GAUSSIAN MODEL BASED APPROACH TO REDUCING DCT COMPUTATIONS IN H.264

Hanli Wang and Sam Kwong

Department of Computer Science, City University of Hong Kong, Hong Kong, P. R. China
E-mail: wanghl@cs.cityu.edu.hk, cssamk@cityu.edu.hk

ABSTRACT

This paper presents an efficient method to predict zero quantized DCT coefficients in order to reduce redundant computations in H.264 encoding. The Gaussian distribution is firstly applied to study the quantized integer DCT coefficients in H.264 and then an adaptive scheme with multiple thresholds is proposed to realize different types of DCT and quantization implementations. Compared with other approaches in the literature, the experimental results demonstrate that the proposed method can achieve the best performance in reducing computations and obtain almost the same rate-distortion performance as the original encoder.

Index Terms— H.264, integer DCT, quantization, Gaussian model.

1. INTRODUCTION

The H.264/AVC video standard [1] has gained much attention recently due to its substantial coding gain over existing coding standards. The complexity of encoder, however, is dramatically increased. Therefore, there is an emerging interest to reduce the computational complexity of H.264 encoding. In the research arena, the efforts are mainly focused on fast mode selection and motion estimation algorithms. After the mode selection and motion estimation are optimized, we also need to optimize other important functions such as DCT and quantization (Q). In video coding, it is quite common that a substantial number of DCT coefficients of the prediction residue become zeros after Q. Hence, considerable computations can be saved if there exists a method which can early detect zero quantized DCT (ZQDCT) coefficients prior to DCT and Q.

Several approaches have been proposed to early detect ZQDCT coefficients for the 8×8 DCT based video encoders such as MPEG-4. Zhou *et al.* [2] perform theoretical analyses on the range of DCT coefficients and derive a sufficient condition to detect all-zero DCT blocks. This approach [2] is further refined by Sousa [3], where a tighter sufficient condition is derived to detect all-zero blocks. In [4], an analytical model is proposed to predict individual ZQDCT coefficients and an adaptive scheme is proposed to perform different kinds of DCT implementations. All of the three approaches [2, 3, 4] obtain the same rate-distortion (R-D) performance as the original encoder since they all provide sufficient conditions to detect ZQDCT coefficients. In a different manner, Pao *et al.* [5] propose a Laplacian distribution based model to predict ZQDCT coefficients. Based on this statistical model, multiple thresholds are derived to detect various sizes of non-zero blocks and thus to skip DCT and Q computations in the other part of an 8×8 block. Similar to [5], a Gaussian distribution based model [6] is proposed to early detect ZQDCT coefficients. Compared with [5], the Gaussian distribution based model [6] is able to detect more ZQDCT coefficients and thus reduce more redundant computations. As far as H.264 encoding is concerned, however, the aforementioned methods can not be applied since H.264 uses the integer 4×4 DCT and a scaling multiplication is integrated into the quantizer to avoid divisions for quantization. In [7], after examining the properties of DCT and Q in H.264, Moon *et al.* propose a sufficient condition to detect all-zero 4×4 DCT blocks.

In this paper, we propose a Gaussian distribution based model by extending our previous work [6] to reduce computations in H.264. Based on this model, four thresholds are derived to determine five kinds of DCT, Q, IQ and IDCT implementations: *Skip*, 1×1, 2×2, 3×3 and 4×4. Experimental results demonstrate that the proposed approach is superior to the approach [7] in terms of reducing computations and obtains almost the same R-D performance as the original encoder. The rest of this paper is organized as follows. The DCT and Q in H.264 are discussed in Section 2. In Section 3, the proposed approach is presented. Experimental results are shown in Section 4. Finally, Section 5 concludes this paper.

2. INTEGER DCT AND Q IN H.264

In H.264, an integer DCT is applied to 4×4 blocks of residual data and avoids inverse transform mismatch problems. For a 4×4 residual block $f(x, y)$, $0 \le x, y \le 3$, the integer transform is

$$F(u,v) = \sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y)\cdot A(x,u)\cdot A(y,v) \tag{1}$$

where

$$A(m,n) = \langle \frac{2.5C(n)}{\sqrt{2}} \cos\frac{(2m+1)n\pi}{8} \rangle \tag{2}$$

This work is supported by City University of Hong Kong Strategic Grant 7002023.

