WARPED DISCRETE COSINE TRANSFORM EXTENSION TO THE H.264/AVC

Sang Heon Lee and Nam Ik Cho

School of Electrical Engineering and Computer Sciences,
Seoul National Univ., S. Korea
email: *leesh007@ispl.snu.ac.kr, nicho@snu.ac.kr*

ABSTRACT

This paper proposes a new video compression algorithm using an adaptive transform that is adjusted depending on the frequency contents of the input signals. The adaptive transform is based on the warped discrete cosine transform (WDCT) which is shown to provide better performance than the DCT at high bit rates, when applied to JPEG compression scheme [1, 2, 3]. The WDCT is applied to the video compression in this paper, as a new feature in the H.264/AVC. The proposed method shows the coding gain over the H.264/AVC at high bit rates. The coding gain is shown over the 35dB PSNR quality, and the gain increases as the bit rate increases. (about 1.0dB at 45dB PSNR quality at maximum)

***Index Terms*—** Video codecs, Video coding, Discrete Cosine Transforms

1. INTRODUCTION

Transform coding has been used in most of compression algorithms for decades. The transform coding compacts most energies into the low-frequencies and thus the transformed result is appropriate for a run-level entropy coding. Theoritically, Karhunen Loeve Transform (KLT) is known to be the optimum energy concentration transform. But it is unrealistic because of its variation according to the input statics and the difficulties in hardware implementation. Thus many alternative fixed transforms have been proposed. Among them, the Discrete Cosine Transform (DCT) is known to best approximate the KLT when the property of the signal matches with the first order Gauss-Markov model.

In the case of the H.264/AVC, an integer transform is used instead of the DCT, which is almost identical to the DCT except that the kernels are slightly modified for the integer computation. By using the DCT-like transform, it is expected that the energy of signal will be compacted to the lower frequencies for many parts of images that can be modeled by the first order Gauss-Markov signal. But, for the parts where this model is not well matched, the efficiency of this transform is decreased.

In order to alleviate the problems of using a fixed transform for every block of an image, new transform methods have also been introduced. For example, in the KTA software project (Advanced version of the H.264/AVC), adaptive prediction error coding tool[5, 6] and mode dependent directional transform tool[7] have been introduced. Adaptive prediction error coding tool selects a prediction error domain (spatial domain or frequency domain) adaptively. In the spatial domain, some signals can be quantized and entropy coded without any transform. Mode dependent directional transform employs a trained KLT according to the intra prediction direction for the intra coding. Above tools show that different transforms can yield better coding gain. Based on these promising results of alternative transforms, we develop another adaptive transform method for the H.264/AVC. We use the WDCT which is shown to be efficient when used in the JPEG image coding. The compression method employs several pre-defined transforms which are generated by warping the frequency responses of each row of the DCT. In this paper, we extend the WDCT method to the inter frames of the H.264/AVC as well as intra frames. Applying the WDCT to the small size blocks in H.264/AVC (size of 4×4) can cause the problem that the side information can override the coding gain. Hence, we adaptively choose to use the WDCT under the RD-optimized selection of H.264/AVC. In summary, we extend the warped transforms for the 4×4 blocks, and also develop an optimal WDCT parameter selection and WDCT index transmission algorithm considering the features of the H.264/AVC.

This paper is organized as follows. Section 2 describe how to establish the 4×4 WDCT transform. The details of the extension of WDCT for the H.264/AVC are presented in Section 3. Simulation results are shown for the 4 test sequences in Section 4, and we conclude the paper in Section 5.

2. INTRODUCTION OF WARPED DISCRETE COSINE TRANSFORM

The WDCT is derived from the cascade of an IIR all pass filter and the DCT. It is approximated to be a single matrix, and we call it as a WDCT matrix [1, 2, 3]. The 4×4 DCT is

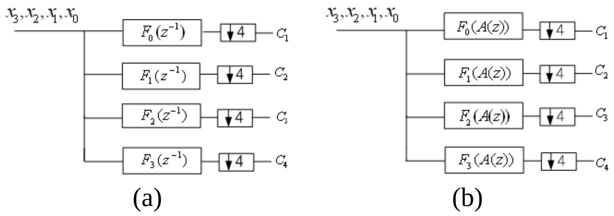


Fig. 1. Representation of DCT and WDCT in the form of filter bank. (a). DCT filter bank, (b). WDCT filter bank.

represented as

$$C_k = U(k) \sum_{n=0}^{3} x_n cos \frac{(2n+1)k}{8} \pi \quad (1)$$

where

$$U(k) = \begin{cases} 1, & k = 0 \\ \sqrt{\frac{1}{2}}, & \text{otherwise.} \end{cases} \quad (2)$$

