Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 25 (2007) 922–933
www.elsevier.com/locate/imavis

Efficient predictive model of zero quantized DCT coefficients for fast video encoding

Hanli Wang [a], Sam Kwong [a,*], Chi-Wah Kok [b]

[a] Department of Computer Science, City University of Hong Kong, Hong Kong, China
[b] Independent Consultant, Hong Kong, China

Received 21 April 2005; received in revised form 23 June 2006; accepted 12 July 2006

Abstract

Discrete cosine transform (DCT), quantization (Q), inverse quantization (IQ) and inverse DCT (IDCT) are the building blocks in video coding standards adopted by ITU-T and MEPG. Under these standards, a lot of computations are required to perform the DCT, Q, IQ and IDCT operations. With this concern, a novel statistical model based on Gaussian distribution is proposed to predict zero quantized DCT (ZQDCT) coefficients in order to reduce the computational complexity of video encoding. Compared with other predictive models in the literature, the proposed model can detect more ZQDCT coefficients. Simulation results demonstrate that the proposed statistical model is superior to others in terms of speeding up video encoders. Moreover, a hybrid model is derived based on the proposed statistical model and mathematical analysis of individual DCT coefficients to further improve the encoding efficiency.

Keywords: Discrete cosine transform (DCT); Quantization (Q); Zero quantized DCT (ZQDCT) coefficients; Video encoding

1. Introduction

With the advent of internet and multimedia systems, the hybrid DCT motion-compensated approach has gained widespread popularity for video encoding standards such as MPEG-4 [1], H.263 [2] and H.264 [3]. The discrete cosine transform (DCT), motion estimation (ME) and motion compensation (MC), quantization (Q), inverse quantization (IQ) and inverse DCT (IDCT) have been the building blocks for these video coding standards, where the temporal redundancy among video frames is removed by taking the difference between the current frame and the reference frame through motion estimation and compensation. Then, the difference is further processed by DCT and Q to remove the spatial redundancy and achieve compression. Such an architecture, commonly used today, has a performance-critical feedback loop consisting of DCT, ME, Q, IQ and IDCT stages. In fact, there is a significant interest and research in reducing these computations. Previously, the efforts to reduce the computations of video encoders are mainly focused on fast ME algorithms. However, as the ME algorithm becomes optimized, we also need to optimize other functions to further speed up video encoding, such as the DCT, and so forth. In digital video coding, especially in the very low bit rate coding, it is quite common that a substantial number of DCT coefficients of the prediction difference are quantized to zeros. Therefore, considerable computations may be saved if there is a method that can detect zero quantized DCT (ZQDCT) coefficients, i.e., the DCT coefficients equal to zero after Q, before implementing the DCT and Q.

A number of early detection techniques of all-zero DCT blocks have been studied for the 8×8 DCT based video encoders such as H.263 and MPEG-4. Chen et al. [4] propose to compare the signal energy with a threshold, and set all the DCT coefficients of one block to zeros if the

* Corresponding author. Tel.: +86 852 2788 7704; fax: +86 852 2788 8614.
E-mail address: cssamk@cityu.edu.hk (S. Kwong).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.07.007

signal energy is less than the threshold. In a similar manner, Yu et al. [5] propose to compare the sum of absolute difference *SAD* available from ME with the product of the quantization parameter Q_p and a predetermined threshold T. If $SAD < T \times Q_p$, then the DCT and Q computations can be skipped, and the quantized DCT coefficients are all set to zeros. This model is shown to be effective in reducing the computational complexity of the H.263 encoder. However, the quality of the encoded video is heavily dependent on the threshold T, where to define a suitable value is not trivial. In order to reduce the degradation of video quality, Yu et al. [6] design the threshold value experimentally to detect all-zero DCT blocks. In a different manner, Zhou et al. [7] perform theoretical analyses on the range of the DCT coefficients and propose a sufficient condition to detect all-zero DCT blocks. Thus, the redundant DCT and Q computations can be avoided without video quality degradation. Zhou's model [7] is further refined by Sousa [8] where a tighter sufficient condition is derived to obtain more reductions of the computational complexity without video quality degradation. Regarding the H.264 encoder, the aforementioned methods can not be directly applied since H.264 uses the integer 4×4 DCT and a scaling multiplication is integrated into the quantizer to avoid divisions for quantization [9]. In [10], after examining the properties of DCT and Q in H.264, Kim et al. propose a sufficient condition to detect all-zero 4×4 DCT blocks. In [11], the authors perform a comprehensive analysis of the dynamic range of DCT coefficients in H.264 and derive a more efficient sufficient condition than that of [10] to early detect all-zero DCT blocks. As a result, more redundant computations can be saved by using the method in [11].

Note that the methods mentioned above only consider the possibility to detect all-zero DCT blocks. Indeed, the prediction scheme should not be limited to the block level detection, and higher prediction efficiency can be achieved if more effective models are applied. In [12], Pao et al. propose a Laplacian distribution based statistical model for ZQDCT coefficients prediction. Based on this statistical model, multiple thresholds are derived to detect different size of non-zero blocks such 1×1 non-zero block, 2×2 non-zero block, etc, and the DCT and Q computations in the other part of the 8×8 block can be skipped accordingly. Therefore, an adaptive method with multiple thresholds is developed to reduce the computations of DCT, Q, IQ and IDCT for 8×8 DCT based video encoder. As a consequence, the computational complexity of video encoding is significantly reduced with very little degradation of video quality. In [13], the Laplacian distribution based model in [12] is extendedly applied to the chrominance component in a similar way as the luminance component. In fact, about the DCT coefficients, various studies [14–22] have been carried out on their distributions for still images. Although it differs in opinion as to what distribution model is the most suitable, the generalized Gaussian distribution remains a popular choice, which includes the Laplacian and Gaussian distributions as two special cases. Traditionally, these studies are concentrated only on fitting the empirical data from some standard pictures, and then comparing their goodness-of-fit. However, for fast video encoding, where the main concern is how to speed up video encoding efficiently, more effective statistical models for predicting ZQDCT coefficients are desired.

In this paper, we propose a novel statistical model based on Gaussian distribution to predict ZQDCT coefficients and optimize the 8×8 DCT based video encoder. Both the theoretical analysis and simulation results demonstrate that the proposed model is superior to other models such as those in [7,8], and the Laplacian distribution based model [12] in terms of the efficiency of ZQDCT coefficients prediction. Furthermore, after a comprehensive study on the dynamic range of DCT coefficients, the proposed Gaussian distribution based statistical model is further refined. As a result, a hybrid model is presented to predict ZQDCT coefficients. Simulation results on several benchmark video sequences demonstrate that the hybrid model can achieve the best encoding efficiency and achieve almost the same rate-distortion performance as the original encoder.