$C(n) = 1/\sqrt{2}$, for $n = 0$; and $C(n) = 1$, otherwise. The operator $\langle x \rangle$ denotes to round the operand x to the nearest integer. Given an integer DCT coefficient $F(u, v)$ and a quantization parameter Q_p ranging from 0 to 51, the quantized coefficient $Z(u, v)$, $0 \leq u, v \leq 3$, is resulted from

$$Z(u,v) = sign(F(u,v)) \cdot (|F(u,v)| \cdot M(u,v) + c) >> qbits \quad (3)$$

where $qbits = 15 + floor(Q_p/6)$, $>>$ indicates the binary shift right, c is $\langle 2^{qbits}/3 \rangle$ for intra blocks or $\langle 2^{qbits}/6 \rangle$ for inter blocks. $M(u, v)$ is the multiplication factor related to $Q_p\%6$, and three categories can be classified depending on the positions listed in Table 1.

Table 1. Multiplication factor M.

$Q_p\%6$	**Category 1** $(1,1), (1,3)$ $(3,1), (3,3)$	**Category 2** Others	**Category 3** $(0,0), (0,2)$ $(2,0), (2,2)$
0	5243	8066	13107
1	4660	7490	11916
2	4194	6554	10082
3	3647	5825	9362
4	3355	5243	8192
5	2893	4559	7282

3. PROPOSED METHOD FOR ZQDCT DETECTION

From (3), the sufficient condition for $F(u, v)$ to be quantized to zero can be given as

$$|F(u,v)| < T(u,v), \quad T(u,v) = \frac{2^{qbits} - c}{M(u,v)} \quad (4)$$

Suppose the residual pixel values $f(x, y)$ at the input of DCT are approximated by a Gaussian distribution with zero mean and variance σ as

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \ -\infty < x < +\infty \quad (5)$$

The expected value of $|x|$ can be calculated as

$$E[|x|] = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}}\sigma \quad (6)$$

Since $E[|x|]$ can be approximated as

$$E[|x|] \approx \frac{SAD}{N} \quad (7)$$

where N is the number of coefficients (i.e., 16 for a 4×4 block), and SAD is the sum of absolute difference, i.e., $SAD = \sum_{x=0}^{3}\sum_{y=0}^{3}|f(x,y)|$. Hence, we can get

$$\sigma \approx \sqrt{\frac{\pi}{2}} \frac{SAD}{N} \quad (8)$$

Note that the variance of the (u, v)th DCT coefficient $\sigma_F^2(u, v)$ can be given as [8]

$$\sigma_F^2(u,v) = \sigma^2 [ARA^T]_{u,u} [ARA^T]_{v,v} \quad (9)$$

where $[\cdot]_{u,u}$ is the (u, u)th component of a matrix, A is given in (2), and R is

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \quad (10)$$

where ρ is the correlation coefficient. Typically, ρ ranges from 0.4 to 0.75, and in this work we set $\rho = 0.6$ which is the same as in [5, 6]. According to the central limit theorem, the DCT coefficients $F(u, v)$ can be approximately distributed as Gaussian and will be quantized to zeros with a probability controlled by γ in the following form:

$$\gamma\sigma_F(u,v) < T(u,v), \ 0 \leq u, v \leq 3 \quad (11)$$

If $\gamma = 3$, then the probability of the DCT coefficient equal to zero after Q is about 99.73%. Derived from (8), (9) and (11), a criterion for ZQDCT coefficient prediction in (u, v) is

$$SAD < TS^g(u,v), \quad TS^g(u,v) = \beta_G(u,v) \times T(u,v) \quad (12)$$

where

$$\beta_G(u,v) = \frac{\sqrt{2}N}{\gamma\sqrt{\pi[ARA^T]_{u,u}[ARA^T]_{v,v}}} \quad (13)$$

