VARIABLE TEMPORAL LENGTH 3D DCT-DWT BASED VIDEO CODING

Jin Li[1], Jarmo Takala[1], Moncef Gabbouj[1], Hexin Chen[2]

Department of Information Technology
Tampere University of Technology, Tampere, Finland[1]
School of Communication Engineering
Jilin University, Changchun, China[2]

ABSTRACT

Almost all variable length 3D-DCT algorithms employ thresholds. Although the selection of threshold considerably affects the algorithm performance, there is no framework to choose them effectively. This paper proposes a hybrid transform for 3D-DCT based video coding. In the proposed model, 3D-DCT and discrete Haar transform are iteratively used to remove the redundancy for each cube and thus an adaptive scheme is derived to realize variable temporal length of DCT implementations. Compared to other methods in the literature, the proposed model can mathematically select the optimal DCT mode and remove the temporal correlations more effectively. Experimental results show that the proposed approach has substantial improvement over the conventional fixed-length 3D-DCT coding and other variable length 3D-DCT coding.

Index Terms — 3D-DCT, Discrete Haar transform, video coding, computational complexity

1. INTRODUCTION

In recent years, there have been great advancements in video coding techniques. Generally, compression efficiency is improved along with an increase of computational cost. As the newest video coding standards, H.264 significantly outperforms other standards in terms of coding efficiency. However, the complexity is greatly increased. This is not appropriate especially for portable digital applications like mobile phone and digital video cameras. Since for these types of applications, low complexity implementation and low power consumption are still the most critical issues.

An alternative approach is to use the three-dimensional discrete cosine transform (3D-DCT). A 3D-DCT based video codec extends 2D-DCT to the temporal dimension and can effectively remove the redundancy among a number of consecutive frames. Therefore, it is able to reach adequate compression efficiency with less computational cost.

One of the major problems in a 3D-DCT based codec is that if only utilizing fixed-length transform regardless the level of motion activity, the coding efficiency is usually inferior to most today's video codec. To solve this problem, some variable length 3D-DCT schemes with multiple thresholds are proposed in [1]-[5]. These techniques utilize variable transform based on certain empirical thresholds. They show significant improvement for sequences with low motion activity. However, since the selection of thresholds is empirically determined, the algorithm performance is not optimized.

In what follows we describe an adaptive hybrid algorithm for 3D-DCT based video coding. This proposed model iteratively utilizes 3D-DCT and discrete Haar transform (DHT) to remove the temporal redundancy and is able to find out the optimal DCT mode for each $8 \times 8 \times 8$ cube. Compared to other schemes, this technique can remove the temporal redundancy more sufficiently and thus achieve better compression performance.

The rest of this paper is organized as follows. In section 2 we briefly review the basics of 3D-DCT and DHT. We describe the hybrid algorithm in section 3. The experimental results are presented in section 4. Finally, we conclude the paper in section 5.

2. 3D-DCT AND DISCRETE HAAR TRANSFORM

2.1 Three Dimensional Discrete Cosine Transform

In a 3D-DCT based codec, a video sequence is divided into a number of M×N×L cubes, where M×N is an image block of pixels, and L is the number of successive frames. The forward 3D-DCT is then defined as

$$F(u,v,w) = C(u,L)C(v,N)C(w,M)\sum_{x=0}^{L-1}\sum_{y=0}^{N-1}\sum_{z=0}^{M-1} f(x,y,t)\times \frac{\cos(2x+1)u\pi}{2L}\times\frac{\cos(2y+1)v\pi}{2N}\times\frac{\cos(2t+1)w\pi}{2M} \quad (1)$$

where

$$C(k,P) = \begin{cases} \sqrt{1/P} & k = 0 \\ \sqrt{2/P} & otherwise \end{cases}$$

The 3D-DCT can be performed by taking one-dimensional transform separately in each of the three dimensions. Although a number of transforms are required, the computational complexity – even without taking the motion estimation into account – is superior to 2D-DCT based video encoder. A typical architecture of 3D-DCT based video codec is described in Fig. 1.

2.2 Discrete Haar Transform

Discrete Haar transform is the most basic kernel of discrete wavelet transform (DWT). The notation of [6] is used to illustrate the fast DHT algorithm for a data vector of $N = 2^L$ elements, $a_i^0 = x(i)$ indexed by $0 \le i \le N$. From this input vector we form two new vectors, each of half its length

$$d_i^1 = \frac{1}{2}[a_{2i}^0 - a_{2i+1}^0] \quad (2)$$

$$a_i^1 = \frac{1}{2}[a_{2i}^0 + a_{2i+1}^0] \quad (3)$$

with $0 \le i \le N$, which contain the high frequency and low frequency coefficients, respectively.

