Adaptive Transform with HEVC Intra Coding

Tao Zhang*, Haoming Chen†, Ming-Ting Sun†, Ankur Saxena**, Debin Zhao* and Wen Gao*
*Harbin Institute of Technology, Harbin, China
E-mail: taozhang.hit@hotmail.com
†University of Washington, Seattle, USA
E-mail: sun@ee.washington.edu
**Samsung Research America, Richardson, USA
E-mail: a.saxena1@samsung.com

Abstract— This paper proposes a content adaptive transform based on singular value decomposition (SVD) with HEVC intra coding. In the latest video coding standard HEVC, a combination of Discrete Cosine Transform (DCT) and Discrete Sine Transform (DST) is adopted to transform the residuals from intra prediction. The DCT and DST are fixed transforms which are derived based on the Gauss-Markov model. Since the residual blocks contain various distributions which may not be compacted well by the fixed DCT/DST, we propose an adaptive transform method based on SVD which is adaptive to residuals for angular intra coding. Experimental results show that the proposed method can achieve rate-distortion performance improvements compared to HEVC intra coding (up to 1.9% bit rate savings).

I. INTRODUCTION

Intra coding is essential for video coding. Current HEVC [1] employs a quadtree structure in which the size of the coding unit (CU) can be from 8×8 to 64×64. Transform unit (TU) is the basic unit in the transform which is also generated by the quadtree structure. In the HEVC angular intra prediction (AIP), there are 35 intra modes in which 33 modes have angular information which are efficient to predict blocks with directional patterns and the other 2 modes, which are DC mode and planar mode, are used to predict smooth blocks.

The efficiency of intra coding is significantly improved by the large number of intra modes and quadtree partition structures. In this paper, we will focus on the transform in intra coding which is an important part for improving the intra coding efficiency.

The Karhunen–Loeve transform (KLT) is the optimal transform for decorrelation if only one fixed transform can be used for all the coding blocks. In [2], DCT has been proved to be close to KLT in terms of energy compaction for coding blocks without any prediction. However, with the highly accurate prediction, the residual block to be transformed after prediction does not contain much correlations compared to the original coding block. In this case, DCT may not be optimal any more for transforming the residual blocks. According to the statistical distribution of the residuals from intra prediction, many new transforms have been proposed.

In [3], mode-dependent directional transform (MDDT) for H.264/AVC intra coding was proposed. In MDDT, 9 different transforms which are trained from an off-line dataset are used for the intra residual transform. DST [4] was proposed to transform residuals in intra coding. In [5], a set of transforms are trained. The best transform for the residual is chosen from these transforms through the RDO (Rate-Distortion Optimization) process. This kind of transforms are dependent on the training dataset which may not be well to capture various residuals in the coding process.

All the above mentioned transforms are fixed transforms which are not adaptive to the coding content. Several adaptive transforms were proposed to improve the efficiency of the fixed transforms. In [6], an image-coding algorithm which combines DCT and the singular value decomposition (SVD) was proposed. In this hybrid transform, either DCT or SVD is used to transform an image block. SVD transform was calculated from the current coding block and the transform coefficients which are the eigenvalues can also be derived through the SVD process. Then, a small number of eigenvalues and the derived eigenvectors are encoded, and vector quantization is used to encode these eigenvectors. The results verified this method could get some coding gain improvement compared to the method which only uses DCT. In [7], an adaptive transform based on SVD was proposed for H.264/AVC intra coding. A residual block was divided into four equal-sized sub-blocks. The first sub-block was transformed by regular DCT, and the other three sub-blocks were transformed by the transform matrices calculated from the reconstructed first sub-block. Another transform method based on SVD was proposed in [8] to decorrelate the inter prediction residuals. The transform for the current residual was derived from the prediction block in the reference frame considering there exists some similarity between the residual block and the original block because of imperfect motion estimation/compensation, and the original block can be approximated by the prediction block.

