Shuffled Discrete Sine Transform in Inter-Prediction Coding

Jun-woo Choi, Nam-Uk Kim, Sung-Chang Lim, Jungwon Kang, Hui Yong Kim, and Yung-Lyul Lee

Video compression exploits statistical, spatial, and temporal redundancy, as well as transform and quantization. In particular, the transform in a frequency domain plays a major role in energy compaction of spatial domain data into frequency domain data. The high efficient video coding standard uses the type-II discrete cosine transform (DCT-II) and type-VII discrete sine transform (DST-VII) to improve the coding efficiency of residual data. However, the DST-VII is applied only to the Intra 4 × 4 residual block because it yields relatively small gains in the larger block than in the 4 × 4 block. In this study, after rearranging the data of the residual block, we apply the DST-VII to the inter-residual block to achieve coding gain. The rearrangement of the residual block data is similar to the arrangement of the basis vector with a the lowest frequency component of the DST-VII. Experimental results show that the proposed method reduces the luma-chroma (Cb+Cr) BD rates by approximately 0.23% to 0.22%, 0.44% to 0.58%, and 0.46% to 0.65% for the random access, low delay B, and low delay P configurations, respectively.

Keywords: HEVC/H.265, Inter-coding, Residual data, DST-VII, DCT-II.

I. Introduction

The video coding standard, high efficient video coding (HEVC) [1], [2], was jointly developed by the ITU-T video coding expert group (VCEG) and ISO/IEC moving picture expert group (MPEG). HEVC adopted many efficient coding technologies. It has a flexible hierarchical block unit representation such as the coding unit (CU), prediction unit (PU), and transform unit (TU). Each block unit can be used independently according to its quad-tree partition rules. The CU is used for region splitting for inter/intra-prediction, the PU for prediction processes and the TU for transform and quantization processes, regardless of the PU partitions.

HEVC exploits transform schemes in a frequency domain to compress the residual block between the original block and its matched prediction block. Transform converts the residual data in the spatial domain to the data in the frequency domain. An efficient transform model can compact more data being scattered in the spatial domain into the low frequency components in the frequency domain. To improve the transform coding performance, the DCT-II [3] or DST-VII [4] is used as the transform kernel in HEVC. The Karhunen–Loève transform (KLT) [5] is known to be the optimal transform in terms of data decorrelation. However, KLT is not used in practical transform coding because of its high complexity. Because DCT-II and DST-VII provide good approximations of KLT, they replace KLT in practical transform coding [6], [7]. DCT-II is extensively used to improve the coding efficiency in video compression fields and to research an image in the image-processing fields [8], [9]. The Hadamard transform is also used in HEVC for fast mode decision processing [10].

In this study, after we rearrange the data of the residual block, DST-VII is applied to the TU of the

Manuscript received Nov. 25, 2016; revised Feb. 23, 2017; accepted Mar. 8, 2017.

Jun-woo Choi (jwchoi@sju.ac.kr), Nam-Uk Kim (nukim@sju.ac.kr), and Yung-Lyul Lee (corresponding author, yllee@sejong.ac.kr) are with the Computer Engineering Department, Sejong University, Seoul, Rep. of Korea.

Sung-Chang Lim (sclim@etri.re.kr), Jungwon Kang (jungwon@etri.re.kr), and Hui Yong Kim(hykim5@etri.re.kr) are with the Broadcasting Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

inter-predicted PU, which is named shuffled DST-VII (S-DST). This is performed because the rearrangement of the residual block data is similar to the arrangement of the basis vector with a low frequency component of the DST-VII. Based on our analysis of the residual data distribution in the inter-predicted PU, DST-VII has more efficient energy compaction than does DCT-II in the inter-predicted PU when the residual block data are rearranged for DST-VII.

The remainder of this paper is organized as follows. Section II reviews related studies regarding the improvement of transform coding. Section III presents our analysis of residual data distribution in the inter-predicted PU. In Section IV, the proposed S-DST, which is designed from the analytical results in Section III, is described. Section V presents experimental results and Section VI concludes the study.

II. Related Work

1. Review of DCT-II and DST-VII

The one-dimensional (1-D) DCT-II is used for transforms of 4 × 4, 8 × 8, 16 × 16, and 32 × 32 residual blocks, except for the 4 × 4 luma intra-prediction residual block in HEVC [11], [12]. The N-point DCT-II is defined by (1) as follows.

$$v_i = \sum_{j=0}^{N-1} u_j \frac{P}{\sqrt{N}} \cos\left[\frac{\pi}{N}\left(j+\frac{1}{2}\right)i\right] \begin{cases} P=1 & i=0 \\ P=\sqrt{2} & i>0 \end{cases}, \quad (1)$$

$$c_{ij} = \frac{P}{\sqrt{N}} \cos\left[\frac{\pi}{N}\left(j+\frac{1}{2}\right)i\right], \quad (2)$$

where v_i and u_j are the transformed coefficient and input sample at each position i and j, respectively, and c_{ij} in (2), which is part of (1), is the DCT-II kernel. Specifically, v_i is the basis vector of the DCT-II kernel in terms of u_j, where i, j = 0, 1, 2, … , N – 1. v_i can be represented as a linear combination of c_{ij} and u_j in Fig. 1 when N = 4.

Figure 1 reveals that the DCT-II generates good transform coding efficiency when the residual data distribution in the transform block is flat or little difference exists between neighboring pixels [13].

The 1-D DST-VII is simply applied to the 4 × 4 luma intra-prediction residual blocks in HEVC because the residual block data following intraD-prediction is similar to the arrangement of the basis vector v_0 of DST-VII as shown in Fig. 2.

The N-point DST-VII is defined in (3), and c_{ij} is defined in (4) as follows.

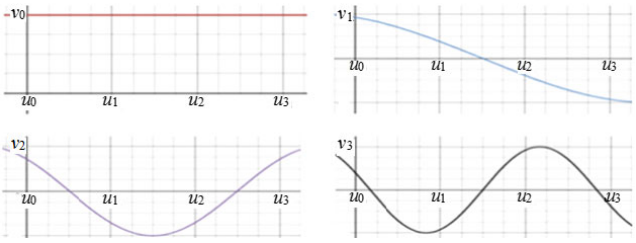


Fig. 1. Basis vectors of DCT-II for input samples.