Then, the algorithm just applies the same process to a_i^1

$$d_i^l = \frac{1}{2}\left[a_{2i}^{l-1} - a_{2i+1}^{l-1}\right] \quad (4)$$

$$a_i^l = \frac{1}{2}\left[a_{2i}^{l-1} + a_{2i+1}^{l-1}\right] \quad (5)$$

with $0 \le i \le 2^{L-1}$ for $1 \le l \le L$.

So, the DHT is given by the sequence

$$H = [a^L, d^L, d^{L-1}, d^{L-2}, \ldots, d^1] \quad (6)$$

where d^l is the sequence d_i^l for $0 \le i \le 2^{L-1}$.

Similarly to DCT, DHT also has a favorable characteristic of energy packing. After transform most of the energy is deposited to the low frequency coefficients in a_i^l and a large amount of high frequency information can be neglected. Moreover, DHT can decompose a signal into a set of basic functions both in time and spatial frequency domain.

The DHT can be applied to 3D-DCT based video coding to remove the temporal correlations more effectively. If the cube to be transformed contains only low frequency contents, the using of DHT can produce tighter information representation along with a smaller cube and thus, obtain better compression performance.

3. PROPOSED HYBRID 3D DCT-DWT MODEL

We propose a hybrid 3D DCT-DWT based video coding with variable length of cube. In the proposed model, DHT and 3D-DCT are iteratively used to remove the redundancy for each $8 \times 8 \times 8$ cube. Moreover, it can adaptively find out the optimal DCT mode based on the local DCT coefficient contents. Totally, four modes are utilized to perform the transform along temporal dimension.

Since one of the most desirable properties of DHT is that the decomposition can be performed both in time and spatial domain, the proposed model is started with DHT for each cube. Given an $8 \times 8 \times 8$ cube $f(x, y, t)$ with $0 \le x, y, t \le 7$, where x, y, t represent the horizontal, vertical and temporal dimension, respectively. The DHT is first applied to the cube along temporal dimension

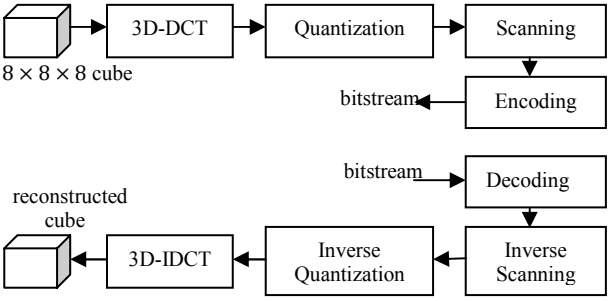


Fig. 1 Block diagram of typical 3D-DCT based video codec

$$f_L^1(x, y, t) = \frac{1}{2}[f(x, y, t) + f(x, y, t+1)] \quad (7)$$

$$f_H^1(x, y, t) = \frac{1}{2}[f(x, y, t) - f(x, y, t+1)] \quad (8)$$

$$0 \le x, y \le 7 \text{ and } t = 0,2,4,6.$$

where $f_L^1(x, y, t)$ is the $8 \times 8 \times 4$ low frequency cube and $f_H^1(x, y, t)$ is the corresponding residual cube containing only high frequency information.

Subsequently, 3D-DCT is performed on the residual cube $f_H^1(x, y, t)$ as (1)

$$F_H^1(u,v,w) = C(u,8)C(v,8)C(w,4)\sum_{x=0}^{7}\sum_{y=0}^{7}\sum_{t=0}^{3} f_H^1(x,y,t)\times \frac{\cos(2x+1)u\pi}{16}\times\frac{\cos(2y+1)v\pi}{16}\times\frac{\cos(2t+1)w\pi}{8} \quad (9)$$

where $F_H^1(u, v, w)$ denotes the transformed residual cube in DCT domain.

Thirdly, the transformed cube $F_H^1(u, v, w)$ is truncated by a uniform quantization parameter QP_1

$$F_{HQ}^1(u, v, w) = F_H^1(u, v, w)/QP_1 \quad (10)$$

$$0 \le u, v \le 7 \text{ and } 0 \le w \le 3.$$

where $F_{HQ}^1(u, v, w)$ is the quantized $F_H^1(u, v, w)$.