The rest of this paper is organized as follows. In Section 2, the novel Gaussian distribution based statistical model is presented and compared with other predictive models theoretically. The comprehensive analysis of DCT coefficients is discussed in Section 3, and the hybrid model is derived. The simulation results are shown in Section 4 where the performances of the proposed statistical model and hybrid model are studied and compared with other models on several benchmark video sequences. Finally, Section 5 concludes this paper and gives the future research directions.

2. Proposed Gaussian distribution based statistical model

2.1. DCT and Q

We first analyze the sufficient condition for the quantized DCT coefficients to be zeros. In this paper, we mainly focus on the 8×8 DCT which is widely used in MPEG-4 and H.263 standards and will consider the 4×4 integer DCT in H.264 standard in our future research. We define $f(x,y)$, $0 \leqslant x,y \leqslant 7$, as the 8×8 residual pixel block, such that

$$f(x,y) = I(x,y) - I_m(x,y), \qquad 0 \leqslant x,y \leqslant 7 \tag{1}$$

where $I(x,y)$ is the current image block and $I_m(x,y)$ is the best-matched block predicted from the reference frame. The best-matched block is obtained in the motion estimation stage to minimize the sum of absolute difference *SAD* which is given by

$$SAD = \sum_{x=0}^{7} \sum_{y=0}^{7} |f(x,y)| \tag{2}$$

The two dimensional 8×8 DCT coefficients $F(u,v)$, $0 \leqslant u,v \leqslant 7$, are computed by

$$F(u,v) = \frac{C(u)C(v)}{4}\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y) \times \cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2y+1)v\pi}{16}\right) \quad (3)$$

where $C(u)$, $C(v) = 1/\sqrt{2}$, for $u,v = 0$, and $C(u)$, $C(v) = 1$, otherwise. The transformed coefficients $F(u,v)$ are quantized for compression, and will be equal to zero if the following condition holds true

$$F(u,v) < \alpha Q_{\mathrm{p}} \quad (4)$$

where Q_{p} is the quantization parameter which is usually equal to half of the quantization step size and ranges from 1 to 31. The parameter α is related to the quantization method applied. For example, the quantization performed in H.263 and MPEG-4 inter mode follows

$$L(u,v) = \mathrm{sign}(F(u,v)) \times \left\lfloor \frac{|F(u,v)| - \frac{Q_{\mathrm{p}}}{2}}{2Q_{\mathrm{p}}} \right\rfloor \quad (5)$$

where $L(u,v)$ is the quantized DCT coefficient. The DCT coefficients are quantized to zeros if $|L(u,v)| < 1$. Therefore, when $|F(u,v)| < 2.5\ Q_{\mathrm{p}}$, the coefficients $F(u,v)$ will be quantized to zeros. As a result, α should be chosen as $\alpha = 2.5$.

2.2. Prediction of ZQDCT coefficients based on Gaussian distribution

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:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < +\infty \quad (6)$$

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

$$E[|x|] = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}}\sigma \quad (7)$$

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

$$E[|x|] \approx \frac{SAD}{N} \quad (8)$$

where N is the number of coefficients (i.e., 64 for a 8×8 block). Hence, we can get

$$\sigma \approx \sqrt{\frac{\pi}{2}}\frac{SAD}{N} \quad (9)$$

Note that the variance of the (u,v)th DCT coefficient $\sigma_F^2(u,v)$ can be written as [23]:

$$\sigma_F^2(u,v) = \sigma^2 \left[ARA^T\right]_{u,u}\left[ARA^T\right]_{v,v} \quad (10)$$

where $[\cdot]_{u,u}$ is the (u,u)th component of a matrix, the uth row of A is the basis vector $\frac{1}{2}C(u)\cos\left(\frac{(2x+1)u\pi}{16}\right)$, and R is

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^7 \\ \rho & 1 & \rho & \cdots & \rho^6 \\ \rho^2 & \rho & 1 & \cdots & \rho^5 \\ & \vdots & & \ddots & \\ \rho^7 & \cdots & \cdots & \cdots & 1 \end{bmatrix} \quad (11)$$

where ρ is the correlation coefficient. In this work, we set ρ equal to 0.6 which is the same as in [12]. By the central limit theorem, the summation of identically distributed random variables can be well approximated as a Gaussian distribution. And the central limit theorem applies even though the random variables are spatially correlated, as long as the correlation coefficient is less than one. Therefore, 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:

$$\gamma\sigma_F < \alpha Q_{\mathrm{p}} \quad (12)$$

If $\gamma = 3$, then the probability of the DCT coefficient equal to zero after quantization is about 99.73%. Derived from (9), (10) and (12), a criterion for ZQDCT coefficients prediction with high probabilities is

$$SAD < \beta_{\mathrm{G}}(u,v) \times \alpha Q_{\mathrm{p}} \quad (13)$$

where

$$\beta_{\mathrm{G}}(u,v) = \frac{\sqrt{2}N}{\gamma\sqrt{\pi\left[ARA^T\right]_{u,u}\left[ARA^T\right]_{v,v}}} \quad (14)$$

Given $\rho = 0.6$, $N = 64$ and $\gamma = 3$, the components of β_{G} are shown in Table 1.

2.3. Implementation and comparison with other models

Based on the above analysis, we propose the following adaptive scheme to reduce the DCT, Q, IQ and IDCT computations. If $SAD < 5.53 \times \alpha Q_{\mathrm{p}}$, the DCT is not performed and all the coefficients are set to zeros. Else if $5.53 \times \alpha Q_{\mathrm{p}} \leqslant SAD < 7.22 \times \alpha Q_{\mathrm{p}}$, the upper left coefficient (DC coefficient) is computed and all other coefficients (AC coefficients) are set to zeros. Else if $7.22 \times \alpha Q_{\mathrm{p}} \leqslant SAD < 14.34 \times \alpha Q_{\mathrm{p}}$, the 4×4 low-frequency DCT coefficients are computed and all other coefficients are set to zeros. Otherwise, all the 64 DCT coefficients are computed using the traditional DCT computation method. Simulation results have shown that the proposed simple ZQDCT prediction scheme works very well to efficiently reduce the computational complexity of the video encoder with negligible video quality degradation. If further reduction in the computational complexity is required, we can additionally add other schemes to the above-mentioned ZQDCT prediction scheme, for instance, the thresholds $\beta_{\mathrm{G}}(0,2) \times \alpha Q_{\mathrm{p}}$ and $\beta_{\mathrm{G}}(0,6) \times \alpha Q_{\mathrm{p}}$ can be used to calculate the 2×2 and 6×6 low-frequency DCT coefficients, respectively.