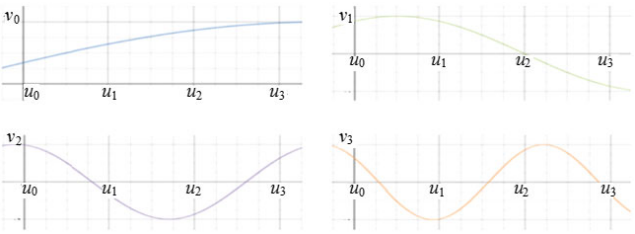


Fig. 2. Basis vectors of DST-VII for input samples.

$$v_i = \sum_{j=0}^{N-1} u_j \frac{2}{\sqrt{2N+1}} \sin\left(\frac{(2i+1)(j+1)\pi}{2N+1}\right), \quad (3)$$

$$c_{ij} = \frac{P}{\sqrt{2N+1}} \sin\left(\frac{(2i+1)(j+1)\pi}{2N+1}\right), \quad (4)$$

where symbols are denoted as those in (1) and (2): v_i is the basis vector of the DST-VII kernel in terms of u_j, i, j = 0, 1, 2, … , N – 1, and v_i is represented as a linear combination of c_{ij} and u_j in Fig. 2 when N = 4.

2. Improvement of Transform Coding Scheme

Many transform schemes have been proposed for improving transform coding efficiency. Some transform schemes attempt to exploit residual data distribution based on a specific intra-direction mode in order to improve efficiency. Ye and Korczewicz [14] proposed a mode-dependent directional transform (MDDT) and Yeo and others [15] improved MDDT by using a simple separable, directional, and anisotropic image correlation model. Zou and others [16] improved the performance of transform coding efficiency with intra-directional prediction residual data by using rate-distortion optimization. These transform methods exploit the characteristics of residual data distribution in intra-prediction. Other transform methods exploit the characteristics of residual data distribution in inter-prediction. When inter-prediction for a PU is performed, prediction errors are usually higher near the PU

boundaries than in the middle of the PU. An and others [17] and Fan and others [18] proposed boundary-dependent transforms for inter-predicted residual signals. They proposed additional transforms based on whether the TU is on the PU boundary. An and others [17] proposed adaptive boundary dependent transforms, such as DST-VII, DCT-IV, and DCT-II for HEVC. Similarly, Fan and others [18] proposed the DST-VII, DCT-IV, DCT-II, and flipped versions of the transforms on the PU.

In this study, common characteristics of the residual data distribution in an inter PU are investigated and the S-DST, which rearranges the residual block data and then applies the DST-VII, is designed to improve transform coding efficiency. The proposed S-DST only requires that the residual block data be rearranged before DST-VII is applied.

III. Residual Distribution in an Inter-predicted PU

To improve transform coding efficiency, applying a transform in the consideration of the characteristics of the residual block data is necessary. This section investigates common characteristics of the residual block data distribution in the inter-predicted PU and verifies that the DST-VII performance improves after the proposed S-DST is applied. Figure 3 shows the average absolute values of residual data at each position in the inter-predicted 8 × 8 PUs from "*Cactus*" sequence in a low delay P (LDP) configuration [19].

The high 30% and high 70% values are marked in grey in Figs. 3(a) and 3(b), respectively. These figures show that the residual absolute values near the PU boundary are higher than those near the PU center position. In other words, the absolute residual values increase from the PU center position to the PU boundary.

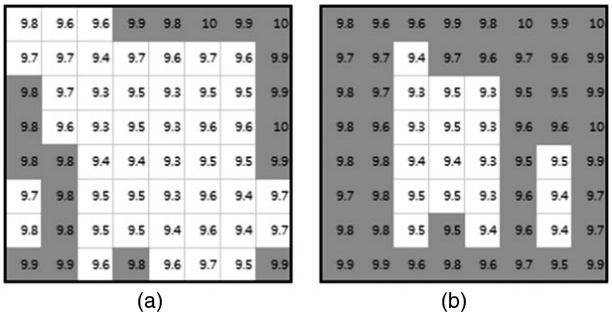


Fig. 3. Average absolute values of residual data at each position in inter-predicted 8 × 8 PUs in a "*Cactus*" sequence: (a) high 30% values and (b) high 70% values are marked in grey.

An analysis of the residual distribution in the inter-predicted PU reveals that the *N*-point DCT-II is not an efficient transform of the residual data in the inter-predicted PU. This is because the *N*-point DCT-II usually provides good performance when the residual distribution is flat like the v_0 basis vector (DC component), as shown in Fig. 1. By contrast, the *N*-point DST-VII is efficient for the residual data distribution having the shape of the v_0 basis vector (low frequency component), as shown in Fig. 2. In the proposed S-DST, the residual data of the inter-predicted PU are rearranged to fit the residual data distribution to the shape of the v_0 basis vector of the DST-VII, as shown in Fig. 2. The DST-VII is then applied. The proposed S-DST is described in the following section.

IV. Proposed S-DST

The proposed S-DST is developed to improve transform coding efficiency for each TU in an inter-predicted PU. The proposed S-DST is performed in two phases. First, the residual data are rearranged to fit the residual data distribution in a PU to the shape of the v_0 basis vector in the DST-VII of Fig. 2. Second, the DST-VII is performed on the rearranged residual block data.

1. Rearrangement of Residual Data in an Inter-Predicted PU

The HEVC standard includes $2N \times 2N$, $2N \times N$, $N \times 2N$, $N \times N$, $2N \times n$U, $2N \times n$D, nL $\times 2N$, and nR $\times 2N$ inter-predicted PUs in a $2N \times 2N$ CU. DCT-II is applied in quad-tree partitions from $2N \times 2N$ TU to 4×4 TU, which is the minimum TU size, regardless of the PU partitions. Therefore, the shuffling process is explained only with respect to a $2N \times 2N$ PU.