In this paper, we propose a content adaptive transform coding framework based on HEVC intra coding considering structure similarity between the residual block and the corresponding prediction block, and the similarity between prediction block and original block. Based on the analysis of the intra prediction residuals and inspired by the work in [6-8], a transform based on the prediction block, which we call P-SVD is proposed to improve the current fixed DCT/DST transform in HEVC angular intra coding.

The rest of the paper is organized as follows. Section II gives the analyses for the residuals from intra prediction and the derivation for the proposed transform. The coding method for the proposed adaptive P-SVD is described in Section III in details. The simulation results are presented in Section IV, and concluding remarks are given in Section V.

II. PROPOSED ADAPTIVE TRANSFROM

A. *Intra Prediction Residual and Transform Analyses*

The DCT/DST transforms are derived for the Gauss Markov model which is a statistical model. The two dimensional Gauss Markov model [4] can be expressed by

$$x_{i,j} = \rho x_{i-1,j} + \rho x_{i,j-1} - \rho^2 x_{i-1,j-1} + e_{i,j} \tag{1}$$

where $x_{i,j}$ denote the pixel in the *ith* row and *jth* column in a coding block, and is assumed to have zero mean and unit variance. ρ is the correlation coefficient between pixels. $e_{i,j}$ are independent identically distributed Gaussian random variables. With this model, the best combinations of DCT and DST for intra prediction residual transform can be derived [4].

According to our observations of the prediction residual blocks, there are large amount of residual blocks with structural information, and the residual distributions do not follow the Gauss-Markov model. For example, some blocks have vertical textures and can be predicted well by the vertical mode in the intra prediction. After the intra prediction, the residual is more or less random, but often still exists some vertical textures in which is not conforming to the Gauss-Markov model. Fig. 1 shows three residual examples which are chosen from the 4×4 residual blocks predicted by the vertical mode in HEVC intra coding. In these cases, the DCT/DST cannot compact the energy of the residual block well. Other transforms which are adaptive to the residual blocks could handle this problem of the DCT/DST.

1	0	-2	-2
3	0	-2	-6
1	0	-1	-2
1	0	-3	0

-2	-3	-4	-2
-4	-3	-2	-1
-4	-3	-2	-1
-4	-4	-2	-1

-3	-2	2	2
-4	-2	1	3
-4	-2	1	2
-3	-5	0	2

Fig. 1. Three residual examples predicted by the vertical mode.

B. *Proposed P-SVD Transform*

It is known, for a set of images, KLT trained from those images is the optimal transform to compact blocks in these images if only one fixed transform can be used. However, for each block in the image, the best transform is SVD transform which is degenerated from the separable KLT.

In intra coding, these SVD transforms can be calculated from the residual block as

$$[U_R, S_R, V_R] = \text{SVD}(R) \tag{2}$$

where S_R is the transform coefficients matrix which is a diagonal matrix and has singular values s_i along the main diagonal; R is current residual block; SVD ( ) represents the process to perform singular value decomposition on residual R in order to get the transform kernels. U_R and V_R are the eigenvector matrices for the residual block R. The transform process based on the SVD basis can be represented by

$$S_R = U_R^T \times R \times V_R \tag{3}$$

Since the SVD transform can diagonalize the residual block, the redundancy of the residual block is fully removed. In [6], the transform matrices U_R and V_R need to encode. It is obvious the bits cost on the transform matrices is significant. Reducing the coding bits on transform matrices U_R and V_R is crucial to SVD based image/video coding. However, previous work [9] [10] showed that it is difficult to find an efficient bit representation for matrices U_R and V_R.

Another way is to compromise the decorrelation capability of the SVD transform in order to decrease the bits cost on coding transform matrices. In the previous analysis, we can see that some residual blocks keep the structural information of the prediction blocks when the prediction is accurate. So, we could try using the prediction block to calculate the SVD transforms and apply these transforms to the residual block.