The computation of DCT can also be represented by a filter bank as shown in Fig. 1(a), where each filter $F_k(z^{-1})$ is represented by

$$F_k(z^{-1}) = U(k)\{cos\frac{k\pi}{8} + cos\frac{3k\pi}{8}z^{-1} + cos\frac{5k\pi}{8}z^{-2} + cos\frac{7k\pi}{8}z^{-3}\}. \quad (3)$$

For warping the frequency response of this "DCT filter bank", a first order IIR all pass filter is used, denoted as

$$A(z) = \frac{-\alpha + z^{-1}}{1 - \alpha z^{-1}}. \quad (4)$$

By changing the "warping" parameter α, the frequency distribution of the input signal can be warped to be appropriate for the DCT. By replacing z^{-1} into $A(z)$, we obtain the IIR WDCT filter bank as shown in Fig. 1(b), where each filter $F_k(A(z))$ is represented as

$$F_k(A(z)) = U(k)\{cos\frac{k\pi}{8} + cos\frac{3k\pi}{8}A(z) + cos\frac{5k\pi}{8}A(z)^2 + cos\frac{7k\pi}{8}A(z)^3\}. \quad (5)$$

The magnitude response of the DCT filter bank and the WDCT filter bank for α=0.5 are shown in Fig. 2. The figure shows that WDCT can warp the frequency response of the DCT by changing the value of α. Thus if an appropriate α is chosen, it can increase the coding efficiency for the cases when the DCT cannot compress input signal effectively. For the practical implementation of the WDCT, the above IIR filter representation

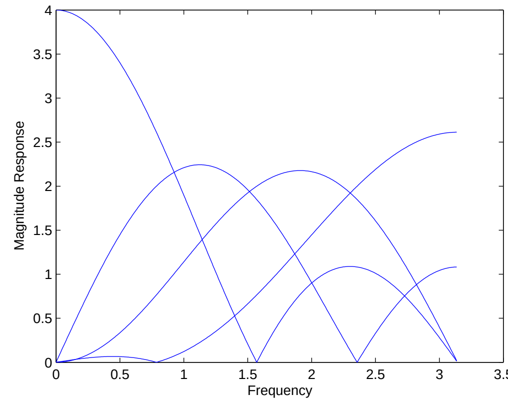


Fig. 2. Frequency response of the DCT.

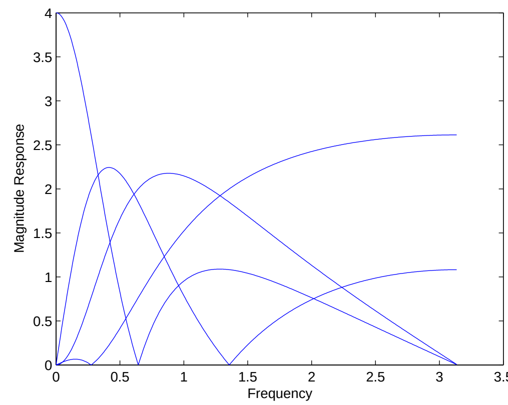


Fig. 3. Frequency response of the WDCT when α=0.5.

should be approximated to be an FIR filter. By using the Parseval's theorem, the IIR filter is approximated to be an FIR form. To be specific, the frequency components are sampled at the frequencies $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ of $F_k(A(z))$, and then the IDFT is performed for it. Below is an example of approximating $F_0(A(z))$ when α is 0.05. By sampling the frequency response of $F_0(z)$ at w=0, $\frac{\pi}{2}$,π, $\frac{3\pi}{2}$, we have

$$[2.000 \;\; 0.1077 + 0.0795i \;\; 0 \;\; 0.1077 - 0.0795i]^T, \quad (6)$$

and the IDFT of this vector is

$$[0.5539 \;\; 0.4653 \;\; 0.4461 \;\; 0.5397]^T, \quad (7)$$

which are the elements of the first row of WDCT matrix for $\alpha = 0.05$. We can generate other rows of the WDCT matrix in the same manner.

