of Image Communications, Journal of Visual Communication and Image Representation. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.