Assume the current coding block is O, the residual block R is calculated by:

$$R = O - P \tag{4}$$

where P is the prediction block. The SVD transforms for prediction block can be derived by:

$$[U_P, S_P, V_P] = \text{SVD}(P) \tag{5}$$

where U_P and V_P are the eigenvector matrices for P, and S_P is a diagonal matrix with singular values along the main diagonal. Similarly, the corresponding SVD transforms for the original block O can be represented by U_O and V_O which are derived by:

$$[U_O, S_O, V_O] = \text{SVD}(O) \tag{6}$$

When we apply transform matrices U_P and V_P to residual R, we can obtain the transform coefficients C for residual R by:

$$\begin{aligned} C &= U_P^T\, R\, V_P = U_P^T (O - P) V_P \\ &= U_P^T\, O\, V_P - U_P^T\, P\, V_P \\ &= S_O' - S_P \end{aligned} \tag{7}$$

If the prediction is accurate enough, we can assume that S_O' is close to S_O which only has singular values along the main diagonal. As a result, large coefficients in matrix C may only concentrate in or near the diagonal. Fig.2 shows an example for a 4×4 block predicted by Horizontal mode (mode 10) from sequence *BasketballPass*. It can be observed that S_O' is close to S_O especially for values in or near the diagonal.

Besides, this P-SVD transforms can capture structural information from prediction blocks which is similar to the structure of some residual blocks as illustrated in Fig. 1. Considering this structure similarity and the derivation of (7), we propose to use P-SVD as an additional transform for residual transform in intra coding.

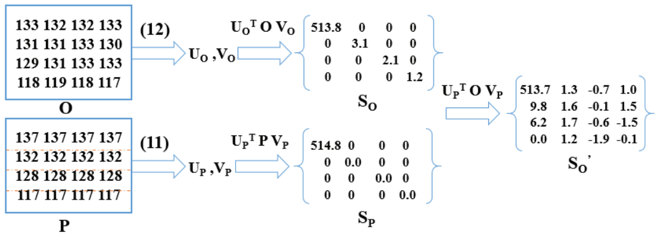


Fig. 2. An example to illustrate SVD transforms for a 4×4 block predicted by Horizontal mode (mode 10) from sequence *BasketballPass*. O is the original block and P is the prediction.

Since these SVD transforms are not derived by the residual block, we do not need to encode these transform matrices and these transforms can be derived in the decoder in the same way in the encoder after getting the intra prediction block. This P-SVD is a good balance between overhead cost on coding the transform matrices and the decorrelation capability compared to the method in [6] where these transform matrices need to be encoded and transmitted. The efficiency of the P-SVD transform will be verified in the experimental section.

III. INTRA CODING BASED ON THE PROPOSED TRANSFORM

In this section, we first introduce the framework of the proposed intra coding method based on the HEVC intra coding structure, then we explain the details of this method. Finally, we perform analyses for the proposed method.

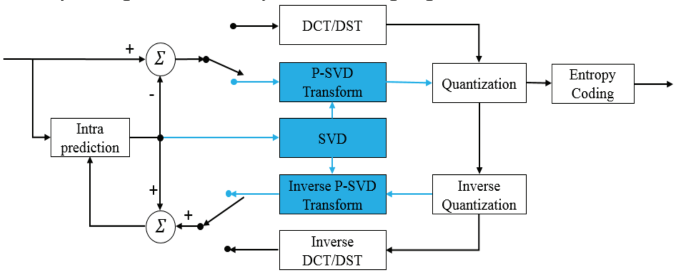


Fig. 3. The framework of the proposed coding method based P-SVD. The blue boxes and arrows are new added parts for our proposed method.

A. *Proposed Coding Method with P-SVD*

Fig. 3 shows the framework of the proposed intra coding method based on P-SVD. The blue boxes are the changes compared to the original HEVC framework. On top of the HEVC intra coding, P-SVD transform is added for the transform of the residual blocks. Here, the SVD matrices U_P and V_P are derived by the prediction block (P) of the current coding block (O), and applied to the residual block (R). The transform coefficients matrix C for the residual block can be derived by:

$$C = U_P^T \times R \times V_P \tag{8}$$

The corresponding inverse transform is represented by:

$$R = U_P \times C \times V_P^T \tag{9}$$

From the analysis in the previous section, we know that for most of the residual blocks which fit the Gauss-Markov model well can be transformed by DCT/DST efficiently, while for other residual blocks, the proposed P-SVD transform is more efficient.