Compared with the models proposed in [7,8], which only consider the case of detecting all-zero DCT blocks, our

Table 1
Threshold matrix β_G, $\rho = 0.6$, $N = 64$ and $\gamma = 3$

5.53	7.22	9.26	11.86	14.34	16.46	18.07	19.07
7.22	9.44	12.10	15.50	18.73	21.51	23.61	24.91
9.26	12.10	15.51	19.87	24.02	27.58	30.26	31.93
11.86	15.50	19.87	25.45	30.76	35.32	38.76	40.90
14.34	18.73	24.02	30.76	37.18	42.69	46.85	49.44
16.46	21.51	27.58	35.32	42.69	49.02	53.79	56.76
18.07	23.61	30.26	38.76	46.85	53.79	59.03	62.29
19.07	24.91	31.93	40.90	49.44	56.76	62.29	65.73

proposed threshold $5.53 \times \alpha Q_p$ is larger than the thresholds $4\alpha Q_p$ [7] and $\frac{4\alpha Q_p}{\cos^2(\pi/16)}$ [8], and besides the block skipping scheme, the proposed model also considers other skipping schemes, hence can achieve more reductions in the computational complexity. Considering the Laplacian distribution based model [12], each of the threshold $\beta_G(u,v)$ derived by our Gaussian distribution based model is larger than the corresponding threshold $\beta_L(u,v)$ proposed in [12], and the relationship can be established as

$$\beta_G(u,v) = \frac{2}{\sqrt{\pi}}\beta_L(u,v), \quad 0 \leqslant u,v \leqslant 7 \tag{15}$$

Therefore, the proposed statistical model is able to predict more ZQDCT coefficients than the model in [12], and consequently more efficient to improve the real time performance of video encoding.

3. Proposed hybrid model

In this section, we exploit to study the dynamic range of DCT coefficients analytically aiming to further boost the ZQDCT prediction capacity of the proposed Gaussian distribution based model. As a result, a hybrid model is derived from the combination of comprehensive analysis of DCT coefficients and the previously proposed Gaussian distribution based model.

3.1. Comprehensive analysis of ZQDCT coefficients

From Eqs. (2) and (3), the range of the DCT coefficient $F(u,v)$ is bounded as

$$F(u,v) \leqslant \frac{C(u)C(v)}{4} \times \max_{x,y}\left\{\left|\cos\left(\frac{(2x+1)u\pi}{16}\right)\right|\left|\cos\left(\frac{(2y+1)v\pi}{16}\right)\right|\right\} \times SAD \tag{16}$$

We start our discussion by considering the case of $u = v = 0$, such that

$$F(0,0) \leqslant \tfrac{1}{8}SAD \tag{17}$$

As a result, the DC term $F(0,0)$ will be quantized to zero if

$$F(0,0) \leqslant \frac{SAD}{8} < \alpha Q_p \Rightarrow SAD < 8\alpha Q_p \tag{18}$$

Therefore, $F(0,0)$ can be predicted as zero by comparing the SAD with the threshold T given by

$$T = 8\alpha Q_p \tag{19}$$

Take $u = 4$ and $v = 2$ for another example, we can get

$$F(4,2) \leqslant \frac{1}{4}\max_{x,y}\left\{\left|\cos\frac{(2x+1)\pi}{4}\right|\left|\cos\frac{(2y+1)\pi}{8}\right|\right\} \times SAD \tag{20}$$

Considering

$$\max_{0\leqslant x\leqslant 7}\left|\cos\frac{(2x+1)\pi}{4}\right| = \cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} \tag{21}$$

$$\max_{0\leqslant y\leqslant 7}\left|\cos\frac{(2y+1)\pi}{8}\right| = \cos\left(\frac{\pi}{8}\right) \tag{22}$$

Eq. (20) is further given as

$$F(4,2) \leqslant \frac{\sqrt{2}}{8}\cos\left(\frac{\pi}{8}\right) \times SAD \tag{23}$$

So $F(4,2)$ can be predicted as zero if

$$SAD < \frac{4\sqrt{2}\alpha Q_p}{\cos(\frac{\pi}{8})} \tag{24}$$

Similarly, other DCT coefficients can be bounded depending on the frequency position that affects the maximum values of the two cosine functions. As a result, the thresholds for SAD that determine the quantized DCT coefficients to be zero-valued are listed in Table 2.

3.2. Proposed hybrid model

According to our comprehensive analysis of DCT coefficients, we find that the proposed statistical model can be further improved. In the proposed Gaussian distribution based statistical model, only the DC coefficient $F(0,0)$ is manipulated through the DCT, Q, IQ and IDCT computations if $5.53 \times \alpha Q_p \leqslant SAD < 7.22 \times \alpha Q_p$. However, based on the mathematical analysis of DCT coefficient in the previous subsection, under such a condition, the DC coefficient is by no means needed to compute since it has been predicted as zero. So the computations related to the DC term can be skipped and the thresholds for prediction can be further optimized. Therefore, the proposed statistical model can be further improved when combined with the comprehensive study of DCT coefficients. In practice, we

Table 2
Threshold T_i, $i = 1,\ldots,6$ for prediction of ZQDCT coefficients

Threshold	DCT Coefficients in (u,v)
$T_1 = \frac{4\alpha Q_p}{\cos^2(\frac{\pi}{16})}$	$u = 1,3,5,7,\ v = 1,3,5,7$
$T_2 = \frac{4\alpha Q_p}{\cos(\frac{\pi}{16})\cos(\frac{\pi}{8})}$	$u = 1,3,5,7,\ v = 2,6$ $u = 2,6,\ v = 1,3,5,7$
$T_3 = \frac{4\alpha Q_p}{\cos^2(\frac{\pi}{8})}$	$u = 2,6,\ v = 2,6$
$T_4 = \frac{4\sqrt{2}\alpha Q_p}{\cos(\frac{\pi}{16})}$	$u = 0,4,\ v = 1,3,5,7$ $u = 1,3,5,7,\ v = 0,4$
$T_5 = \frac{4\sqrt{2}\alpha Q_p}{\cos(\frac{\pi}{8})}$	$u = 0,4,\ v = 2,6$ $u = 2,6,\ v = 0,4$
$T_6 = 8\alpha Q_p$	$u = 0,4,\ v = 0,4$

propose a hybrid model on the basis of the proposed statistical model. The motivation for this effort is to further improve the encoding efficiency of our statistical model. To nicely state the hybrid model, Table 3 is given for description. The hybrid model only considers three types of DCT, Q, IQ and IDCT implementations: *Skip*, 4×4 and 8×8.

