LAPLACIAN MODELING OF DCT COEFFICIENTS FOR REAL-TIME ENCODING

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]

ABSTRACT

Digital image/video coding standards such as JPEG, H.264 are becoming more and more important for multimedia applications. Due to the huge amount of computations, there are significant efforts to speed up the encoding process. This paper proposes a Laplacian based statistical model to predict zero-quantized DCT coefficients in JPEG and to reduce the computations of encoding process. Compared with the standard JPEG and the reference in the literature, the proposed model can significantly simplify the computational complexity and achieve the best real-time performance at the expense of negligible visual degradation. Moreover, it can be directly applied to other DCT-based image/video codec. Computational reduction also implies longer battery lifetime and energy economy for digital applications.

***Index Terms*—** Discrete cosine transform (DCT), quantization, computational complexity, real-time encoding

1. INTRODUCTION

As a key compression technique, the discrete cosine transform (DCT) is widely used in image/video coding standards. However, as most DCT coefficients are quantized to zeros, a huge number of redundant computations are introduced. On the other hand, high computational complexity limits the real-time performance of the encoder as well as its application in digital portable devices, as they are still suffering from the lack of computational power. Thus, there is great significance to reduce the complexity for fast encoding.

Previously, the efforts to speed up the calculation of DCT were mainly focused on utilizing more efficient transform structure [1]-[3]. However, it doesn't reduce the redundant computations. As the structure for calculation of DCT is optimized, more efforts are focused on reducing the redundant computations of DCT coefficients. Most of these effects are on motion-compensated DCT blocks [4]-[5] and significant reductions are obtained. But they cannot be directly applied to the normal DCT in JPEG. Y. Nishida proposed a zero-value prediction for fast DCT calculation [6]. If several consecutively zero elements are produced during the DCT operation, the remaining transform is skipped. Although this method reduces the computations by 29% for DCT when applied to video coding, the video quality is degraded by 1.6dB averagely.

In this paper, we extend Pao's results to the normal DCT which is widely used in JPEG and intra block in video coding, aiming to reduce the encoding computations without much video quality degradation. Although the proposed model is implemented based on the 8×8 DCT in JPEG, it can be applied to other DCT based image/video standards. As a result, high prediction efficiency and good computational savings are achieved by the proposed model.

The rest of this paper is organized as follows. The mathematical decomposition of DCT is performed in Section 2. Laplacian distribution based statistical model is presented in section 3. Section 4 shows the experimental results. Finally, Section 5 concludes this paper.

2. MATHEMATICAL DECOMPOSITION OF DCT

In this paper, we mainly consider the 8×8 2-D DCT which is widely used in image/video coding standards. If we define $f(x,y)$ as the pixel value, $0 \le x,y \le 7$, the DCT coefficient $F(u,v), 0 \le u,v \le 7$, is computed by

$$F(u,v) = \frac{c(u)c(v)}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\frac{(2x+1)u\pi}{16} \cos\frac{(2y+1)v\pi}{16} \quad (1)$$

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

Alternatively, the DCT in (1) can be expressed in matrix form as $\boldsymbol{F} = \boldsymbol{A f A^T}$, where the uth row of $\boldsymbol{A}$ is the basis vector $1/2\ C(u) cos\,(2x+1)u\pi/16$.

If $\bar{f}$ is the mean value of the 64 pixels in an 8×8 DCT block and $f'(x,y)$ is the residual pixel value, we define

$$\bar{f} = \frac{1}{64} \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y), \quad f'(x,y) = f(x,y) - \bar{f} \quad (2)$$

Therefore, an 8×8 DCT block is decomposed into a mean value $\bar{f}$ and an 8×8 residual block $f'(x,y)$. Then, each DCT coefficient can be respectively computed by $\bar{f}$ and the residual pixel values $f'(x,y)$ as (3).

$$F(u,v) = \begin{cases} 8\bar{f} & for\, u,v = 0 \\ \dfrac{c(u)c(v)}{4}\displaystyle\sum_{x=0}^{7}\sum_{y=0}^{7} f'(x,y) cos\frac{(2x+1)u\pi}{16} cos\frac{(2y+1)v\pi}{16} & otherwise \end{cases} \quad (3)$$

From (3), 63 out of the 64 DCT coefficients can be directly calculated from the residual block. Basically, if we can efficiently predict the zero-quantized DCT (ZQDCT) coefficients for the 63 coefficients, a lot of computations will be saved.