If not all coefficients in $F_{HQ}^1(u, v, w)$ are truncated to zeros, we repeat (9) on the low frequency cube and encode the two $8 \times 8 \times 4$ cubes separately, which is regarded as mode 1.

Otherwise, if all the coefficients are quantized to zeros, which means that the correlations in the low frequency cube are still high, we again apply DHT on the $8 \times 8 \times 4$ low frequency cube to further remove the redundancy

$$f_L^2(x, y, t) = \frac{1}{2}[f_L^1(x, y, t) + f_L^1(x, y, t+1)] \quad (11)$$

$$f_H^2(x, y, t) = \frac{1}{2}[f_L^1(x, y, t) - f_L^1(x, y, t+1)] \quad (12)$$

$$0 \le x, y \le 7 \text{ and } t = 0,2.$$

We repeat 3D-DCT on $f_H^2(x, y, t)$ with the transform size of $8 \times 8 \times 2$ and finally obtain the quantized cube as

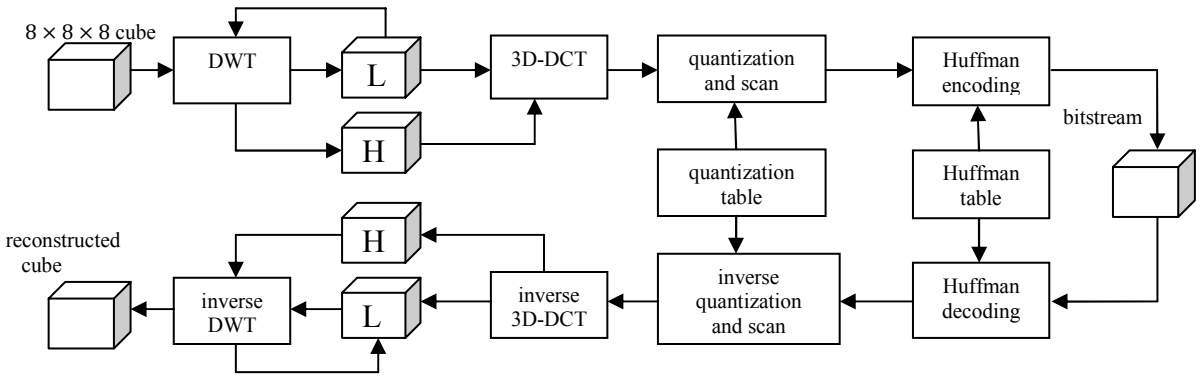


Fig. 2 Architecture of proposed hybrid 3D DCT-DWT based video codec, L: low frequency cube, H: high frequency cube

$F_{HQ}^2(u,v,w)$.Similarly, if there are still nonzero coefficients in the quantized cube, we take 3D-DCT on the corresponding low frequency cube $f_L^2(x,y,t)$ and encode the two $8 \times 8 \times 2$ cubes separately. This can be regarded as mode 2.

If all coefficients in $F_{HQ}^2(u,v,w)$ are truncated to zeros, we sequentially repeat DHT and 3D-DCT to the $8 \times 8 \times 2$ low frequency cube and thus obtain two 8×8 blocks: $f_L^3(x,y)$ with low frequency only and $f_H^3(x,y)$ with high frequency.

We take DCT and quantization on these two blocks separately. If all coefficients in $F_{HQ}^3(u,v)$ are truncated to zeros, we only encode the first block as mode 3. Otherwise, we encode both the low frequency and high frequency blocks, which can be regarded as mode 4.

The proposed algorithm to encode an $8 \times 8 \times 8$ cube can be summarized as

1) Take DHT to separate the original cube into two cubes f_L^1 and f_H^1 containing only low frequency and high frequency respectively;
2) Take 3D-DCT and quantization on f_H^1 ;
3) If not all coefficients are truncated to zeros, take 3D-DCT on f_L^1 and encode the two cubes as mode 1;
4) Otherwise, sequentially repeat DHT on f_L^1 and decompose it into f_L^2 and f_H^2;
5) Take 3D-DCT and quantization for f_H^2 ;
6) If not all coefficients are truncated to zeros, take 3D-DCT on f_L^2 and encode the two cubes as mode 2;
7) Otherwise, continuously take DHT on f_L^2 and decompose it into two blocks f_L^3 and f_H^3;
8) Take 3D-DCT and quantization for both the blocks;
9) If not all coefficients are truncated to zeros, encode these two blocks as mode 3;
10) Otherwise, only encode F_{LQ}^3 as mode 4.

