ICSP2008 Proceedings

Simplified Video Coding for Digital Mobile Devices

Jin Li[1], Moncef Gabbouj[1], Jarmo Takala[1], and Hexin Chen[2]
Department of Information Technology, Tampere University of Technology, Tampere, Finland[1]
Department of Communication Engineering, Jilin University, Changchun, China[2]
FirstName.LastName@tut.fi[1], chx@email.jlu.edu.cn[2]

Abstract

Most mobile devices are currently suffering from the lack of computational power and energy-consumption constraints, there is significant interest and research in reducing the computational complexity for fast video coding. This paper proposes a statistical model to predict the zero-quantized DCT (ZQDCT) coefficients for 3-D DCT video coding and thus to speech up the encoding process. First, each $8 \times 8 \times 8$ *pixel cube at the input of DCT is decomposed into a mean value and a residual* $8 \times 8 \times 8$ *pixel cube. Subsequently, a statistical Laplacian model is mathematically developed to predict the ZQDCT coefficients. Finally, the redundant calculations for these ZQDCT coefficients are skipped. Compared to the baseline encoding method, the proposed model can significantly simplify the complexity and achieve better real-time performance at the expense of negligible visual degradation. Computational reduction also implies longer battery lifetime and energy economy for portable devices.*

1. Introduction

Previously, the efforts to advance 3-D DCT algorithms are mainly focused on the improvement of compression efficiency. These proposed approaches can reduce the correlations more effectively by using the scene change detector [1], threshold for different local activity [2]-[3] and multiple description method [4]. The latest results [5] indicate that 3-D DCT coding outperforms MPEG-2 while is slightly worse than MPEG-4 at high bit rate applications.

Although 3-D DCT coding is much superior to H.264 and MPEG-4 in terms of computational complexity, it is still desirable to further reduce the computations with minimal visual quality degradation. As the most complex part, 3-D DCT consumes more than half of the computations in the entire encoding process while a huge number of coefficients will be finally quantized to zeros, which can be regarded as redundant computations. Thus, if we can reduce these redundant computations, the encoding process will be much faster. This also implies longer battery lifetime for portable digital applications.

This paper proposes a statistical model to skip redundant DCT and quantization with minimum negligible visual quality degradation. Although the proposed method is implemented based on the baseline 3-D DCT video coding, it can be widely used on other 3-D DCT schemes in [1]-[4]. As a result, high prediction efficiency and significant computational savings are achieved by the proposed model.

The rest of this paper is organized as follows. The statistical model is mathematically analyzed and developed in Section 2. The experimental results are presented in Section 3. Finally, Section 4 concludes this paper.

2. Statistical model of 3-D DCT computation

2.1 Mathematical decomposition of DCT

In this paper, we mainly consider the $8 \times 8 \times 8$ 3-D DCT. If we define $\boldsymbol{f}$ as the pixel cube and $f(x,y,z)$ is the pixel value, $0 \le x,y,z \le 7$, the coefficient $F(u,v,w)$ is computed by

$$F(u,v,w) = \frac{c(u)c(v)c(w)}{8}\sum_{x=0}^{7}\sum_{y=0}^{7}\sum_{z=0}^{7} f(x,y,z)\cos\frac{(2x+1)u\pi}{16} \times \cos\frac{(2y+1)v\pi}{16}\cos\frac{(2z+1)w\pi}{16} \quad (1)$$

where $c(u), c(v), c(w) = 1/\sqrt{2}$, for $u,v,w = 0$, and $c(u)$, $c(v), c(w) = 1$, otherwise.