4. Simulation results

In this work, the XVID codec [24] is implemented for simulation, which is an MPEG-4 compliant video codec. Eight benchmark video sequences are used for testing, each of which is of CIF format (352×288). They are "Foreman", "News", "Silent", "Table Tennis", "Garden", "Bus", "Stefan", 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: 3, 7, 14, 21 and 28, are used in our experiments. For comparison, the models discussed in [7,8,12] are implemented. For notational simplicity, we let (O) indicate the performance of the original MPEG-4 encoder, (G) and (H) indicate the performance of the proposed Gaussian distribution based model and the hybrid model. All the simulations are running on a PC with Intel Pentium 3.0 GHz CPU and 512 Mb of RAM.

4.1. Encoding computational complexity

The computational complexity of video encoding is studied. First, we will compare the computational complexity of DCT, Q, IQ and IDCT of the test models in Figs. 1–8. In these figures, the required computational complexity of DCT, Q, IQ and IDCT of the test predictive model is evaluated as

$$C = \frac{T_d}{T_d^o} \times 100\% \tag{25}$$

where T_d is the encoding time of DCT, Q, IQ and IDCT of the test model, and T_d^o is the encoding time of these four stages in the original encoder. From these figures, it is evident that the proposed statistical model and hybrid model can achieve better performance in reducing the computational complexity of DCT, Q, IQ and IDCT than the other models [7,8,12]. It reveals that the proposed models can effectively eliminate redundant computations which are impossible to detect in [7,8,12]. In general, for different Q_p values and different video sequences, the average computations of DCT, Q, IQ and IDCT have been decreased by about 40 and 45 percent as compared with the original encoder when applying the proposed statistical model and hybrid model, respectively.

For all the predictive models, an additional overhead is introduced to compare the *SAD* with the thresholds for each block to make a decision for DCT, Q, IQ and IDCT implementations. And this overhead has been considered into the encoding time of DCT, Q, IQ and IDCT. So it can be observed that the computational complexity of DCT, Q, IQ and IDCT for some video sequences based on the models of [7] and [8] are more than that of the original encoder when Q_p is small, for example, $Q_p = 3$ and 7 for "Mobile Calendar". However, the computation reduction due to the proposed statistical model and hybrid model outperforms this overhead, so the real time performance improves. We also notice that when Q_p is small, the thresholds are relatively small so that some ZQDCT coefficients might not be recognized. When Q_p increases, more ZQDCT coefficients can be accurately predicted resulting in a greater reduction of the encoding computational complexity.

In addition, the entire encoding time which includes the processing time of all the encoder's procedures is evaluated. Similar to that of the DCT, Q, IQ and IDCT, the computational complexity of entire video encoding is defined and plotted in Figs. 9–16, where we can easily see the advantage of the proposed statistical model and hybrid model in speeding up video encoders. From experiments, the average encoding time have been decreased by about 17 and 20 percent as compared with the original encoder when applying the proposed statistical model and hybrid model, respectively. In conclusion, the proposed hybrid model can

Table 3
Hybrid model for DCT, Q, IQ and IDCT implementation

Type	Condition	DCT, Q, IQ and IDCT
Skip	$SAD < 8\alpha Q_p$	Not performed
4×4	$8\alpha Q_p \leqslant SAD < 14.34\times\alpha Q_p$	Only calculate 4×4 low frequency DCT
8×8	$14.34\times\alpha Q_p \leqslant SAD$	Performed to all the 64 DCT coefficients

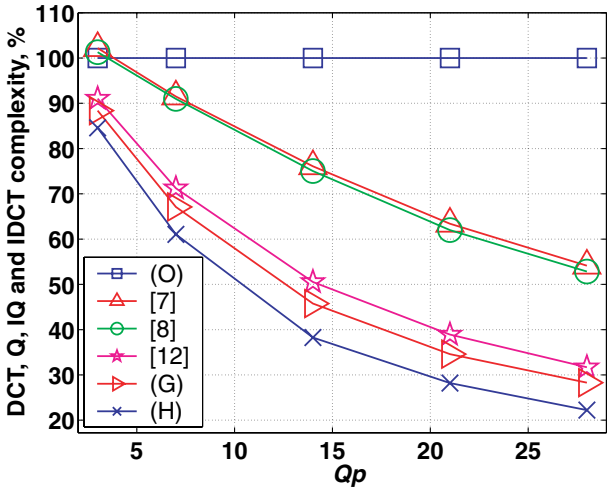

Fig. 1. Computational complexity of DCT, Q, IQ and IDCT, "Foreman" video sequence.

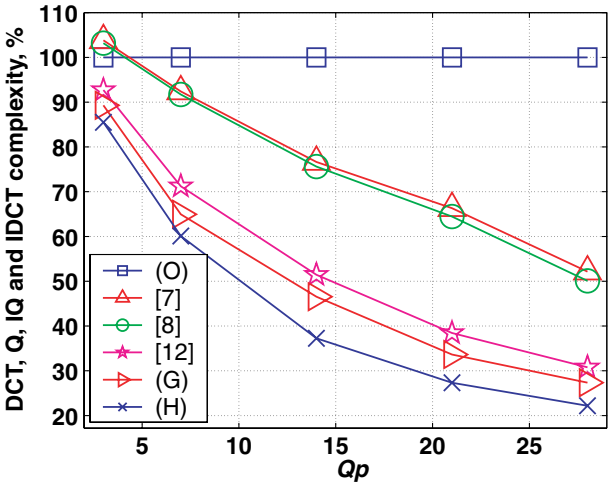

Fig. 4. Computational complexity of DCT, Q, IQ and IDCT, "Table Tennis" video sequence.

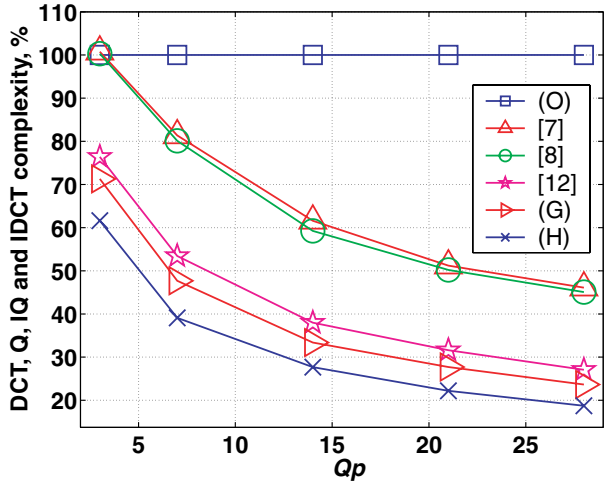

Fig. 2. Computational complexity of DCT, Q, IQ and IDCT, "News" video sequence.