As shown in (4), $T(u, v)$ is a function of Q_p and (u, v). Therefore, the thresholds $TS^g(u, v)$ in (12) are dependant on Q_p and (u, v). After observing $TS^g(u, v)$ values for each Q_p, we find that TS^g is a symmetric matrix (i.e., $TS^g(u, v) = TS^g(v, u)$) and the following formula holds true.

$$m > n \Rightarrow TS^g(u,m) > TS^g(u,n), \quad 0 \leq m, n \leq 3 \quad (14)$$

On the other hand, according to (1), (2) and (4), we can get a sufficient condition to detect all-zero DCT blocks in H.264 as

$$SAD < TS^*, \quad TS^* = \frac{T(1,1)}{4} \quad (15)$$

And we observe that for each Q_p, the following relation is always true.

$$TS^g(0,1) > TS^* > TS^g(0,0) \quad (16)$$

Therefore, we can employ four thresholds TS^*, $TS^g(0, 1)$, $TS^g(0, 2)$ and $TS^g(0, 3)$ to implement five types of DCT, Q, IQ and IDCT implementations as described in Table 2.

In practice, the row-column butterfly based flow structure is applied to implement the 4×4 DCT/IDCT [9]. Since it is only necessary to calculate parts of the coefficients in the 1×1, 2×2 and 3×3 types, i.e., the computations related to the ZQDCT coefficients are skipped, the DCT/IDCT implementation can be optimized accordingly. Regarding Q/IQ, the elements which are predicted as zeros are directly set to zeros to save computations. To illustrate the computational complexity of different types of DCT/Q/IQ/IDCT implementations occurred in the reference software JM9.5 [10], the number of addition (ADD), multiplication (MUL), shift (SFT) and comparison (CMP) operations are listed in Table 3.

Table 2. DCT, Q, IQ and IDCT Implementation.

Type	Condition	DCT/Q/IQ/IDCT
Skip	$SAD < TS^{*}$	Not performed
1×1	$TS^{*} \leq SAD < TS^{g}(0,1)$	Only calculate the DC coefficient.
2×2	$TS^{g}(0,1) \leq SAD < TS^{g}(0,2)$	Only calculate 2×2 low frequency DCT.
3×3	$TS^{g}(0,2) \leq SAD < TS^{g}(0,3)$	Only calculate 3×3 low frequency DCT.
4×4	$TS^{g}(0,3) \leq SAD$	Calculate all 16 coefficients.

Table 3. Computational complexity of DCT/Q/IQ/IDCT.

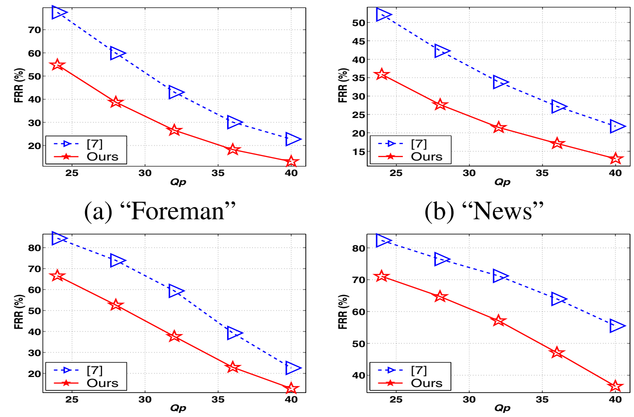
	ADD	MUL	SFT	CMP
4×4	192	32	112	64
3×3	147	18	76	50
2×2	108	8	58	40
1×1	57	2	37	34

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed approach, the reference software JM9.5 [10] is used to carry out experiments. The fast motion estimation is enabled and the number of reference frames is set to 1. All the block search sizes are enabled for mode selection. Four benchmark video sequences are used, each of which is of CIF format (352×288). They are "Foreman", "News", "Table Tennis" and "Mobile Calendar". These video sequences have different motion types from low to high and different spatial details from simple to complex. In order to examine the performance at different bit rates, five Q_p values, 24, 28, 32, 36 and 40, are used in our experiments. For comparison, the approach [7] is implemented.