Alternatively, the 3-D DCT can be calculated in matrix form as [6]

$$\boldsymbol{F} = (\boldsymbol{B}^{T3}(\boldsymbol{B}^{T2}(\boldsymbol{B}^{T1}\boldsymbol{f}\boldsymbol{B})\boldsymbol{B})\boldsymbol{B})$$

where

$$\boldsymbol{B}_{(\boldsymbol{w})} = \begin{cases} \frac{1}{2}cos\frac{(2x+1)u\pi}{16} & u \ne 0 \\ \frac{1}{2\sqrt{2}} & u = 0 \end{cases}$$

And especially, $\boldsymbol{B}^{T3}$ is transpose of $\boldsymbol{B}$ in temporal direction. If $\bar{f}$ is the mean value and $f'(x,y,z)$ is the residual pixel, then we define

$$\bar{f} = \frac{1}{512}\sum_{x=0}^{7}\sum_{y=0}^{7}\sum_{z=0}^{7} f(x,y,z), \quad f'(x,y,z) = f(x,y,z) - \bar{f} \quad (2)$$

978-1-4244-2179-4/08/$25.00 ©2008 IEEE

$$F(u,v,w) = \begin{cases} 16\sqrt{2}\bar{f} & for\ u,v,w=0 \\ \frac{1}{8}\sum_{x=0}^{7}\sum_{y=0}^{7}\sum_{z=0}^{7} f'(x,y,z) cos\frac{(2x+1)u\pi}{16} cos\frac{(2y+1)v\pi}{16} cos\frac{(2z+1)w\pi}{16} & otherwise \end{cases} \quad (3)$$

TABLE I MATRIX $\beta(u,v,0)$

9.58	5.59	3.35	2.05	1.42	1.06	0.88	0.79
5.59	3.28	2.01	1.21	0.83	0.63	0.52	0.47
3.35	1.98	1.21	0.73	0.50	0.38	0.32	0.28
2.05	1.21	0.73	0.45	0.31	0.23	0.19	0.17
1.42	0.83	0.50	0.31	0.21	0.16	0.13	0.12
1.06	0.63	3.38	0.23	0.16	0.12	0.10	0.09
0.88	0.52	0.32	0.19	0.13	0.10	0.08	0.07
0.79	0.47	0.28	0.17	0.12	0.09	0.07	0.06

Each 3-D DCT coefficient can be calculated only by the mean value and the residual pixels as (3).

2.2 Statistical modeling of ZQDCT coefficients

Experiments show that the distribution of the residual pixel values can be modeled by Laplacian distribution with a significant peak at zero. To investigate the distribution of these residual values, we collected the residual pixel values from several benchmark sequences. The data suggest that the distribution of the residual pixel values yields the Laplacian distribution. As an example, Fig. 1 shows the distribution of the residual pixel values of Miss American and Akiyo and the ideal Laplacian with zero mean.

Like the motion-compensated pixels in video standards [7], the residual pixels $f'(x,y,z)$ are approximated by a Laplacian distribution with zero mean and a separable covariance variance σ. The variance of the (u,v,w)th DCT coefficient $\sigma_F^2(u,v,w)$ can be written as [8]

$$\sigma_F^2(u,v,w) = \sigma^2 \times \beta(u,v,w) \quad (4)$$

if we defin

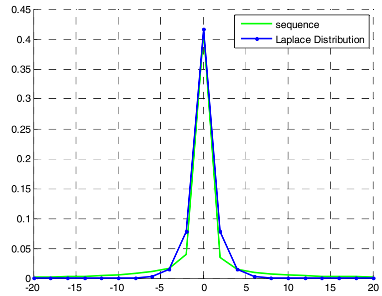
$$\beta(u,v,w) = [\boldsymbol{BRB^{T1}}]_{u,u}[\boldsymbol{BRB^{T2}}]_{v,v}[\boldsymbol{ARA^{T3}}]_{w,w} \quad (5)$$

where $[\cdot]_{u,u}$ is the (u,u)th component of a matrix and $\boldsymbol{R}$ is

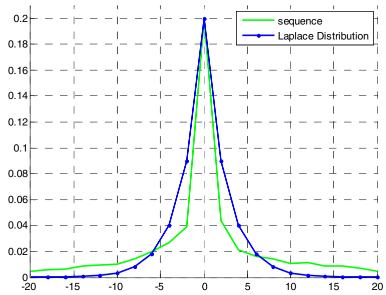
$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho^7 \\ \rho & 1 & \cdots & \rho^6 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^7 & \rho^6 & \cdots & 1 \end{bmatrix}$$

where ρ is the correlation coefficient. In this work, we set $\rho = 0.6$ in accordance with [7]. As an example, the matrix $\beta(u,v,0)$ is shown in Table I.