3. MODELING OF THE RESIDUAL DCT BLOCK

The distribution of the residual pixel value $f'(x,y)$ 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 images (Couple, Airplane, Peppers and Girl). The data suggest that the distribution of the residual pixel values yields a Laplacian distribution. As an example, Fig. 1 shows the distribution of Airplane and Couple.

Like the motion-compensated pixels in video standards [4], the residuals $f'(x,y)$ are approximated by a Laplacian distribution with zero mean and a separable covariance variance σ. The variance of the (u,v)th DCT coefficient $\sigma_F^2(u,v)$ can be written as [7]

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

if we define

$$\beta(u,v) = [\boldsymbol{ARA^T}]_{u,u}[\boldsymbol{ARA^T}]_{v,v} \quad (5)$$

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

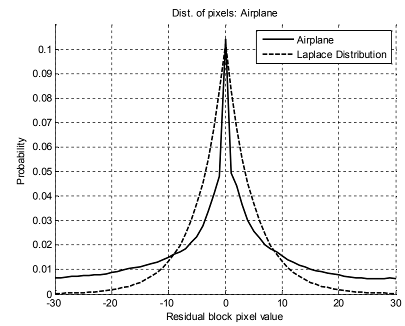
$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho^7 \\ \rho & 1 & \cdots & \rho^6 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^7 & \rho^6 & \cdots & 1 \end{bmatrix}$$

where ρ is the correlation coefficient. In this work, we set $\rho = 0.6$ in accordance with [4] and [5]. The matrix $\boldsymbol{\beta}$ is shown in Table I.

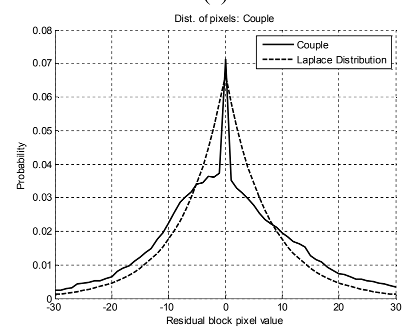
Eq. (4) shows that the variances of the DCT coefficients can be estimated by the variances of the residual pixel value $f'(x,y)$. Moreover, it also shows that the up-left DCT coefficients have larger variances, which indicates that the probability of these coefficients to be quantized to zeros is smaller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

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 residuals (SAD)

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



(a)



(b)

Fig1. Distribution of residual pixel $f^r(x,y)$ of (a) Airplane at and (b) Couple. The dashed line shows the ideal Laplacian distribution with a zero mean and a variance approximate to that of the collected data.

If the SAD is small, it indicates that the energy is small and the DCT coeffcients will have a higher probability to be quantized to zero. This justifies setting the thresholds based on SAD.

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

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

where N is the number of coefficients, i.e. 64 in this work.

Together with (4), (5) and (7), we get

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

By the Laplacian theorem, the DCT coefficients $F'(u,v)$ will fall within $(-\gamma\sigma_F(u,v),\ \gamma\sigma_F(u,v))$ with a probability controlled by the confidence parameter γ. Therefore, $F'(u,v)$ will be truncated to zero if the quantization parameter $Q_p(u,v) > \gamma\sigma_F(u,v)$, $\forall u,v \in \{0,\ldots,7\}$. If $\gamma = 3$, the probability of $F'(u,v)$ to be quantized to zero is more

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

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

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

4.87	6.38	8.24	10.55	12.67	14.62	16.04	16.93
6.38	8.33	10.63	13.74	16.57	19.05	20.92	22.07
8.24	10.72	13.74	17.64	21.27	24.46	26.85	28.27
10.55	13.74	17.64	22.60	27.30	31.28	34.39	36.25
12.67	16.57	21.27	27.21	32.97	37.84	41.56	43.78
14.62	19.05	24.46	31.28	37.84	43.45	47.68	50.34
16.04	20.92	26.85	34.39	41.56	47.68	52.29	55.21
16.93	22.07	28.27	36.25	43.78	50.34	55.21	58.23

than 99%. Derived from (8) a criterion for zero quantized DCT coefficient $F'(u,v)$ with high probabilities is

$$SAD < N \times Q(u,v)/(\sqrt{2\beta(u,v)} \times \gamma) \qquad (9)$$

Given $N = 64, \gamma = 3$, the threshold T is shown in Table II.