The residual data in a PU is divided into four groups. Figure 4 shows the four-group classification of the residual data in a $2N \times 2N$ PU. A quad-tree partitioning structure of TU depth 1 is used, but excludes the residual data rearrangement. The arrows from the center position indicate the directions of average absolute values increase of the residual data shown in Fig. 3.

Figure 4 shows each group of residual data in a $2N \times 2N$ PU, in which *Group*0 (*G*0), *Group*1 (*G*1), *Group*2 (*G*2), and *Group*3 (*G*3) consist of the left-top, right-top, left-bottom, and right-bottom $N \times N$ residual blocks in the PU, respectively. After the residual data in the $2N \times 2N$ PU are classified into four groups, each $N \times N$ residual data in each group except *G*3 is rearranged before transform. Because the absolute values of the residual data

https://doi.org/10.4218/etrij.17.0116.0867

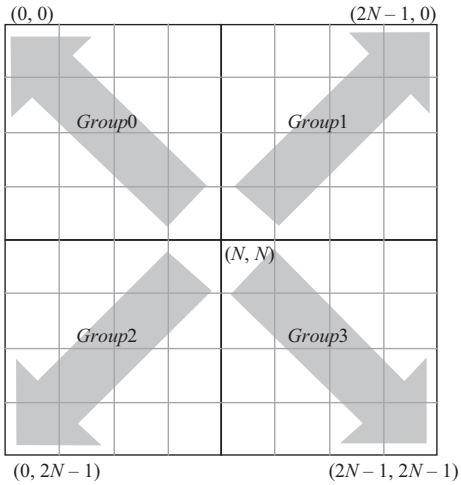


Fig. 4. Grouping of the residual data distribution according to the residual data value increment in a $2N \times 2N$ PU: the residual data are rearranged in *Group*0, *Group*1, and *Group*2. Arrows indicate that the average absolute values of the residual data increment.

in each $N \times N$ group increase from the center (N, N) position as shown in Fig. 4 to the boundary as shown in Fig. 3, the residual data in each group are rearranged in a different direction to fit the residual data distribution to the shape of the v_0 basis vector (low frequency component) of DST-VII in one-dimensional (1-D) view, depending on the group in (5) through (8) as follows.

Rearrangement of the residual data in $G0$:

$$a'(x,y) = a(N-1-x, N-1-y). \tag{5}$$

Rearrangement of the residual data in $G1$:

$$b'(x,y) = b(x, N-1-y). \tag{6}$$

Rearrangement of the residual data in $G2$:

$$c'(x,y) = c(N-1-x, y). \tag{7}$$

Rearrangement of the residual data in $G3$:

$$d'(x,y) = d(x,y), \tag{8}$$

where $0 \le x < N, 0 \le y < N$.

Note that $a(x, y)$, $b(x, y)$, $c(x, y)$, and $d(x, y)$ indicate the $N \times N$ original residual data values in $G0$, $G1$, $G2$, and $G3$, respectively, and $a'(x, y)$, $b'(x, y)$, $a'(x, y)$, and $d'(x, y)$ are those rearranged in $G0$, $G1$, $G2$, and $G3$, respectively. N indicates the width and height of each group and (x, y) is the position of residual data in each group.

Figure 5 gives an example of the 4×4 residual data rearrangement in $G0$ when the PU size is 8×8.

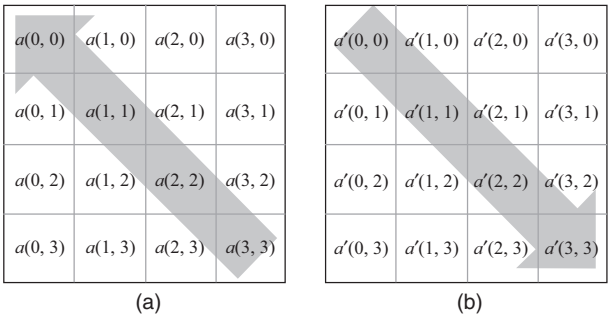


Fig. 5. Example of the 4×4 residual data rearrangement in G0 in an inter-predicted 8×8 PU. Arrows indicate that the average absolute values of the residual data increase.

Figure 5(a) shows the original residual data and Fig. 5(b) shows the rearrangement in $G0$ in the 8×8 PU.

The residual data values in $G0$ in Fig. 5(b) after rearrangement will gradually increase from the (0, 0) to the (3, 3) position. The horizontal and vertical 1-D view of the rearranged residual value data in each group have the residual data distribution of the shape of the v_0 basis vector shown in Fig. 2, which is the lowest frequency component of the DST-VII.

This shuffling process rearranges the $G0$, $G1$, $G2$, $G3$ residual data independently to fit their residual data distribution to the shape of the v_0 basis vector in DST-VII. After each group is shuffled, the DST-VII is applied to the rearranged data of each group independently.

In Fig. 4, when the TU depth is equal to 1, the $2N \times 2N$ PU is divided into four $N \times N$ blocks and residual data rearrangement is performed for each $N \times N$ $G0$, $G1$, $G2$, and $G3$ by using (5) through (8). After shuffling, DST-VII is applied to each $N \times N$ rearranged $G0$, $G1$, $G2$, and $G3$ residual data. If the TU depth is equal to 2, each $N \times N$ block in the $2N \times 2N$ PU is divided into four $N/2 \times N/2$ residual blocks in a quad-tree partition, and the similar rearrangement using (5) through (8) is applied to each $N/2 \times N/2$ residual block data by changing the width and height. For example, if the TU depth equals 2, four $N/2 \times N/2$ residual blocks in the $N \times N$ $G0$ residual block are rearranged according to (5) to achieve the high v_0 basis vector (low frequency component) shown in Fig. 2. The quad-tree-based TU partition continues until the minimum TU size is obtained (the minimum TU size is 4×4 when the CU size is 8×8 or 16×16; it is 8×8 when the CU size is 32×32; and it is 16×16 when the CU size is 64×64).

As a reference, when the TU depth equals 0, the $2N \times 2N$ DCT-II is applied without rearranging the residual data as in the HEVC standard.