Eq. (4) and (5) show that the variances of the DCT coefficients can be estimated by the variances of the residual pixel values $f'(x,y,z)$. Moreover, it also shows that the up-left DCT coefficients have larger variances, which indicates that the probability of these coefficients to be quantized to zeros is smaller than those down-right coefficients.



(a)



(b)

Fig. 1. Distribution of the residual pixel $f'(x,y,z)$ of (a) Miss American, and (b) Akiyo. The blue line shows the ideal Laplacian distribution having a zero mean and a variance approximate to that of the collected data.

Since DCT is a unitary transform, the energy of both at the input of DCT and the output of DCT can be approximately expressed as the sum of absolute value of the residual pixel values (SAD)

$$SAD = \sum_{x=0}^{7}\sum_{y=0}^{7}\sum_{z=0}^{7}\left|f'(x,y,z)\right| \quad (6)$$

If the SAD is small, it indicates that the residual pixel vlaues are very close to zeros and the block contains little energy. That is, the DCT coeffcients have a higher probability to be zeros after quantization. This justifies setting the thresholds based on SAD.

Practically, the variance of pixel values with Laplacian distribution and zero mean can be estimated by SAD. They approximately satisfy

$$\sigma \approx \frac{\sqrt{2}SAD}{N} \quad (7)$$

where N is the number of coefficients, i.e. 512 in an $8 \times 8 \times 8$ DCT block.

Together with (4), (6) and (7), we get

$$\sigma_F(u,v,w) = \frac{\sqrt{2}SAD \times \sqrt{\beta(u,v,w)}}{N} \quad (8)$$

By the Laplacian theorem, the DCT coefficient $F^{'}(u,v,w)$ will fall within $(-\gamma\sigma_F(u,v,w),\ \gamma\sigma_F(u,v,w))$ with a probability controlled by the confidence parameter γ. Therefore, $F^{'}(u,v,w)$ will be truncated to zero if the quantization $Q_p(u,v,w) > \gamma\sigma_F(u,v,w)$, $\forall u,v,w \in \{0,\ldots,7\}$ with a very high probability. For instance, if $\gamma = 3$ and $Q_p(u,v,w) > \gamma\sigma_F(u,v,w)$, the probability of $F^{'}(u,v,w)$ to be quantized to zero is more than 99%. Derived from (8) a criterion for the ZQDCT coefficient $F^{'}(u,v,w)$ with high probabilitiy is

$$SAD < \frac{N \times Q(u,v,w)}{\sqrt{2\beta(u,v,w)} \times \gamma} \quad (9)$$

Given $N = 512, \gamma = 3$, the thresholds for ZQDCT coefficients are shown in Table II.

Based on the above analysis, we propose the following adaptive scheme to reduce the DCT and quantization computations. If the SAD satisfies (9), the calculations of $F^{'}(u,v,w)$ and the subsequent quantization is just omitted. For instance, if $SAD < 18.38 \times Q_p(1,1,0)$, we directly set $F^{'}(1,1,0)$ as zero. Otherwise, if $SAD \geq 444.3 \times Q_p(7,7,7)$, all the 511 DCT coefficients of the residual pixel values, $\forall u,v,w \in \{1,\ldots,7\}$, have to be computed.