Based on the above analysis, we propose the following adaptive scheme to reduce the DCT and quantization computations. If SAD satisfies (9), the computation of $F'(u,v)$ and the corresponding quantization is just omitted. For instance, if $SAD < 8.33 \times Q_p(u,v)$, we directly set $F'(1,1)$ as zero. Otherwise, if $SAD > 58.23 \times Q_p(u,v)$, all the DCT coefficients are computed.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed model, a series of experiments were carried out with JPEG. Four benchmark images are tested. All the simulations are running on a PC with Intel Pentium 3.2G and 1.5Gbytes of RAM. The quantization strategy is in accordance with [8] where two quantization tables are used for luminance and chrominance transform. Moreover, a scaling factor p is used to get various size of compressed bit stream.

First, we will study the computational complexity of the proposed model. The comparison of the complexity about DCT and quantization are illustrated in Table III. The computational reduction is assessed as

$$C = \frac{T_d}{T_d^o} \times 100\% \qquad (10)$$

where T_d and T_d^o are the required encoding time of DCT and quantization for the proposed model and the baseline codec.

It is obvious that the proposed algorithm can effectively reduce the redundant computations and achieve better performance in terms of computational cost. In general, the

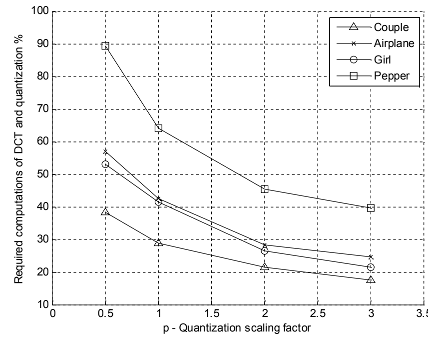


Fig.2 Computational reduction of DCT and quantization ($C\%$)

average required computations of DCT and quantization by the proposed model have been decreased by 60%.

As two important evaluation parameters, the false acceptance rate (FAR) and the false rejection rate (FRR) are provided to evaluate the proposed model. Normally, 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 efficient predictive model.

From the experimental results in Table III, 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 20-85% 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. Take Couple as an example, the FRR is 41.53% when $Q_p = 0.5$, and then decreases to 14.57% when $Q_p = 3.0$. This means that the proposed model is especially suitable for low bit-rate encoding cases. Thirdly, the proposed model has a FAR arranging from 0.60 - 9.88%, 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 III, the Couple image has the most approximate shape to the ideal Laplacian distribution, therefore it has the smallest FAR.

Finally, we continue to study the visual quality and the encoding time of the proposed algorithm. Table IV shows the visual degradation measured by the Peak Signal to Noise Ratio (PSNR). In Table IV, a negative value actually means the PSNR degradation. Experiments show that the falsely classified non-zero coefficients are usually the high frequency coefficients, thus it does not result in obvious PSNR degradation. Moreover, along with a non-zero FAR, the skipped calculations of DCT not only reduce the computations but also the bits required to code these coefficients. Therefore, the compression efficiency of the proposed model is even slightly higher than the baseline encoder at the same quantization.