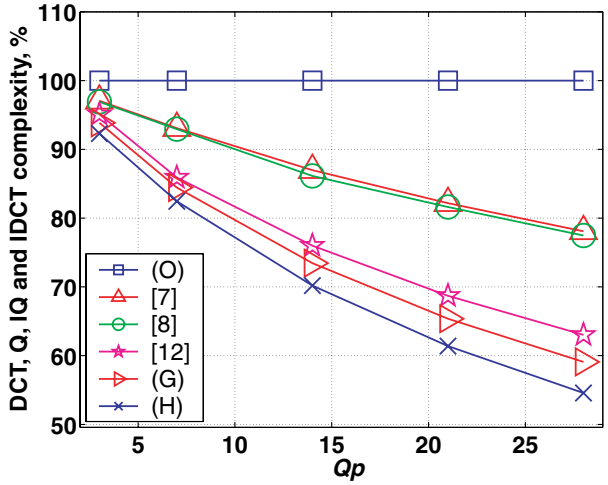

Fig. 5. Computational complexity of DCT, Q, IQ and IDCT, "Garden" video sequence.

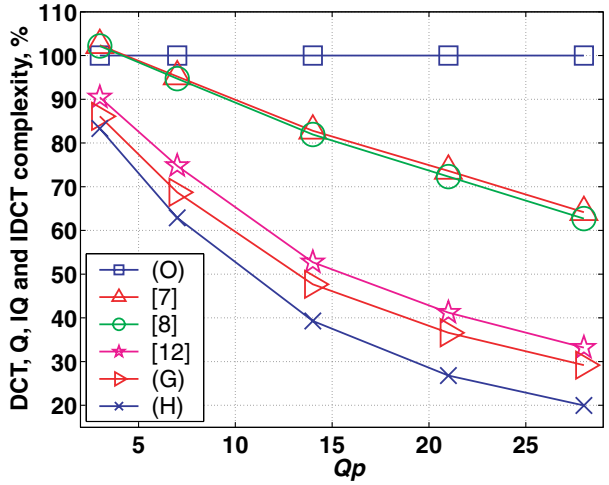

Fig. 3. Computational complexity of DCT, Q, IQ and IDCT, "Silent" video sequence.

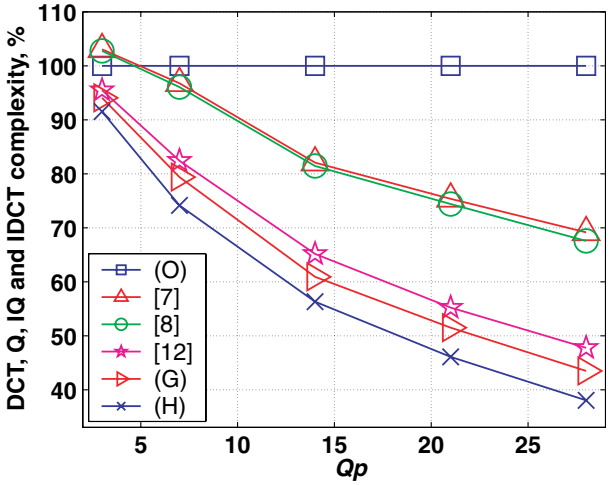

Fig. 6. Computational complexity of DCT, Q, IQ and IDCT, "Bus" video sequence.

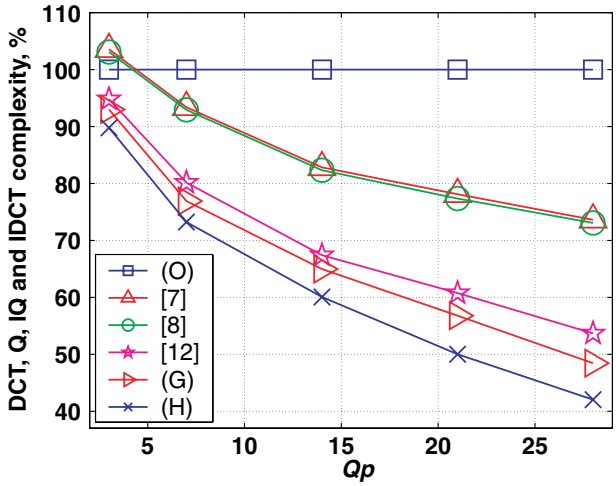

Fig. 7. Computational complexity of DCT, Q, IQ and IDCT, "Stefan" video sequence.

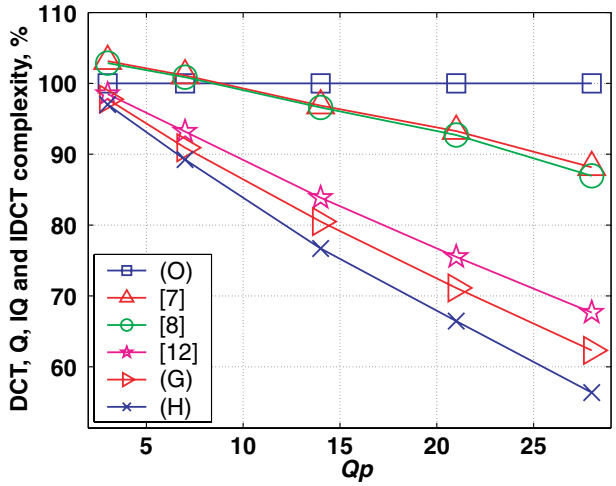

Fig. 8. Computational complexity of DCT, Q, IQ and IDCT, "Mobile Calendar" video sequence.

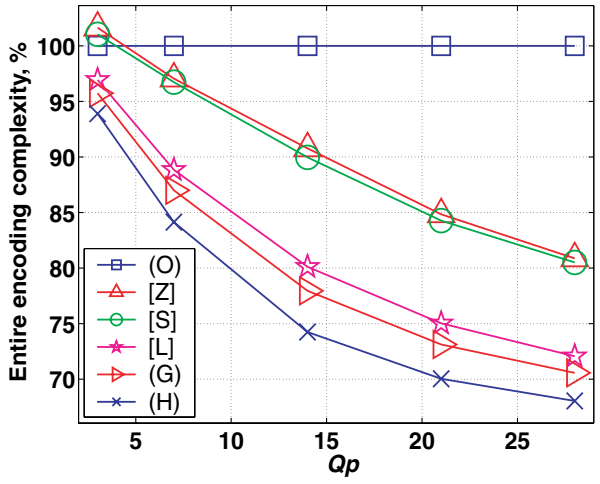

Fig. 9. Computational complexity of the entire video coding, "Foreman" video sequence.