TABLE II THRESHOLD MATRIX $T(u,v,0)$

10.74	14.06	18.05	23.15	27.96	32.10	35.22	37.16
14.06	18.38	23.61	30.28	36.57	41.98	46.06	48.60
18.05	23.61	30.32	38.89	46.97	53.92	59.15	62.42
23.15	30.28	38.89	49.87	60.24	69.15	75.86	80.06
27.96	36.57	46.97	60.24	72.76	83.53	91.63	96.69
32.10	41.98	53.92	69.15	83.52	95.89	105.2	111.0
35.22	46.06	59.15	75.86	91.63	105.2	115.4	121.8
37.16	48.60	62.42	80.06	96.69	111.0	121.8	128.5

TABLE III THRESHOLD MATRIX $T(u,v,7)$

37.16	48.60	62.42	80.05	96.69	111.0	121.7	128.5
48.60	63.57	81.64	104.7	126.5	145.2	159.3	168.1
62.42	81.64	104.8	134.5	162.4	186.5	204.5	215.8
80.06	104.7	134.5	172.5	208.3	239.1	262.3	276.8
96.69	126.5	162.4	208.3	251.6	288.8	316.8	334.4
111.0	145.2	186.5	239.1	288.8	331.5	363.7	383.8
121.8	159.3	204.5	262.3	316.8	363.7	399.0	421.0
128.5	168.1	215.8	276.8	334.4	383.8	421.1	444.3

3. Simulation results

A series of experiments were carried out with the baseline 3-D DCT codec. Four benchmark QCIF (176×144) sequences are tested. The quantization strategy in accordance with [5] is defined as

$$q\overrightarrow{[k]} = \lfloor Q_p(1 + k_1^p + k_2^p + k_3^p) \rfloor \quad (10)$$

where $q\overrightarrow{[k]}$ is the quantization step for 3-D DCT coefficient at position $\vec{k}[k_1,k_2,k_3]$, Q_p is the quantization parameter, and p is the parameter to control the rate of delay in the quantization volume. Here, it is fixed to 0.3.

The comparisons of the computational complexity about 3-D DCT and quantization between the proposed model and the original encoder are illustrated in Fig.2. In this figure, the required computational complexity for the proposed model is defined as

$$C = \frac{T_d}{T_d^o} \times 100\% \quad (11)$$

where T_d and T_d^o are the required computations of DCT and quantization for the proposed model and the baseline codec. Fig.2 gives the overall computational savings in 3-D DCT and quantization. It is obvious that the proposed model can effectively reduce redundant computations and achieve better performance than the original codec in terms of computational cost. In general, the average computations of 3-D DCT have been decreased by 55% compared to the original encoder, although the extent is slightly different for different video sequences and quantization parameters.

As two evaluation parameters, the false acceptance rate (FAR) and the false rejection rate (FRR) are provided to evaluate the proposed analytical model. The smaller the FAR, the less the video quality degrades and the smaller the FRR, the more efficient the predictive model. Therefore, it is desirable to have both small FAR and FRR for an efficiently predictive model.

From the experimental results in Table IV, some obvious conclusions can be drawn. Firstly, the proposed model can efficiently predict the ZQDCT coefficients. Compared to the baseline encoder, the proposed model can detect 70-85% of ZQDCT coefficients and thereby, it is desirable to avoid redundant computations. Secondly, the proposed model becomes more efficient along with increase of the quantization. Thirdly, the proposed model has a FAR arranging from 0.01 – 8.53%, which indicates that video quality degradation is observed. Usually, the more the distribution of the residual pixel values is approximate to the ideal Laplacian modeling, the smaller the FAR is. Together with Fig.1 and Table IV, the Miss American sequence has a more approximate shape to the ideal Laplacian distribution, therefore it has a smaller FAR.