2. Discrete Sine Transform Type-VII

The purpose of shuffling the residual data (that is, rearrangement into four groups) is to rearrange the residual data distribution to obtain the transform coding efficiency of the DST-VII. The residual shuffling process is based on the hypothesis that DST-VII achieves higher transform coding efficiency than does DCT-II when the residual data are rearranged to produce a higher magnitude in the shape of the v_0 basis vector (the lowest frequency component) shown in Fig. 2 than that of other basis vectors in terms of energy compaction. Because the DST-VII has a separable property, the 2-D DST-VII is performed by 1-D horizontal and vertical DST-VII.

Figure 6 gives an example of an 8 × 8 ($2N \times 2N$) PU having the residual data and its transformed blocks by the 4 × 4 DCT-II and 4 × 4 S-DST. Figures 6(a) through 6(c) show the residual distribution of the 8 × 8 inter-predicted PU, the coefficients distributions by each 4 × 4 DCT-II, and the coefficients distribution by each 4 × 4 DST-VII after shuffling the 4 × 4 blocks of Fig. 6(a), respectively. In this example, the proposed S-DST is better than DCT-II in terms of energy compaction into a low frequency component when the residual data are rearranged. When each 4 × 4 residual block is transformed by the 4 × 4 S-SDT, most transform coefficients are located at the lowest frequency component, whereas more high frequency components remain after the DCT-II is applied.

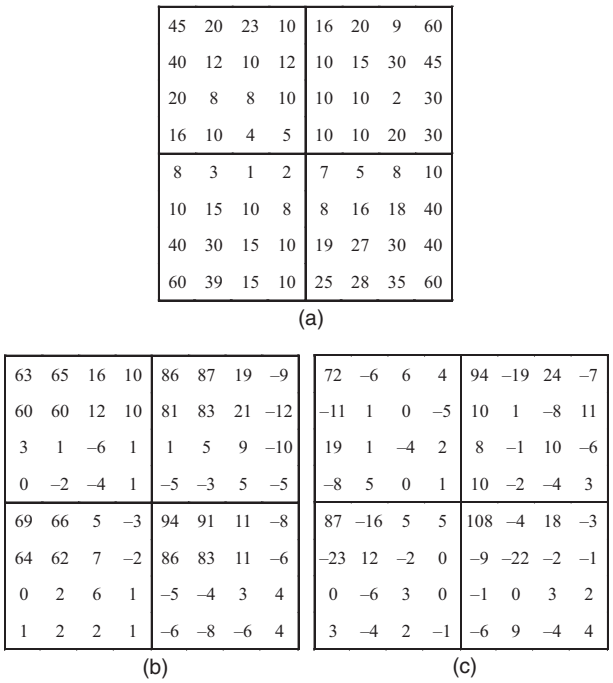


Fig. 6. Example of the 8 × 8 ($2N \times 2N$) PU: (a) 8 × 8 residual block, (b) its transformed blocks by 4 × 4 DCT-II, and (c) its transformed blocks by 4 × 4 S-DST.

In this experiment, the floating point DCT-II and DST-VII are scaled-up by 2^n for integer operation, in which n = 7, 7.5, 8, and 8.5 for 4-point DCT-II and DST-VII, 8-point DCT-II and DST-VII, 16-point DCT-II and DST-VII, and 32-point DCT-II and DST-VII, respectively.

The efficiency of transform coding derives from the difference in the frequency characteristics of the DST-VII and DCT-II. As shown in Figs. 1 and 2, DST-VII can achieve better transform coding results than DCT-II when the rearrangement of the residual block data is similar to the arrangement of the basis vector v_0 of DST-VII.

3. Implementation Details

The proposed S-DST, which rearranges the residual data of an inter-predicted PU to fit the residual data distribution to the shape of the v_0 basis vector in DST-VII and applies the DST-VII, uses the integer transform derived from the DST-VII by using its separable property [6]. DCT-II or S-DST is selected for each TU based on the rate-distortion (R-D) cost [20]. One signaling bit, *sdst_flag*, for each TU, indicating the use of the S-DST, is required to identify the selected transform.

Figure 7 shows the decoding process of a TU in an inter-predicted PU. If the TU depth in a CU is 0, then the S-DST is not used because the S-DST is performed from TU depth 1 to max TU depth. If the TU transform mode is skip or the coded block flag (cbf) of the TU equals 0, then the *sdst_flag* is not needed. If the *sdst_flag* in a TU is "1," then the inverse DST-VII is performed and the reconstructed residual data after inverse transform is rearranged into the original residual positions prior to shuffling. The inverse shuffling of the residual block associated with (5) through (8) is performed in the inverse equations. The shuffled residual block data are rearranged after the inverse quantization and DST-VII to the original residual block data as described in (9).

$$\begin{aligned} a(x,y) &= a'(N-1-x, N-1-y), \\ b(x,y) &= b'(x, N-1-y), \\ c(x,y) &= c'(N-1-x, y), \\ d(x,y) &= d'(x,y), \\ &\text{where } 0 \le x,y \le N-1, \end{aligned} \tag{9}$$

where $a(x, y)$, $b(x, y)$, $c(x, y)$, and $d(x, y)$ indicate the inverse shuffled residual data in $G0$, $G1$, $G2$, and $G3$ at position (x, y), respectively, and $a'(x, y)$, $b'(x, y)$, $c'(x, y)$, and $d'(x, y)$ are the shuffled residual data in $G0$, $G1$, $G2$, and $G3$, respectively.

https://doi.org/10.4218/etrij.17.0116.0867

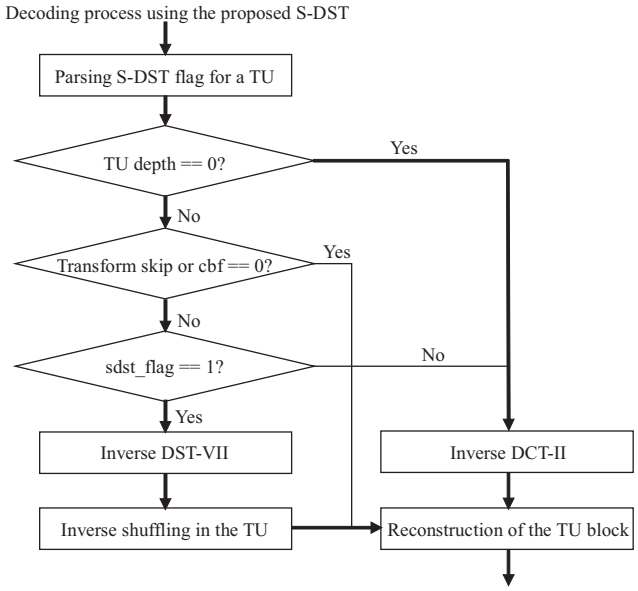


Fig. 7. Decoding process using the proposed S-DST.