Based on this, we propose to use the RDO process to decide which transform is the best transform for the current residual block. To reduce the overhead for signaling which transform is chosen, we implement our proposed P-SVD scheme on the CU level. It means all the TUs in the CU will choose one transform which is the transform that minimizes the RD cost of that CU.

In current reference software of HEVC intra coding, several selected intra prediction modes from rough mode decision (RMD) will go through the RDO process, and the best intra prediction mode and TU partition can be derived by minimizing the RD cost. In the proposed coding method, for each of the selected intra modes, both the original DST/DCT and the proposed P-SVD are tested in the RDO process, and the best transform is determined for each intra mode. Finally, the best combination of intra mode, transform type and TU partition is derived according to the RD cost.

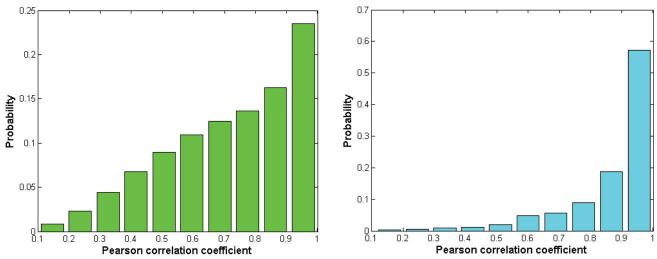
In the decoder side, the best transform flag for a CU is parsed first, then the residuals in each TUs will be derived by the inverse transform process with the corresponding transform type. If the selected transform is P-SVD, the SVD process will be conducted to derive the adaptive transform matrices based on the prediction block for TUs in the current coding block.

B. *Analyses for the Proposed P-SVD on AIP*

We select three of the test sequences to perform analyses for the proposed method. The correlation coefficients between the original blocks and the prediction blocks are calculated by the Pearson correlation coefficient (PCC):

$$PCC = \frac{\sum_i^N (O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_i^N (O_i - \overline{O})^2}\sqrt{\sum_i^N (P_i - \overline{P})^2}} \tag{10}$$

where O_i is the original block sample and P_i is the prediction block sample; N is the number of samples in the blocks. $\overline{O}$ and $\overline{P}$ are the mean values of the original block and prediction block, respectively.



a) For all blocks b) For blocks coded by P-SVD

Fig. 4. Distributions for Pearson correlation coefficients between original blocks and the corresponding angular intra prediction blocks for: a) all the 8×8 blocks, and b) blocks coded by P-SVD. The QP is 22 for test sequences.

We calculate the Pearson correlation coefficients between each 8×8 block and the corresponding best prediction block.

As shown in Fig.4 (a), for all the coding blocks, the PCCs vary from 0.1 to 1.0 but most of the blocks have the PCCs larger than 0.6. We also calculate the PCCs for the 8×8 blocks which use the P-SVD as the best transform, and the distribution is shown in Fig. 4 (b). It is obvious in this figure more than 90% of the P-SVD encoded blocks have PCCs larger than 0.7 which verifies P-SVD transform is efficient when the prediction is accurate and the residuals contain some structural information as the prediction blocks.

IV. EXPERIMENTAL RESULTS

A. *Experimental Settings*

To evaluate the performance of the proposed method, we implemented our method on the recent HEVC reference software (HM 14.0). The results are provided using test sequences from JCT-VC encoding with the All-Intra main configuration. 10 frames are used in our test. QP values of 22, 27, 32, and 37 are used for the evaluation. We measure the coding performance by the Bjontegaard-Delta Bit Rate (BD-BR). The performance of the proposed method is compared to the anchor which is the angular intra coding in HM 14.0 with default settings.