The false rejection rate (FRR) and false acceptance rate (FAR) are provided to compare the prediction capacity of ZQDCT coefficients. The smaller the FRR is, the more efficiently the approach detects ZQDCT coefficients. The smaller the FAR is, the less the video quality degrades. Thus, it is more desirable to have small FAR and FRR values for efficient prediction approaches. The FRR curves are shown in Fig. 1, where we can see that the proposed approach obtains the lowest FRR results in all the cases indicating that it is more efficient to predict ZQDCT coefficients than the approach [7].

Regarding the FAR results, since the approach [7] provides the sufficient condition to detect all-zero blocks, zero FAR values are obtained by it. So we only list the FAR results obtained by the proposed approach in Table 4. In addition, the percentage for ZQDCT coefficients (PZQ) is listed in Table 4. From the results shown in Table 4, we can see that the ZQDCT coefficients occupy a great portion of the whole. And along with the increase of Q_p, more DCT coefficients are quantized to zeros. As shown in Table 4, the FRR value of the proposed approach is so insignificant that almost no video quality degradation is observed (we will present the video quality performance as follows).



(a) "Foreman" (b) "News"

(c) "Table Tennis" (d) "Mobile Calendar"

Fig. 1. FRR curves.

Table 4. FAR and PZQ.

	"Foreman"		"News"	
Q_p	PZQ	FAR	PZQ	FAR
24	92.74%	0.23%	94.53%	0.11%
28	95.73%	0.27%	96.25%	0.14%
32	97.55%	0.36%	97.52%	0.16%
36	98.57%	0.41%	98.42%	0.14%
40	99.16%	0.20%	98.99%	0.10%
	"Table Tennis"		"Mobile Calendar"	
Q_p	PZQ	FAR	PZQ	FAR
24	88.70%	0.05%	73.98%	0.04%
28	93.49%	0.10%	82.36%	0.07%
32	96.59%	0.15%	89.24%	0.15%
36	98.25%	0.20%	93.80%	0.23%
40	99.07%	0.23%	96.59%	0.26%

Now, we study the encoded video quality and bit rates resulted from the proposed approach. The video quality is objectively evaluated in terms of the peak signal-to-noise ratio (PSNR, dB). The performances of PSNR and bit rates are presented in the following form:

$$\Delta P = P - P_{org}, \quad \Delta R = \frac{R - R_{org}}{R_{org}} \times 100\% \tag{17}$$

where P and P_{org} are the PSNR criterion of the test approach and the original encoder; R and R_{org} are the encoded bit rates of the test approach and the original encoder, respectively. From the experimental results, the approach [7] does not degrade the video quality since it provides the sufficient condition to detect all-zero blocks and obtains the same PSNR and bit rates performances as the original encoder. Therefore, we only present the average PSNR and bit rates results obtained by our approach in Table 5. As shown in Table 5, the average PSNR loss is 0.002 dB and thus negligible. Moreover, the average bit rate has been reduced by 0.07% on average, which indicates that the proposed approach can achieve better bit rate performance than the original encoder. Therefore, the proposed approach accomplishes a trade-off between the encoded video quality in terms of PSNR and the compression efficiency in terms of bit rate, and retains almost the same R-D performance as the original encoder. We also evaluate the video quality subjectively and observe that there is no difference between the proposed approach and the original encoder.

Table 5. Average ΔP (**dB**) and ΔR.

Sequence	ΔP	ΔR
"Foreman"	-0.001	-0.10%
"News"	-0.005	-0.11%
"Table Tennis"	0	-0.08 %
"Mobile Calendar"	-0.002	0.01%
Average	-0.002	-0.07%

Finally, the reduction in computational complexity of DCT, Q, IQ and IDCT procedures is studied. In this work, the computational complexity of DCT, Q, IQ and IDCT of the test approaches is evaluated as $C = O_t/O_o \times 100\%$, where O_t is the number of one of the following operations: ADD, MUL, SFT and CMP, required in DCT, Q, IQ and IDCT of the test approach and O_o is the number of the corresponding operation used in DCT, Q, IQ and IDCT of the original encoder. Note that the thresholds for implementing our approach are pre-computed and used through a look-up table. The only overheads introduced by the proposed approach are the CMP operations, which are also considered into the complexity computations. Due to the space limit, we only present the average results for each of the test sequences in Fig. 2, where we can see that the proposed approach achieves better performances in reducing H.264 encoder computations than that of [7].