3. EXTENSIONS TO THE H.264/AVC

In this section, we explain how the WDCT can be applied to the H.264/AVC. In this paper, we change α between [-0.1, 0.1] with the interval of $\frac{1}{40}$, and prepare nine WDCT matrices. When the α is 0, the WDCT matrix is identical to the DCT matrix, and we use the original integer transform in this case. Thus for every 4×4 block, 9 transforms are tested and the optimal WDCT matrix for this block is founded. Because several candidates exist for the transform of the input signal, RD-optimized WDCT matrix selection is an important problem. It can be solved by using an already existing RD-optimization technique in the H.264/AVC. Optimal WDCT matrix can be determined according to the following existing RD-equation,

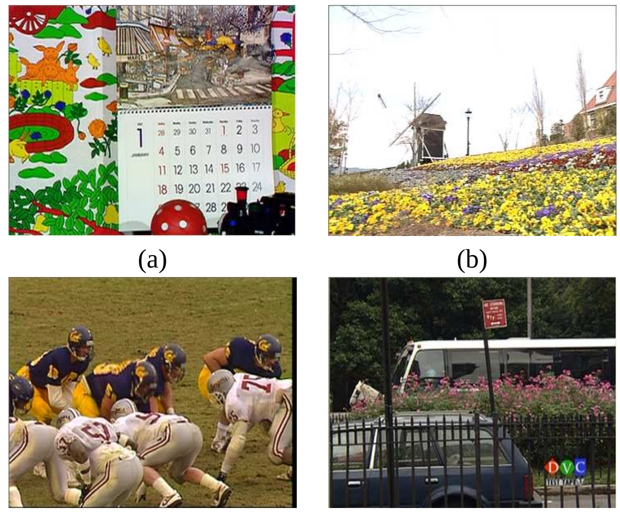
$$\begin{aligned} &J(s, c, index|QP, \lambda_{mode}) \\ &= SSD(s, c, index|QP) + \lambda_{mode} R(s, c, index|QP), \end{aligned} \quad (8)$$

where *s* means a value of the original signal, *c* is the reconstructed signal, *index* represents each WDCT matrix, SSD is the square sum of residuals between *s* and *c*, and R means the sum of required bits to transmit the transformed and quantized residual coefficients and a WDCT index.

Transmitting the selected WDCT index by the above RD-optimization technique is also an important problem. WDCT index is a big burden for using the WDCT with the H.264/AVC, when the small size blocks (4×4) are frequently used. Because there are many subblocks that do not transmit any residual signals (i.e., inverse transform is not required at the decoder side), WDCT index should be transmitted considering the existence of the residual signals. For that, we exploit the existing syntax CBP. In the MPEG video coders, there is a concept so called as CBP (coded bit pattern), which can be used to detect the existence of the non-zero coefficients for each subblock (normally for the 8×8). Thus, WDCT index syntax should be placed after the CBP syntax. If CBP indicates that there is no non-zero coefficients, WDCT index can be skipped. Other unskipped WDCT index is binarized as Table. 1. Because the conventional integer transform is a dominant transform in the low bit-rates, we allocate the shortest codeword for the warping parameter 0. Also, we use a CABAC to transmit the WDCT index. Because CABAC can reduce the occurred bits when its probability is biased to the specific mode, WDCT can be used without significant PSNR drops in the low bit-rates.

Table 1. Binarization of the WDCT Index

WDCT index	Warping Parameter (α)	Binarization
0	0	1
1	-0.1	000
2	-0.075	001
3	-0.05	010
4	-0.025	011
5	0.025	100
6	0.05	101
7	0.075	110
8	0.1	111



(a) (b) (c) (d)

Fig. 4. Test sequences for simulation. (a) Mobile, (b) Flower, (c) Football, (d) Bus

4. SIMULATION RESULTS

The proposed method is implemented on the H.264/AVC JM [9] version 12.0. We use mobile, flower, football and bus CIF sequences for the simulation and use the high profile. Test sequences have 4:2:0 color formats and the proposed method is simulated for 100 frames with a IPPP GOP structure and GOP size 15. Then we test the proposed algorithm for different 4 QP values from QP17, QP20, QP23, and QP26. Fig. 5 shows the RD-results of the each sequence. It can be observed that the proposed method gives higher coding performance over the 35dB PSNR quality. Also, the PSNR gains increases as the bit-rates increase, and the maximum gains are more than 1.0dB at the 45dB PSNR quality. In summary, the WDCT yields coding gain over the integer transform at high bit rates..

5. CONCLUSIONS

In this paper, we have presented a method to employ the WDCT to the H.264/AVC video compression. By using several different transforms that can match the properties of input signal, we can have coding gain over the conventional fixed transforms. When applied to the H.264/AVC it shows some coding gains over the 35dB PSNR quality with the IPPP GOP structure. It shows that WDCT yields coding gain over the integer transform despite of excessive side-information required for the transmission of WDCT index.

6. REFERENCES

[1] N. I. Cho and S. K. Mitra, “Warped Discrete Cosine Transform and Its Application in Image Compression,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1364-1373, Dec, 2000.

[2] I. K. Kim, N. I. Cho and S. K. Mitra, “Rate-distortion optimization of the image compression algorithm based on the warped discrete cosine transform,” *Signal Processing*, vol. 83, issue. 9, pp. 1919-1928, Sep. 2003.