TABLE I
CODING PERFORFANCE OF THE DIFFERENT PARTS OF THE PROPOSED METHOD

	Test Sequences	Sequence Resolution	Proposed method
LR Sequences	BasketballDrill	832×480	-0.5%
	BasketballPass	416×240	-1.9%
	BlowingBubbles	832×480	-0.8%
	BQMall	832×480	-1.6%
	BQSquare	416×240	-1.2%
	PartyScene	832×480	-1.4%
	RaceHorses	416×240	-0.4%
	RaceHorses	832×480	-0.3%
HR Sequences	BasketballDrive	1920×1080	-1.7%
	BQTerrace	1920×1080	-0.7%
	Cactus	1920×1080	-0.7%
	Kimono1	1920×1080	0.0%
	ParkScene	1920×1080	-0.1%
	Traffic	2560×1600	-0.2%
Average (LR Sequences)			-1.0%
Average(HR Sequences)			-0.7%
Total Average			-0.9%

B. *Performance for the Proposed Method*

The results for the proposed methods are shown in Table I. From this table, we can see that the average bitrate reduction for the proposed method is about 0.9%. The average gain is 1.0% for the low resolution (LR) sequences and 0.7% for the high resolution (HR) sequences. The gain can be up to 1.9% for sequence *BasektballPass*. For sequence *Kimono1*, there has no improvement. This is because most blocks in this sequence are predicted by DC or planar modes so that the prediction block is smooth which cannot be used to get an efficient P-SVD transform. It is noted that the performance of P-SVD transform is better for LR sequences compared to the HR sequences. The reason is that the proposed P-SVD transform is more efficient to be applied on small TUs and a larger proportion of small TUs are used in LR sequences than those in HR sequences. The residual block is more likely to contain structures which are similar to the prediction block and the prediction is more accurate when the TU is small.

Because of the additional RD checks and singular value decomposition, the encoding times of the proposed method increase up to twice higher than the original HEVC. On the other hand, the decoding times increase 1%~30% which are dependent on the number of CUs coded by SVD. The complexity of the proposed method can be reduced by considering the characteristics of the prediction blocks which will be our future work.

V. CONCLUSIONS

This paper investigated the residual characteristics of intra prediction and the deficiency of DCT/DST for the residual blocks. Based on the residual analyses, an adaptive transform P-SVD which is derived from the prediction block, is proposed to compact the residual from the angular intra prediction. The experimental results show the proposed transform is efficient for intra coding. Experimental results suggest that the proposed method outperforms HEVC reference software with a bit-rate reduction of 0.9% on average, and the bit rate reduction can be up to 1.9%.

REFERENCES

[1] G. J. Sullivan, J. 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.

[2] K. R. Rao and P. Yip, Discrete Cosine Transform-Algorithms, Advantages and Applications. New York: Academic, 1990.

[3] 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 Processing, (ICIP), 2008, pp. 2116–2119.

[4] A. Saxena and F. C. Fernandes, “DCT/DST-based transform coding for intra prediction in image/video coding,” IEEE Trans. Image Process., vol. 22, no. 10, pp. 3974–3981, Oct. 2013.

[5] X. Zhao, L. Zhang, S. W. 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.

[6] A. Dapena and S. Ahalt, “A hybrid DCT-SVD image-coding algorithm,” IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 2, pp.114–121, 2002.

[7] M. Wang, K. N. Ngan, and L. Xu, “Efficient H.264/AVC Video Coding with Adaptive Transforms,” IEEE Trans. Multimedia., vol. 16, no. 4, pp. 933–946, Jun. 2014.

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

[9] H. C. Andrews and C. L. Patterson, “Singular value decomposition (svd) image coding,” IEEE Trans. Comniun., vol. 24, pp. 425-432, Apr. 1976.

[10] P. Waldemar and T. A. Ramstad, “Image compression using singular value decomposition with bit allocation and scalar quantization,” in Proc. NORSlG Conf., pp. 83-86, 1996.