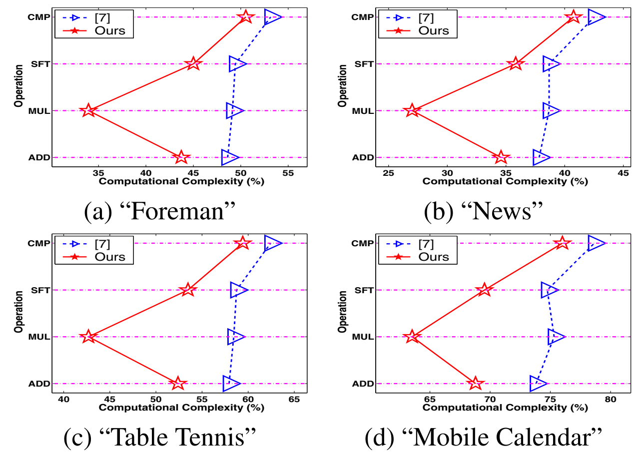


Fig. 2. Computational complexity of ADD, MUL, SFT and CMP.

5. CONCLUSIONS

In this paper, the Gaussian distribution model is applied to predict ZQDCT coefficients in order to reduce redundant DCT and Q computations in H.264 encoding. Derived from the proposed model, four thresholds are afforded to perform five kinds of DCT, Q, IQ and IDCT implementations: *Skip*, 1×1, 2×2, 3×3 and 4×4. The experimental results demonstrate that the proposed approach outperforms the approach [7] to reduce encoding computational complexities and achieves almost the same R-D performance as the original encoder. In the future, we will extend the current work into studying the 8×8 DCT which is used in the Fidelity Range Extensions (FRExt) [11] of H.264.

6. REFERENCES

[1] "Advanced Video Coding for Generic Audiovisual Services", *ISO/IEC 14496-10:2005(E) ITU-T Rec. H.264(E)*, Mar. 2005.

[2] X. Zhou, Z. Yu, and S. Yu, "Method for Detecting All-Zero DCT Coefficients Ahead of Discrete Cosine Transformation and Quantization", *Electro. Lett.,* vol. 34, no. 19, pp. 1839-1840, Sept. 1998.

[3] L. A. Sousa, "General Method for Eliminating Redundant Computations in Video Coding", *Electron. Lett.,* vol. 36, no. 4, pp. 306-307, Feb. 2000.

[4] H. Wang, S. Kwong, and C. W. Kok, "Analytical Model of Zero Quantized DCT Coefficients for Video Encoder Optimization", *IEEE Int. Conf. Multimedia & Expo (ICME'06)*, pp. 801-804, July 2006.

[5] I. M. Pao and M. T. Sun, "Modeling DCT Coefficients for Fast Video Encoding", *IEEE Trans. Circuits Syst. Video Technol.,* vol. 9, no. 4, pp. 608-616, Jun. 1999.

[6] H. Wang, S. Kwong, and C. W. Kok, "Fast Video Coding Based on Gaussian Model of DCT Coefficients", *IEEE Int. Symp. Circuits Syst. (ISCAS'06)*, pp. 1703-1706, May 2006.

[7] Y. H. Moon, G. Y. Kim, and J. H. Kim, "An Improved Early Detection Algorithm for All-Zero Blocks in H.264 Video Encoding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1053-1057, Aug. 2005.

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

[9] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-Complexity Transform and Quantization in H.264/AVC", *IEEE Trans. Circuits Syst. Video Technol.,* vol. 13, no. 7, pp. 598-603, Jul. 2003.

[10] H.264/AVC Reference Software JM9.5, [Online] Available: http://iphome.hhi.de/suehring/tml/.

[11] "Draft Text of H.264/AVC Fidelity Range Extensions Amendment", *Document JVT-L047*, Sept. 2004.