TABLE III FAR AND FRR (%)

Image	Qp - p	FRR	FAR
Couple	0.5	41.53	2.25
	1.0	29.18	0.68
	2.0	19.77	0.60
	3.0	14.57	2.19
Airplane	0.5	48.53	5.74
	1.0	36.47	6.83
	2.0	25.65	3.27
	3.0	20.38	4.24
Girl	0.5	54.11	2.57
	1.0	39.06	8.63
	2.0	27.74	3.70
	3.0	22.30	4.42
Peppers	0.5	81.16	7.60
	1.0	60.58	9.88
	2.0	42.86	4.25
	3.0	34.66	0.86

Since additional computations are introduced for the calculation of mean values and residual pixels, we also compared the entire encoding time as Fig.3. The encoding time reduction ∇T is defined as

$$\nabla T = \frac{T}{T_{org}} \times 100\% \qquad (11)$$

where T and T_{org} are the entire encoding time of the proposed model and the JPEG encoder.

Fig.3 shows that our analytical model achieves better real-time performance than original codec. This validates that the proposed model can reduce the computational complexity of the encoder, which is more suitable for real-time encoding and digital portable devices.

A series of experiments were carried out to compare the proposed model with the reference encoder [6]. In the reference encoder, if two consecutive zero coefficients are produced, we just skip the following DCT and quantization. Experimental results show that although they have a comparative FAR, the proposed model greatly outperforms the reference encoder in terms of computational reduction, FRR and the encoding time, e.g. the average FRR in the reference encoder is around 70% and obviously inferior to the proposed Laplacian model. Compared to the reference encoder, the proposed method has a higher prediction efficiency and precision to detect ZQDCT coefficients.

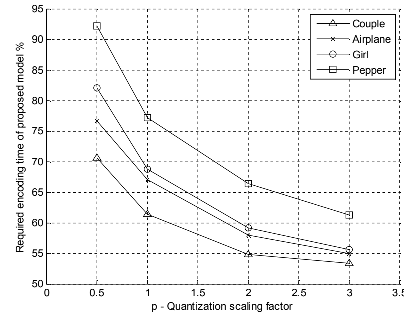


Fig.3 Comparison of required encoding time

TABLE IV VISUAL QUALITY COMPARISON

$Q_p - p$	Couple	Airplane	Girl	Peppers
0.5	-0.191	-0.261	-0.217	-0.631
1.0	-0.029	-0.407	-0.776	-0.804
2.0	-0.029	-0.238	-0.235	-0.238
3.0	-0.132	-0.283	-0.386	-0.092

5. CONCLUSION

This paper proposes a Laplacian based statistical model to predict ZQDCT coefficients, aiming to reduce the redundant computations and to achieve better real-time performance. Based on the mathematical decomposition of the pixel values at the input of DCT, we derive a prediction algorithm under which each DCT coefficient becomes zero at high probability. Finally, the transform and quantization of ZQDCT coefficients is skipped. Experiments show that the proposed model can improve the encoding process at the expense of negligible video quality degradation. Moreover, it outperforms the reference encoder [6] in the literature. The proposed model can be also directly applied to other DCT-based image/video standards. Computational reduction also implies longer battery lifetime and energy economy for digital applications.

6. ACKNOWLEDGEMENT

This work is supported partly 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 the research work.

7. REFERENCES

[1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90–93, Jan. 1974.

[2] Y. Chan and W. Siu, "Mixed-radix discrete cosine transform," *IEEE Trans. Signal Process.*, vol. 41, pp.3157–3161, Nov. 1993.

[3] K. Yamatani and N. Saito, "Improvement of DCT-Based Compression Algorithms Using Poisson's Equation," in *IEEE Tran. Image Process.*, pp. 3272–3289, Dec.2006.

[4] I.M. Pao and M.T. Sun, "Modeling DCT Coefficients for Fast Video Encoding," *IEEE Tran. Circuit and Systems for Video Tech.*, Vol. 9, No. 4, pp. 608-616, 1999.

[5] H.L. Wang, S. Kwong and W.C. Kok, "Fast Video Coding Based on Gaussian Model of DCT Coeffcients," *IEEE Proc. ISCAS*, 2006.

[6] Y. Nishida, K. Inoue and V. G. Moshnyaga, "A Zero-Value Prediction Technique for Fast DCT Computation," *IEEE Workshop on Sig. Proc. Syst.*, pp. 165-170, Aug. 2003.

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

[8] W. Pennebaker and J. Mitchell, "*JPEG, Still Image Data Compression Standard*," Van Nostrand Reinhold, 1993.