If the *sdst_flag* in a TU is "0," then the normal inverse DCT-II is performed.

V. Experimental Results and Analysis

To verify the efficiency of the proposed S-DST, several experiments were performed on the HEVC reference software, HM16.6 [21]. The experimental conditions were the same as the common test conditions in the HEVC test using random access (RA), LDP, and low delay B (LDB) [19]. In the RA configuration, hierarchical B pictures (IBBBBBBBP), which have a group of pictures (GOP) size of 8, is used. The coding efficiency achieved by the bi-directional hierarchical prediction structure was higher than in the other configurations. However, it experienced a longer delay as a result of the picture reordering. In the LDP configuration, the first picture was encoded as I and subsequent pictures were encoded as P. The P pictures in the LDB were generalized P and B pictures (GPBs), in which the P pictures were replaced by B pictures having the same two reference pictures. In the LDB configuration, the first picture was encoded as I and subsequent pictures were encoded as B. In general, higher coding efficiency was achieved in the LDB than in the LDP configuration because of its bi-prediction structure. Table 1 lists the coding tools, which are used in RA, LDB, and LDP configurations. Not only a flexible coding structure, such as CU and TU, but also various coding tools, such as CABAC, SAO, and a deblocking filter are used to improve the coding efficiency of HM16.6. All test conditions are listed in Table 2. The proposed S-DST was applied to the TUs of luma and chroma components (Cb, Cr) in an inter-predicted CU, except at a TU depth of 0.

Table 1. Coding tools used in RA, LDB, and LDP configurations.

RA	LDB(P)
Bi-directional hierarchical prediction structure	First picture: I picture subsequent pictures: B(P) pictures
Coding unit 8 × 8 up to 64 × 64 in tree structure	
Prediction units	
Transform unit tree (3 level max)	
Transform block size of 4 × 4 to 32 × 32	
8-bit depth	
Entropy coding: CABAC	
Sample adaptive offset (SAO), deblocking filter	

Table 2. Experimental conditions in the HM16.6.

Test sequence	All sequences in classes A to F
Profile	LDP, LDB, RA
Reference codec	HM16.6
QP	22, 27, 32, 37

Table 3 lists the average selection percentage (%) of the DCT-II and S-DST for luma and chroma components at each quantization parameter (QP) of the video sequences in the RA configuration. The table indicates that the selection percentage of S-DST was similar to that of the DCT-II at each *QP*. The percentage of S-DST selection rate was even greater than that of DCT-II in some sequences. Because the experimental results in LDB and LDP configurations were similar to those of the RA configuration, the experimental results in the RA configuration are shown. Table 4 shows the coding efficiency according to the BD-rate measurement [22] in which the negative sign of the BD rate refers to the bit saving of the proposed method compared with HM16.6 in the same PSNR reference. The BD-rate results of the proposed S-DST combined with DCT-II were compared with those of HM16.6. Table 4 lists that the coding efficiency of the proposed S-DST combined with the DCT-II was better than that of HM 16.6, especially in the high resolutions sequences such as Class A and B and the natural video sequences except those of Class F.

The proposed method reduces the luma/chroma (Cb+Cr) BD rates by approximately 0.23% to 0.22%, 0.44% to 0.58%, and 0.46% to 0.65% for RA, LDB, and LDP

Table 3. Selection % of DCT-II and S-DST at QP = 22, 27, 32, 37 in each sequence in the RA configuration.