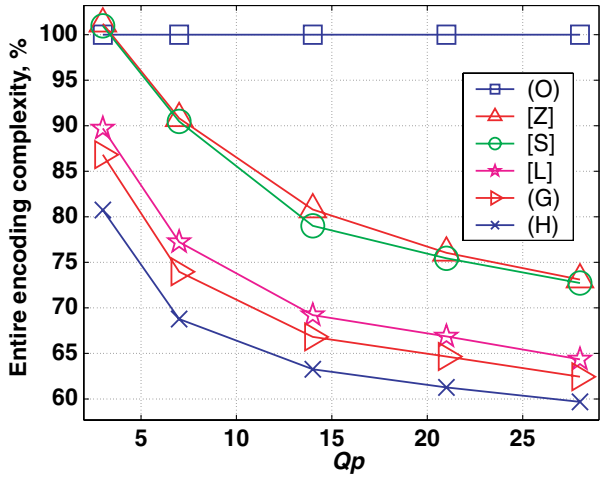

Fig. 10. Computational complexity of the entire video coding, "News" video sequence.

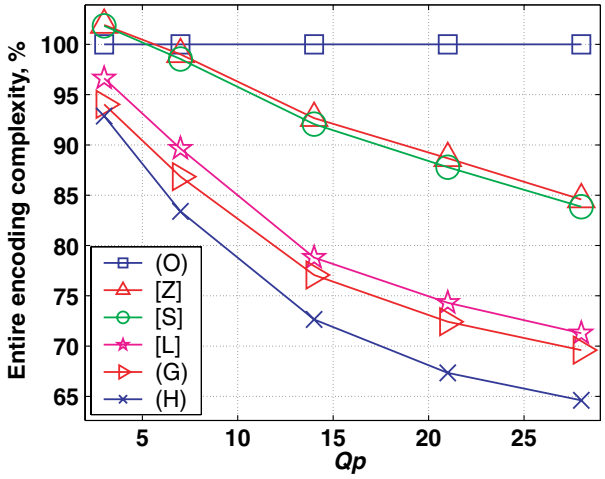

Fig. 11. Computational complexity of the entire video coding, "Silent" video sequence.

achieve the best performance in reducing the computational complexity of video encoding.

4.2. Video quality and bit rates

Now, we will study the encoded video quality and bit rates resulted from the proposed Gaussian distribution based statistical model and hybrid model. 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:

$$\Delta P = P - P_{\text{org}} \tag{26}$$

$$\Delta R = \frac{R - R_{\text{org}}}{R_{\text{org}}} \times 100\% \tag{27}$$

where P and P_{org} are the PSNR criterion of the test model and the original encoder; R and R_{org} are the encoded bit rates of the test model and the original MPEG-4 encoder,

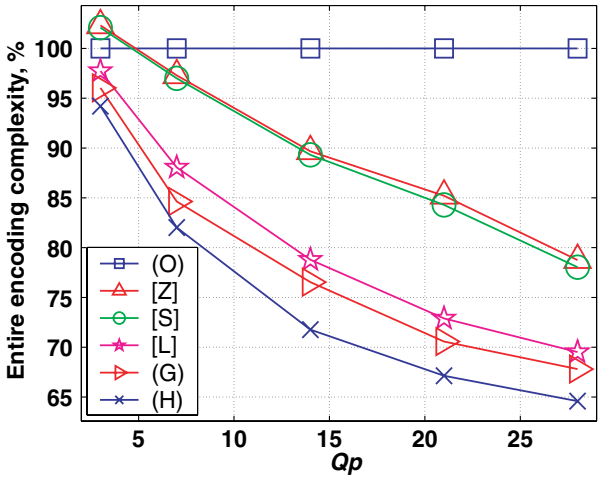

Fig. 12. Computational complexity of the entire video coding, "Table Tennis" video sequence.

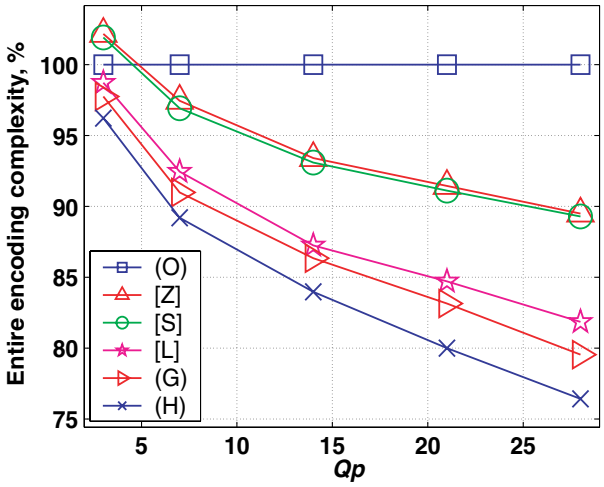

Fig. 15. Computational complexity of the entire video coding, "Stefan" video sequence.

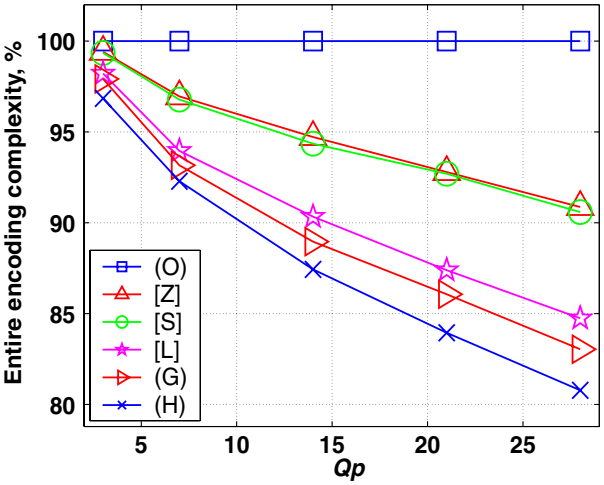

Fig. 13. Computational complexity of the entire video coding, "Garden" video sequence.