In the proposed hybrid model, the mode decision is totally based on the truncated high frequency information. Only if all high frequency contents in the residual cube are quantized to zeros, the hybrid transform is repeated to form a smaller cube and thus produces a tighter information representation. Compared to other variable length 3D-DCT schemes, the proposed model adaptively find out the optimal mode decision without notable loss of energy. Therefore, it can achieve higher coding efficiency.

Fig. 2 gives the architecture of the proposed hybrid codec. In addition, two extra bits are encoded for each cube to indicate the mode decision in the encoder. In the decoder, all steps from the encoding process, except the mode analysis, are implemented in the reverse order.

4. EXPERIMENTAL RESULTS

The proposed hybrid algorithm was tested against the baseline 3D-DCT codec and two reference codec in [1, 2]. Video sequences with various motion activities were encoded and decoded. The Peak Signal to Noise Ratio (PSNR) versus compression ratio (CR) curve is plotted based on the obtained results.

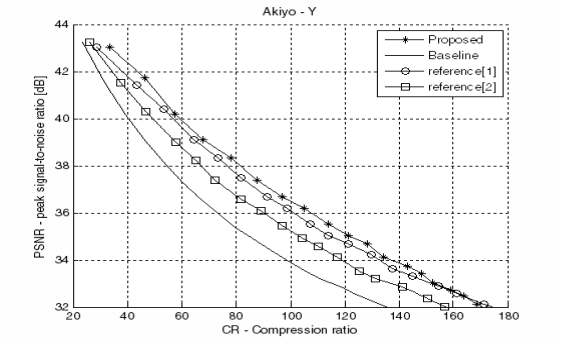
The quantization parameter (QP) for DC coefficients in the low frequency cubes is fixed to 10 for all the modes. The QP for the coefficients in high frequency cubes is uniform, and they satisfy the following relationship for different modes

$$QP_1 = \frac{3}{2}QP_2 = 2QP_3 = 2QP_4$$

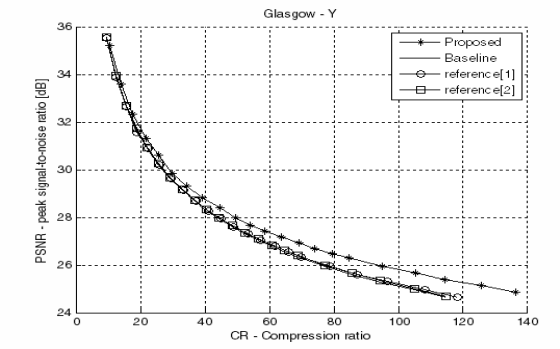
where QP_i denotes the QP of mode i.

Experimental results show that the proposed algorithm can give better compression performance for different types of sequences. Best improvements can be expected for those with low motion activity. Fig. 3 shows the luminance PSNR versus CR curves of Akiyo and Glasgow. According to the results, the proposed scheme can give about 0.5-2.5dB improvement over the baseline codec and the reference codec. In addition, the similar improvements can be also obtained for chrominance components.

Table I shows the utilization ratio of the four modes in the proposed hybrid video codec. Since the mode decision is



(a)



(b)

Fig. 3 Luminance PSNR vs. CR for (a) Akiyo and (b) Glasgow based on four different codec: the baseline 3D-DCT codec, the proposed codec and the reference codec in [1, 2].

related to QP, the utilization of different modes changes with QP. Fig. 4 shows the statistical results on frame 130 of Akiyo sequence at and the enhanced differential image between frame 130 and 131.

In the experiments, we also evaluated the subjective visual quality with the benchmarks [7]. The proposed model works well as is small (e.g. 12,30,52) and no temporal coding artifacts are observed. However, as is increasing, visible temporal artifacts appear due to the hard truncation on high frequency. Thus, future work includes further improvement of quantization strategy and utilization of more efficient DWT algorithm.