Class	Sequence	QP = 22		QP = 27		QP = 32		QP = 37	
		DCT-II (%)/ S-DST (%)		DCT-II (%)/ S-DST (%)		DCT-II (%)/ S-DST (%)		DCT-II (%)/ S-DST (%)	
		Luma	Cb+Cr	Luma	Cb+Cr	Luma	Cb+Cr	Luma	Cb+Cr
Class A 4 K (2,560 × 1,600)	Traffic	49/51	50/50	48/52	48/52	48/52	46/54	49/51	68/32
	PeopleOnStreet	47/53	49/51	48/52	49/51	47/53	44/56	51/49	60/40
	Nebuta	44/56	51/49	49/51	48/52	46/54	48/52	50/50	58/42
	SteamLocomotive	44/56	52/48	49/51	48/52	45/55	44/56	49/51	65/35
Class B 1,080 p (1,920 × 1,080)	Kimono	45/55	47/53	53/47	57/43	51/49	42/58	49/51	65/35
	ParkScene	41/59	45/55	53/47	55/45	48/52	37/63	52/48	62/38
	Cactus	39/61	48/52	53/47	51/49	46/54	35/65	51/49	56/44
	BasketballDrive	38/62	50/50	53/47	55/45	43/57	36/64	52/48	63/37
	BQTerrace	49/51	50/50	53/47	53/47	48/52	50/50	53/47	45/55
Class C WVGA (832 × 480)	BasketballDrill	41/59	54/46	50/50	56/44	49/51	47/53	54/46	47/53
	BQMall	40/60	49/51	48/52	55/45	50/50	54/46	55/45	47/53
	PartyScene	44/56	54/46	47/53	61/39	48/52	38/62	51/49	41/59
	RaceHorses	45/55	51/49	52/48	43/57	50/50	54/46	58/42	54/46
Class D WQVGA (416 × 240)	BasketballPass	50/50	64/36	51/49	43/57	50/50	51/49	57/43	53/47
	BQSquare	51/49	61/39	52/48	40/60	48/52	53/47	56/44	53/47
	BlowingBubbles	56/44	77/23	50/50	43/57	46/54	50/50	57/43	53/47
	RaceHorses	47/53	46/54	51/49	50/50	48/52	43/57	71/29	56/44
Class E 720 p (1,280 × 720)	FourPeople	44/56	42/58	49/51	47/53	46/54	40/60	71/29	63/37
	Johnny	42/58	41/59	48/52	50/50	46/54	42/58	67/33	60/40
	KristenAndSara	44/56	46/54	46/54	52/48	43/57	42/58	61/39	69/31
Class F	BasketballDrillText	47/53	46/54	51/49	51/49	49/51	65/35	58/42	54/46
	ChinaSpeed	45/55	45/55	50/50	48/52	48/52	61/39	57/43	57/43
	SlideEditing	43/57	49/51	49/51	49/51	47/53	51/49	58/42	55/45
	SlideShow	42/58	47/53	45/55	47/53	46/54	56/44	59/41	57/43
Average	Class A	46/54	51/50	49/52	48/52	47/54	46/55	50/50	63/37
	Class B	42/58	48/52	53/47	54/46	47/53	40/60	51/49	58/42
	Class C	43/58	52/48	49/51	54/46	49/51	48/52	55/46	47/53
	Class D	51/49	62/38	51/49	44/56	48/52	49/51	60/40	54/46
	Class E	43/57	43/57	48/52	50/50	45/55	41/59	66/34	64/36
	Class F	44/56	47/53	49/51	49/51	48/53	58/42	58/42	56/44
	All (except class F)	**45/55**	**51/49**	**50/50**	**50/50**	**47/53**	**45/55**	**56/44**	**57/43**
	All	45/55	51/49	50/50	50/50	47/53	47/53	56/44	57/43

configurations, respectively. More specifically, the average BD rates of the luma-chromas (Cb+Cr) in the Class B sequences was reduced by 0.34% to 0.40%, 0.54% to 1.18%, and 0.50% to 1.32% for RA, LDB, and LDP configurations, respectively.

The proposed S-DST was applied from TU depth 1 to max TU depth. As a reference, if the DST-VII is alternatively used with DCT-II only in TU depth 0 and the proposed S-DST is used from TU depth 1, the average BD rates of the luma/chroma (Cb+Cr) improve by 0.15% to 0.2% compared with the proposed S-DST. In other words, the proposed S-DST can improve coding efficiency even more if it is combined with other transforms.

Tables 5 and 6 show the geometric mean of the encoding and decoding times that compare HM16.6 with the proposed S-DST combined with DCT-II for all sequences. The encoding time increased by as much as 119% with the LDP configuration. The encoding time increment was

https://doi.org/10.4218/etrij.17.0116.0867

Table 4. BD-rates of the proposed S-DST compared with HM16.6 in the RA, LB, and LP configurations.

Class	Sequence	BD Rate								
		RA			LDB			LDP		
		Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
Class A (2,560 × 1,600)	Traffic	−0.17%	−0.61%	−0.27%	−0.43%	−1.41%	−1.06%	−0.58%	−1.66%	−0.82%
	PeopleOnStreet	−0.57%	−0.47%	−0.61%	−0.72%	−0.91%	−0.98%	−0.76%	−0.89%	−0.70%
	Nebuta	−0.38%	−1.44%	−1.68%	−0.27%	−3.83%	−3.86%	−0.15%	−2.82%	−3.06%
	SteamLocomotive	−0.15%	−0.01%	−0.56%	−0.29%	1.01%	−1.09%	−0.41%	−0.10%	−0.24%
Class B 1,080 p (1,920 × 1,080)	Kimono	−0.46%	−0.76%	−0.77%	−0.75%	−2.96%	−1.98%	−0.63%	−1.73%	−2.09%
	ParkScene	−0.30%	−0.80%	−0.40%	−0.58%	−2.65%	−2.00%	−0.57%	−2.18%	−1.58%
	Cactus	−0.22%	−0.38%	−0.26%	−0.45%	−1.12%	−1.19%	−0.46%	−1.17%	−0.98%
	BasketballDrive	−0.25%	−0.03%	−0.04%	−0.32%	0.53%	−0.04%	−0.30%	−0.17%	−0.23%
	BQTerrace	−0.45%	−0.43%	−0.05%	−0.62%	−1.43%	1.04%	−0.56%	−1.83%	−1.23%
Class C WVGA (832 × 480)	BasketballDrill	−0.08%	−0.08%	−0.25%	−0.48%	−0.44%	−0.56%	−0.39%	−1.24%	−0.20%
	BQMall	−0.20%	0.09%	−0.08%	−0.49%	−0.85%	−0.25%	−0.51%	−0.90%	−0.50%
	PartyScene	−0.23%	−0.14%	−0.11%	−0.42%	−0.66%	−0.55%	−0.46%	−0.55%	−0.61%
	RaceHorses	−0.29%	−0.34%	−0.30%	−0.43%	−1.29%	−0.98%	−0.45%	−1.65%	−0.99%
Class D WQVGA (416 × 240)	BasketballPass	−0.14%	0.26%	−0.16%	−0.30%	−0.75%	−0.42%	−0.16%	0.07%	−0.28%
	BQSquare	−0.23%	0.41%	0.64%	−0.37%	0.78%	2.44%	−0.45%	1.42%	0.95%
	BlowingBubbles	−0.26%	−0.22%	−0.06%	−0.55%	−0.86%	−0.25%	−0.53%	−0.85%	−0.81%
	RaceHorses	−0.26%	−0.34%	−0.44%	−0.30%	−0.27%	−0.29%	−0.44%	0.21%	−0.34%
Class E 720 p (1,280 × 720)	FourPeople	−0.01%	0.20%	0.24%	−0.33%	−0.67%	0.10%	−0.42%	−1.26%	−0.99%
	Johnny	0.04%	0.39%	0.54%	−0.37%	0.77%	0.72%	−0.42%	1.17%	1.55%
	KristenAndSara	−0.02%	0.28%	0.18%	0.02%	−0.48%	0.15%	−0.18%	0.19%	−0.41%
Class F	BasketballDrillText	−0.11%	0.04%	−0.04%	−0.32%	−0.61%	−0.36%	−0.36%	0.18%	0.33%
	ChinaSpeed	−0.07%	0.03%	−0.06%	−0.16%	−0.10%	−0.56%	−0.12%	−0.11%	−0.71%
	SlideEditing	0.04%	0.07%	0.09%	−0.31%	−0.30%	−0.26%	0.41%	0.25%	0.31%
	SlideShow	−0.13%	0.03%	−0.16%	−0.55%	0.27%	−0.15%	−0.79%	−1.77%	−1.05%
Average	Class A	−0.32%	−0.63%	−0.78%	−0.43%	−1.29%	−1.75%	−0.47%	−1.37%	−1.21%
	Class B	**−0.34%**	**−0.48%**	**−0.31%**	**−0.54%**	**−1.53%**	**−0.83%**	**−0.50%**	**−1.42%**	**−1.22%**
	Class C	−0.20%	−0.12%	−0.18%	−0.46%	−0.81%	−0.59%	−0.45%	−1.09%	−0.57%
	Class D	−0.22%	0.03%	−0.01%	−0.38%	−0.27%	0.37%	−0.40%	0.21%	−0.12%
	Class E	0.00%	0.29%	0.32%	−0.23%	−0.13%	0.32%	−0.34%	0.03%	0.05%
	Class F	−0.07%	0.04%	−0.04%	−0.33%	−0.18%	−0.33%	−0.21%	−0.36%	−0.28%
	All (except class F)	**−0.23%**	**−0.22%**	**−0.22%**	**−0.44%**	**−0.82%**	**−0.34%**	**−0.46%**	**−0.72%**	**-0.57%**
	All	−0.20%	−0.18%	−0.19%	−0.41%	−0.76%	−0.52%	−0.40%	−0.72%	−0.61%