The objective video quality is measured by the Peak Signal to Noise Ratio (PSNR). In Table V, a negative value actually means the PSNR degradation. Experiments show that the falsely classified non-zero coefficients are usually

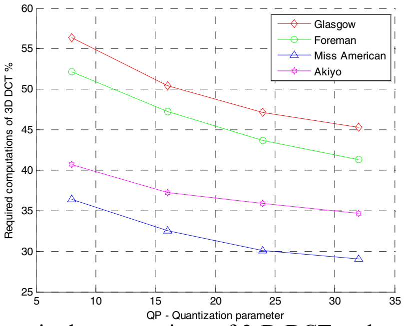


Figure.2 Required computations of 3-D DCT and quantization compared to the baseline 3-D DCT approach.

TABLE IV FAR AND FRR (%)

Image	Qp - p	FRR	FAR
Glassgow	8	33.16	8.53
	16	28.58	7.88
	24	26.86	4.32
	32	25.66	1.26
Foreman	8	30.53	5.47
	16	26.47	5.39
	24	24.65	2.35
	32	24.38	1.04
Akiyo	8	24.11	4.57
	16	20.06	2.63
	24	18.74	0.70
	32	17.30	0.02
Miss American	8	21.27	4.60
	16	19.38	1.88
	24	18.83	0.25
	32	18.66	0.01

the high frequency coefficients, thus it does not result in obvious PSNR degradation. According to the experiments, a bit visual quality degradation ranging from 0 to 0.572dB is observed, which is negligible.

As additional computations are introduced for the calculations of mean values and residual pixels, the entire encoding time is compared as shown in Fig.3. It is obvious that the proposed model achieves better real-time performance. This validates that the proposed model can reduce the computational complexity of the encoder and is much superior to the baseline encoder and the reference encoder.

4. Conclusions

This paper proposes a statistical model to predict zero-quantized 3-D DCT coefficients to avoid redundant DCT and quantization computations. The experimental results show that the proposed model can significantly improve the encoding efficiency with negligible visual degradation. Moreover, it can be directly applied to any other existing 3-D DCT schemes. Computational reduction also implies longer battery lifetime and energy economy for digital applications. Potential applications could be for portable digital devices with restrict battery lifetime and other areas with real-time requirement.

TABLE V VISUAL QUALITY COMPARISON

$Q_p - p$	Glassgow	Foreman	Akiyo	Miss American
8	-0.572	-0.318	-0.294	-0.231
16	-0.459	-0.297	-0.128	-0.072
24	-0.202	-0.149	-0.045	-0.013
32	-0.063	-0.062	-0.001	0

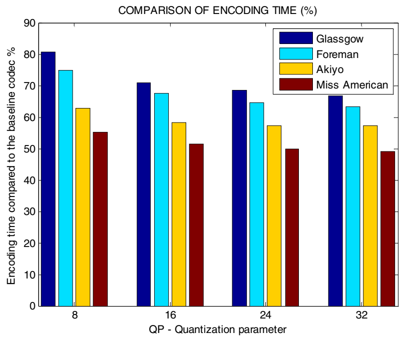


Fig. 3 Required encoding time compared to the baseline encoder

5. Acknowledgment

This work is supported by Chinese Science & Technology Ministry under Grant 2005DFA10300 and by Academy of Finland under Grant 117065. The authors thank Nokia Foundation for the support of research work.

6. References

[1] 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.

[2] 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.

[3] 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.

[4] A. Nokin, A. Gotchev, K. Egiazarian and J. Astola, "A Low-Complexity Multiple Description Video Coder Based on 3D-Transforms," EURASIP *Journal on Embedded Systems*, 2007.

[5] N. Bozinovic, and J. Konrad, "Motion Analysis in 3D DCT Domain and Its Application to Video Coding," EURASIP *Signal Process., Image Commun*., June,2005.

[6] Y. Zhao and H. Chen, "Study on Four-dimensional Matrix Model Theory and Compression Coding of Color Video," Doctoral thesis, Jilin University, March, 2003.

[7] I.M. Pao and M.T. Sun, "Modeling DCT Coefficients for Fast Video Encoding," IEEE *Transactions on Circuit and Systems for Video Technology*, VOL 9, No. 4, pp. 608-616, 1999.

[8] A.K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice.Hall, 1989.