[3] Urhan. O and Erturk. S, “Parameter Embedding Mode and Optimal Post-Process Filtering for Improved WDCT Image Compression,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, issue. 4, pp. 528-532, Apr. 2008

[4] “Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC),” JVT of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, 2003.

[5] M. Narroschke, “Adaptive Prediction Error Coding in Spatial and Frequency Domain for H.264//AVC,” ITU-T SG16/Q6, VCEG-AB06, Jan, 2006.

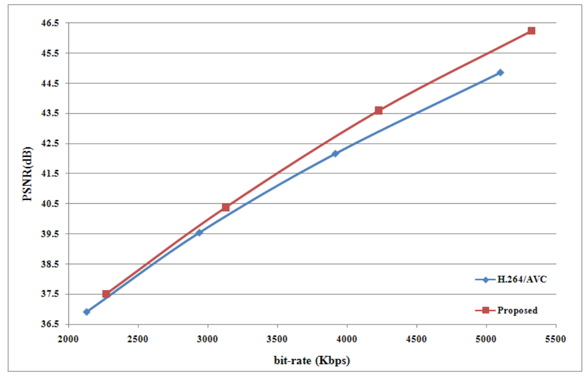
[6] M. Narroschke, “Adaptive Prediction Error Coding in Spatial and Frequency Domain with a Fixed Scan in the Spatial Domain,” ITU-T SG16/Q6, VCEG-AD07, Oct, 2006.

[7] Y. Ye and M. Karczewicz, “Improved Intra Coding,” ITU-T VCEG, VCEG-AG11, Shenzhen, China, Oct, 2007.

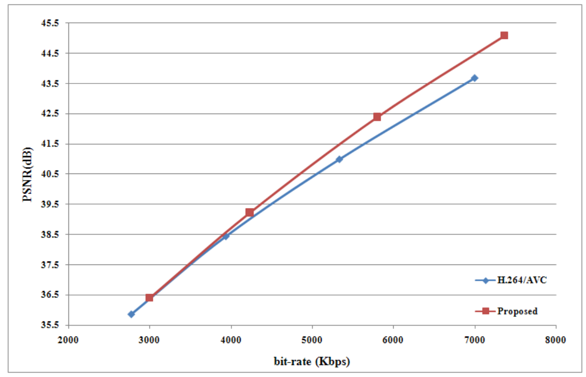
[8] D. Marpe, H. Schwarz and T. Wiegand, “Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard,” *IEEE CSVT*, vol. 13, pp. 620-636, 2003.

[9] http://iphome.hhi.de/suehring/tml/download/, H.264 JM reference model.

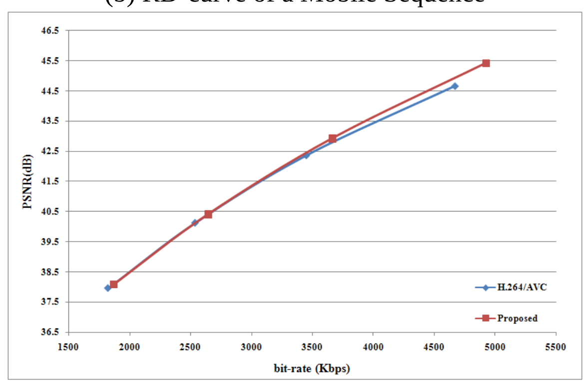
[10] G. Bjontegaard, “Calculation of average PSNR differences between RD curves,” ITU-T VCEG, VCEG-M33, Austin, USA, Apr, 2001.



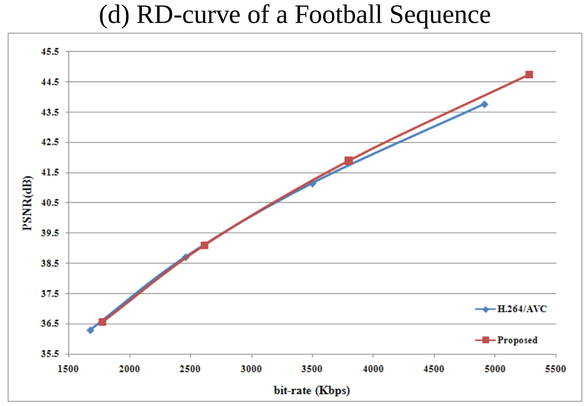
(a) RD-curve of a Flower Sequence



(b) RD-curve of a Mobile Sequence



(d) RD-curve of a Football Sequence



(d) RD-curve of a Bus Sequence

Fig. 5. RD-curves of the proposed method.