derived from the RD cost computations between the DCT-II and S-DST, whereas the decoding time slightly increased because the decoder simply performed the DCT-II or S-DST according to the *sdst_flag*. The proposed method achieved coding gains for RA, LDB, and LDP configurations with a small amount of computational complexity increment. The improvement in coding efficiency was generally less than 1%, with the encoding and decoding times increasing by as much as 19%.

VI. Conclusion

In this study, we proposed the S-DST to improve the transform coding performance in HEVC. Because the absolute values of residual data in the inter-predicted PU boundary are higher compared to those of the PU center position, the S-DST, which rearranges the residual block data and then applies the DST-VII, was designed to improve transform coding efficiency. Because the

Table 5. Comparisons of computational complexity with HM16.6 and S-DST in the RA, LDB, and LDP configurations in the encoder (in hour units).

RA			LDB			LDP		
HM16.6	S-DST	Encoding time increment	HM16.6	S-DST	Encoding time increment	HM16.6	S-DST	Encoding time increment
0.48 h	0.56 h	116%	0.51 h	0.58 h	113%	0.35 h	0.42 h	119%

Table 6. Comparisons of the computational complexity with HM16.6 and S-DST in the RA, LDB, and LDP configurations in the decoder (in second units).

RA			LDB			LDP		
HM16.6	S-DST	decoding time increment	HM16.6	S-DST	decoding time increment	HM16.6	S-DST	decoding time increment
2.91 s	3.01 s	104%	3.26 s	3.30 s	101%	3.02 s	3.07 s	102%

rearrangement of the residual data fit the residual distribution into the shape of the v_0 basis vector shown in Fig. 2, the DST-VII improved coding efficiency.

Experimental results reveal that the proposed S-DST improved the coding performance for HM 16.6. When the proposed S-DST was combined with other transforms while considering residual characteristics, the transform coding efficiency improved.

In addition, we believe that the proposed S-DST can be combined with a coefficient scanning method such as the characteristics distribution of transform and quantization [23]. This topic remains for a future study.

Acknowledgements

This work was in part supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (2016-000572, Development and Standardization of 5th Generation Video/Audio Coding Technology for Ultra High Quality Media Services). This research was in part supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science, ICT and Future Planning) (NRF-2015R1A2A2A01006085).

References

[1] B. Bross et al., "High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & consent)," Proc. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-L1003, Geneva, Switzerland, Jan. 2013.

[2] G.J. Sullivan, et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1649–1668.

[3] N. Ahmed, T. Natarajan, and K.R. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, Jan. 1974, pp. 90–93.

[4] A. Saxena and F. Fernandes, "CE7: Mode-Dependent DCT/DST without Full Matrix Multiplication for Intra Prediction," in Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-E125, Geneva, Switzerland, Mar. 2011.

[5] R.M. Noor and T. Vladimirova, "Integer KLT Design Space Exploration for Hyperspectral Satellite Image Compression," *Proc. Int. Conf. Convergence Hybrid Inform. Technol.*, Daejeon, Rep. of Korea, Sept. 22–24, 2011, pp. 661–668.

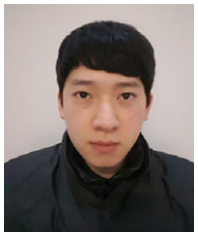
[6] A.K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1989, pp. 132–188.

[7] Y. Kang et al., "High-Performance 4 × 4 Intra Prediction Hardware for High-Efficiency Video Coding," *Electron. Lett.*, vol. 51, no. 1, Jan. 2015, pp. 35–36.

[8] K.H. Kim et al., "Facial Feature Extraction Based on Private Energy Map in DCT Domain," *ETRI J.*, vol. 29, no. 2, Apr. 2007, pp. 243–245.