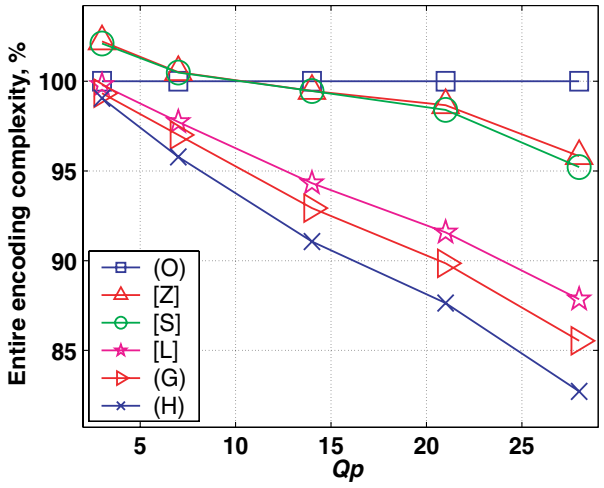

Fig. 16. Computational complexity of the entire video coding, "Mobile Calendar" video sequence.

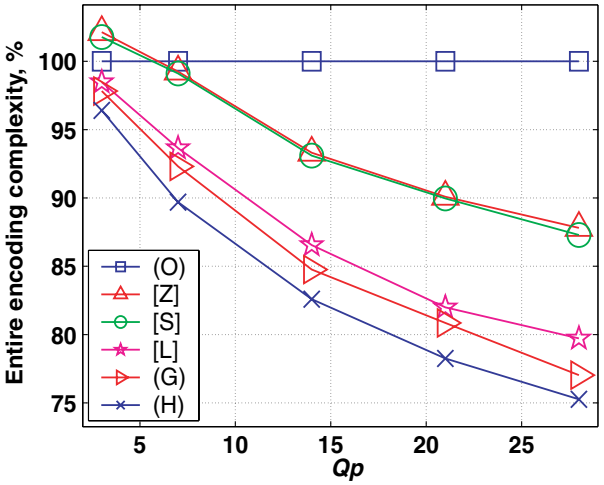

Fig. 14. Computational complexity of the entire video coding, "Bus" video sequence.

respectively. From the experimental results, the models of [7] and [8] do not degrade the video quality since both of these two models provide sufficient conditions to detect all-zero DCT blocks and have the same PSNR and bit rates performances as the original encoder. Therefore, we only present the PSNR and bit rates results for the model of [12], the proposed Gaussian distribution based model and the hybrid model in Tables 4 and 5.

From the results shown in Table 4, the video quality degradation in terms of PSNR drop resulted from the proposed statistical model is compatible to that of [12], and the deterioration of video quality is so insignificant that it does not cause any subjective artifact from human being's point of view. Considering the proposed hybrid model, it incurs almost the same video quality degradation as the proposed statistical model. As far as the bit rates performance is concerned (see Table 5), all of the three models will reduce the bit rates as compared with the original encoder. Both the

Table 4
Comparison of PSNR degradation ΔP (dB)

Q_p		Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
3	[12]	−0.003	−0.005	−0.003	−0.003	−0.003	−0.007	−0.005	−0.002
	(G)	−0.007	−0.015	−0.004	−0.007	−0.007	−0.014	−0.010	−0.006
	(H)	−0.008	−0.016	−0.006	−0.007	−0.008	−0.015	−0.011	−0.006
7	[12]	−0.003	−0.005	−0.001	−0.002	−0.002	−0.006	−0.006	−0.003
	(G)	−0.009	−0.011	−0.003	−0.006	−0.005	−0.013	−0.015	−0.007
	(H)	−0.010	−0.012	−0.003	−0.006	−0.005	−0.014	−0.016	−0.007
14	[12]	−0.003	−0.005	−0.001	−0.002	−0.002	−0.007	−0.007	−0.003
	(G)	−0.011	−0.013	−0.002	−0.006	−0.005	−0.019	−0.016	−0.010
	(H)	−0.012	−0.013	−0.002	−0.006	−0.005	−0.019	−0.017	−0.010
21	[12]	−0.005	−0.006	0	−0.006	−0.003	−0.009	−0.009	−0.004
	(G)	−0.010	−0.014	−0.002	−0.012	−0.007	−0.026	−0.023	−0.012
	(H)	−0.012	−0.014	−0.002	−0.012	−0.007	−0.026	−0.023	−0.012
28	[12]	−0.005	−0.005	−0.001	−0.003	−0.003	−0.017	−0.010	−0.005
	(G)	−0.011	−0.009	−0.003	−0.010	−0.010	−0.037	−0.024	−0.015
	(H)	−0.012	−0.009	−0.004	−0.010	−0.010	−0.038	−0.025	−0.015

Table 5
Comparison of bit rates reduction ΔR (%)

Q_p		Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
3	[12]	−0.04	−0.17	−0.07	−0.07	−0.02	−0.08	−0.05	−0.02
	(G)	−0.16	−0.42	−0.15	−0.19	−0.05	−0.14	−0.09	−0.05
	(H)	−0.18	−0.43	−0.18	−0.20	−0.06	−0.15	−0.10	−0.05
7	[12]	−0.10	−0.11	−0.05	−0.06	−0.03	−0.10	−0.10	−0.04
	(G)	−0.29	−0.29	−0.16	−0.24	−0.08	−0.20	−0.20	−0.09
	(H)	−0.32	−0.30	−0.17	−0.25	−0.08	−0.21	−0.20	−0.10
14	[12]	−0.15	−0.14	−0.11	−0.18	−0.06	−0.16	−0.17	−0.10
	(G)	−0.52	−0.35	−0.24	−0.36	−0.16	−0.47	−0.40	−0.25
	(H)	−0.55	−0.37	−0.23	−0.38	−0.16	−0.48	−0.41	−0.25
21	[12]	−0.25	−0.21	−0.13	−0.24	−0.14	−0.41	−0.31	−0.17
	(G)	−0.65	−0.52	−0.22	−0.42	−0.34	−0.94	−0.65	−0.50
	(H)	−0.69	−0.54	−0.20	−0.43	−0.34	−0.95	−0.66	−0.50
28	[12]	−0.26	−0.21	−0.07	−0.25	−0.23	−0.49	−0.56	−0.28
	(G)	−0.59	−0.44	−0.13	−0.42	−0.62	−1.14	−1.01	−0.91
	(H)	−0.55	−0.45	−0.17	−0.42	−0.62	−1.12	−1.04	−0.92

proposed Gaussian distribution based model and the hybrid model can achieve better bit rates performance than the model of [12]. Therefore, we can draw the conclusion that all of the three models including the model of [12], the Gaussian distribution based model and the hybrid model can achieve almost the same rate-distortion performance as the original encoder.