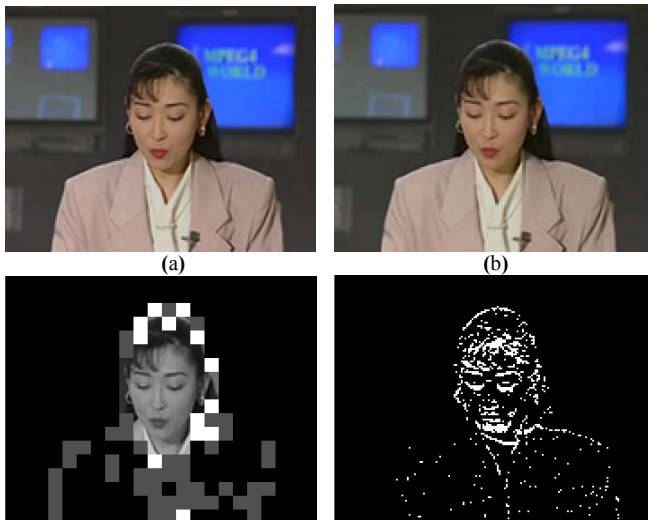
5. CONCLUSION

A hybrid transform algorithm is proposed for 3D-DCT video process. The proposed model iteratively utilizes DCT and DHT to exploit the redundancy and mathematically determine the optimal DCT mode. A series of experiments show that the proposed algorithm outperforms the baseline codec and the reference codec in terms of coding efficiency.

Although the compression efficiency is still lower than H.264, the encoding process of 3D-DCT is much faster. This makes 3D-DCT based video codec especially suitable for such devices with restrict computational power. Potential applications could be for portable digital devices such as mobile phone and digital video cameras. Moreover, since the proposed algorithm requires less computations, it is suited for applications with restrict real-time requirement.

Table I Utilization Ratio of Proposed Modes

QP	Glasgow				Akiyo			
	M1 (%)	M2 (%)	M3 (%)	M4 (%)	M1 (%)	M2 (%)	M3 (%)	M4 (%)
12	63.5	7.8	3.5	25.2	2.0	4.8	5.1	88.1
30	44.4	16.6	7.3	31.7	0.2	2.1	3.7	94.0
52	27.6	20.8	12.2	39.4	0.1	0.8	2.3	96.8
86	14.2	15.6	19.1	51.1	0.0	0.2	1.3	98.5
120	8.5	14.0	16.1	61.4	0.0	0.1	0.5	99.4
160	5.0	9.8	14.8	70.4	0.0	0.0	0.1	99.9



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

Fig. 4 (a) original frame, (b) reconstructed frame, (c) utilization of different mode on luminance components - Original: Mode 1, White: Mode 2, Grey: Mode 3 and Black: Mode 4, (d) enhanced differential luminance image between frame 130 and 131 of Akiyo.

6. ACKNOWLEDGEMENT

This work is supported partly by Chinese Science & Technology Ministry under Grant 2005DFA10300 and by Academy of Finland under Grant 117065.

7. REFERENCES

[1] N.P. Sgouros, S.S. Athineos, P.E. Mardaki and *etc.*, "Use of an Adaptive 3D-DCT Scheme for Coding Multiview Stereo Images," IEEE *Proceedings of ISSPIT*, pp.180-185, 2005

[2] B. Furht, K. Gustafson, H. Huang, and O. Marques, "An Adaptive Three-Dimensional DCT Compression Based on Motion Analysis," *Proceedings of ACM*, pp. 765-768, 2003.

[3] Y.L. Chan and W.C. Siu, "Variable Temporal-length 3-D Discrete Cosine Transform Coding," IEEE *Transactions on Image Processing*, VOL 6, No. 5, pp. 758-763, 1997

[4] S.C. Tai, Y.G. Wu and C.W. Lin, "An Adaptive 3-D Discrete Cosine Transform Coder for Medical Image Compression," IEEE Transactions *on Information Technology in Biomedicine*, Vol.4, pp. 259-264, 2000

[5] J.J. Koivusaari, and J.H. Takala, "Simplified Three-Dimensional Discrete Cosine Transform Based Video Codec," SPIE *Proceedings of Multimedia on Mobile Devices*, pp. 11-21, 2005.

[6] J.R. Macias and A.G. Exposito, "Efficient Computation of the Running Discrete Haar Transform," IEEE *Tran. on Power Delivery,* pp.1-2, 2005.

[7] H.R. Wu and K.R. Rao, Eds., Digital Video Image Quality and Perceptual Coding, CRC Press (ISBN: 0-8247-2777-0), Nov. 2005