[9] A.G. Yepes et al., "Harmonic Identification Algorithms Based on DCT for Power Quality Applications," *ETRI J.*, vol. 32, no. 1, Feb. 2010, pp. 33–43.
[10] Y. Kim et al., "A Fast Intra-Prediction Method in HEVC Using Rate-Distortion Estimation Based on Hadamard Transform," *ETRI J.*, vol. 35, no. 2, Apr. 2013, pp. 33–43.
[11] A. Saxena and F.C. Fernandes, "DCT/DST-Based Transform Coding for Intra Prediction in Image/Video Coding," *IEEE Trans. Image Proc.*, vol. 22, no. 10, Oct. 2013, pp. 3974–3981.
[12] M. Budagavi et al., "Core Transform Design for the High Efficiency Video Coding (HEVC) Standard," *IEEE J. Sel. Topics Signal Proc.*, vol. 7, no. 6, Dec. 2013, pp. 1029–1041.
[13] K.R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Boston, MA, USA: Academic Press, 1990, pp. 27–47.
[14] Y. Ye and M. Karczewicz, "Improved H.264 Intra Coding Based on Bi-Directional Intra Prediction, Directional Transform, and Adaptive Coefficient Scanning," *IEEE Int. Conf. Image Proc.*, San Diego, CA, USA, Oct. 2008, pp. 2116–2119.
[15] C. Yeo et al., "Mode-Dependent Transforms for Coding Directional Intra Prediction Residuals," *IEEE Trans. Circuits Syst.*, vol. 22, no. 4, Apr. 2012, pp. 545–554.
[16] F. Zou et al., "Rate-Distortion Optimized Transforms Based on the Lloyd-Type Algorithm for Intra Block Coding," *IEEE J. Sel. Topic Signal Proc.*, vol. 7, no. 6, Dec. 2013, pp. 1072–1083.
[17] J. An et al., "Boundary-Dependent Transform for Inter-Predicted Residue," JCT-VC Document, JCTVC-G281, Geneva, Switzerland, Nov. 2011.
[18] J. Fan et al., "Adaptive Boundary Dependent Transform Optimization for HEVC," *Picture Coding Symp.*, Cairns, Australia, May 31–June 3, 2015, pp. 149–153.
[19] F. Bossen, "Common HM Test Conditions and Software Reference Configurations," JCT-VC Document, JCTVC-L1100, Geneva, Switzerland, Jan. 2013.
[20] A. Ortega and K. Ramchandran, "Rate–Distortion Methods for Image and Video Compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, Nov. 1998, pp. 23–50.
[21] HM Reference Software, Accessed 2016. https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/tags/HM-16.6/
[22] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-Curves," ITU-T VCEG Meeting, Tech. Rep. SG 16 Q.6 Doc., VCEG-M33, Austin, TX, USA, Apr. 2001.
[23] Y.-L. Lee et al., "Adaptive Scanning for H.264/AVC Intra Coding," *ETRI J.*, vol. 28, no. 5, Oct. 2006, pp. 668–671.

Jun-woo Choi received his BS degree in Computer engineering from Sejong University, Seoul, Rep. of Korea, in 2015. He is currently pursuing his MS degree in computer engineering from Sejong University. His current research interests include image and video compression, image processing, and future video coding technologies.



Nam-Uk Kim received his BS degree in Computer engineering from Sejong University, Seoul, Rep. of Korea, in 2015. He is currently pursuing a joint BS and PhD degree in Computer Engineering from Sejong University. His current research interests include video compression and future video coding technologies.

Sung-Chang Lim received his BS (with highest honor) and MS degrees in computer engineering from Sejong University, Seoul, Rep. of Korea, in 2006 and 2008, respectively. Since 2008, he has been a senior researcher in Broadcasting and Media Research Laboratory of ETRI, Daejeon, Rep. of Korea. His research interests include video coding and image processing.

Jung-Won Kang received her BS and MS degrees in electrical engineering in 1993 and 1995, respectively, from Korea Aerospace University, Seoul, Rep. of Korea. She received her PhD degree in electrical and computer engineering in 2003 from the Georgia Institute of Technology, Atlanta, GA, USA. Since 2003, she has been a principal member of the research staff in the Broadcasting Media Research Laboratory, ETRI, Daejeon, Rep. of Korea. Her research interests are in the areas of video signal processing and video compression.

Hui Yong Kim received his BS, MS, and PhD degrees from Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 1994, 1998, and 2004, respectively. From 2003 to 2005, he was the leader of the Multimedia Research Team of AddPac Technology Co. Ltd., Seoul, Rep. of Korea. In November 2005, he joined the Broadcasting and Telecommunications Media Research Laboratory of ETRI, Daejeon, Rep. of Korea, and currently serves as the Managing Director of Realistic AV Research Group. From 2006 to 2010, he was also an affiliate professor at the University of Science and Technology (UST), Daejeon, Rep. of Korea. From September 2013 to August 2014, he was a visiting scholar at the Media Communications Lab of the University of Southern California (USC), Los Angeles, USA. He has made many contributions to the development of international standards such as MPEG Multimedia Application Format (MAF) and JCT-VC High Efficiency Video Coding (HEVC), as an active technology contributor, editor, and ad-hoc group co-chair in many areas. His current research interest is in image, video, and audio signal processing and compression for realistic media applications such as UHD, 3D, VR and HDR.

Yung-Lyul Lee received his BS and MS degrees in electronic engineering from Sogang University, Seoul, Rep. of Korea, in 1985 and 1987, respectively, and his PhD degree in electrical and electronic engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1999. He was a principal researcher at Samsung Electronics R&D center, Seoul, Rep. of Korea from 1987 to 2001. He has been a professor with the Department of Computer Engineering, Sejong University, Seoul, Rep. of Korea, since 2001. He was a visiting scholar at the University of Texas at Arlington, USA from September 2006 to August 2007. He has contributed many documents to AVC/H.264 and High Efficiency Video Coding (HEVC) standards. Prof. Lee received a Minister Prize from the Ministry of Commerce, Industry and Energy in Korea in November 2006, and a Korea Science Technology Superiority Paper Prize in October 2006. He is now a vice-chairman of the Korea Institute of Broadcasting and Media Engineering. As an IEEE senior member, his current research interests include video compression, image processing, 3-D video coding, and multimedia systems.

https://doi.org/10.4218/etrij.17.0116.0867