4.3. False acceptance rate and false rejection rate

In this subsection, the false acceptance rate (FAR) and false rejection rate (FRR) are provided to compare the ZQDCT prediction capacity of the proposed statistical model, hybrid model and other models discussed in [7,8,12]. The comparative results are given in Tables 6–10 for Q_p equal to 3, 7, 14, 21, and 28, respectively, where the FAR and FRR are defined as

$$FAR = \frac{N_{mn}}{N_n} \times 100\% \tag{28}$$

$$FRR = \frac{N_{mz}}{N_z} \times 100\% \tag{29}$$

N_{mn} is the number of non-ZQDCT coefficients being miss classified as ZQDCT coefficients, N_n is the total number of non-ZQDCT coefficients,N_{mz} is the number of ZQDCT coefficients being miss classified as non-ZQDCT coefficients, and N_z is the total number of ZQDCT coefficients. The smaller the FAR is, the less the video quality degrades.

Table 6
Comparison of FAR and FRR, $Q_p = 3$

	Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
FAR								
N_n	4,155,631	1,106,007	1,481,192	2,660,405	9,527,933	3,603,688	7,363,011	11,335,435
[7]	0%	0%	0%	0%	0%	0%	0%	0%
[8]	0%	0%	0%	0%	0%	0%	0%	0%
[12]	0.05%	0.35%	0.12%	0.09%	0.03%	0.09%	0.06%	0.02%
(G)	0.16%	0.86%	0.23%	0.23%	0.06%	0.18%	0.13%	0.06%
(H)	0.20%	0.89%	0.30%	0.24%	0.07%	0.19%	0.13%	0.06%
FRR								
N_z	36,901,649	39,951,273	39,576,088	38,244,811	24,078,211	16,924,952	33,694,269	29,721,845
[7]	96.46%	95.41%	97.66%	98.00%	89.19%	96.68%	98.25%	99.54%
[8]	96.05%	94.99%	97.39%	97.83%	88.67%	96.44%	98.04%	99.49%
[12]	71.66%	48.38%	72.09%	72.90%	74.26%	74.96%	74.45%	91.31%
(G)	66.78%	40.73%	66.25%	67.17%	71.56%	71.26%	70.90%	88.68%
(H)	66.01%	38.61%	65.43%	66.23%	71.18%	70.53%	69.95%	88.46%

Table 7
Comparison of FAR and FRR, $Q_p = 7$

	Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
FAR								
N_n	1,044,922	367,923	374,811	793,326	4,161,194	1,444,062	3,108,486	5,045,851
[7]	0%	0%	0%	0%	0%	0%	0%	0%
[8]	0%	0%	0%	0%	0%	0%	0%	0%
[12]	0.13%	0.36%	0.10%	0.12%	0.04%	0.11%	0.11%	0.05%
(G)	0.35%	0.87%	0.32%	0.36%	0.09%	0.24%	0.27%	0.13%
(H)	0.41%	0.93%	0.37%	0.38%	0.10%	0.25%	0.29%	0.13%
FRR								
N_z	40,012,358	40,689,357	40,682,469	40,111,890	29,444,950	19,084,578	37,948,794	36,011,429
[7]	85.19%	75.16%	88.97%	86.38%	83.41%	89.80%	87.64%	96.86%
[8]	84.59%	73.78%	88.47%	85.63%	83.08%	89.21%	86.90%	96.66%
[12]	48.00%	29.05%	51.71%	48.46%	67.39%	61.69%	59.99%	80.06%
(G)	42.48%	23.37%	43.37%	40.16%	64.12%	57.39%	55.10%	75.56%
(H)	41.17%	21.79%	42.10%	39.17%	63.63%	56.44%	54.43%	75.09%

Table 8
Comparison of FAR and FRR, $Q_p = 14$

	Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
FAR								
N_n	221,531	116,920	96,698	236,897	1,393,875	551,749	1,167,527	1,756,404
[7]	0%	0%	0%	0%	0%	0%	0%	0%
[8]	0%	0%	0%	0%	0%	0%	0%	0%
[12]	0.32%	0.66%	0.32%	0.26%	0.06%	0.23%	0.27%	0.14%
(G)	1.02%	1.74%	0.74%	0.73%	0.17%	0.62%	0.57%	0.36%
(H)	1.08%	1.80%	0.79%	0.76%	0.17%	0.63%	0.58%	0.36%
FRR								
N_z	40,835,749	40,940,360	40,960,582	40,668,319	32,212,269	19,976,891	39,889,753	39,300,876
[7]	70.05%	54.94%	76.79%	70.28%	78.71%	76.65%	76.28%	91.35%
[8]	68.83%	53.66%	75.84%	69.55%	78.27%	75.79%	75.72%	90.98%
[12]	28.33%	17.85%	27.15%	27.36%	57.88%	45.95%	47.33%	65.62%
(G)	23.06%	13.79%	20.70%	22.04%	54.22%	40.53%	43.79%	59.56%
(H)	21.68%	12.77%	18.93%	20.17%	53.57%	39.65%	42.72%	58.77%

The smaller the FRR is, the more efficiently the model detects ZQDCT coefficients. Therefore, it is more desirable to have small FAR and FRR for an efficient predictive model of ZQDCT coefficients.

Table 9
Comparison of FAR and FRR, $Q_p = 21$

	Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
FAR								
N_n	85,693	53,736	42,052	111,656	626,349	289,126	580,594	786,335
[7]	0%	0%	0%	0%	0%	0%	0%	0%
[8]	0%	0%	0%	0%	0%	0%	0%	0%
[12]	0.83%	1.38%	0.50%	0.56%	0.14%	0.63%	0.57%	0.31%
(G)	2.12%	2.96%	1.08%	1.43%	0.43%	1.48%	1.22%	0.80%
(H)	2.30%	3.08%	1.19%	1.49%	0.44%	1.50%	1.24%	0.81%
FRR								
N_z	40,971,587	41,003,544	41,015,228	40,793,560	32,979,795	20,239,514	40,476,686	40,270,945
[7]	56.88%	44.12%	67.04%	59.52%	74.95%	68.55%	71.46%	87.24%
[8]	55.40%	42.96%	65.69%	57.59%	74.42%	67.66%	70.87%	86.65%
[12]	17.93%	12.89%	15.64%	17.44%	50.95%	34.72%	40.70%	54.56%
(G)	13.70%	9.42%	11.71%	13.75%	46.67%	29.78%	35.87%	48.24%
(H)	12.55%	8.52%	9.81%	12.70%	45.87%	28.76%	34.52%	47.26%

Table 10
Comparison of FAR and FRR, $Q_p = 28$

	Foreman	News	Silent	Table Tennis	Garden	Bus	Stefan	Mobile Calendar
FAR								
N_n	41,627	27,384	20,978	59,947	295,589	160,045	308,676	363,612
[7]	0%	0%	0%	0%	0%	0%	0%	0%
[8]	0%	0%	0%	0%	0%	0%	0%	0%
[12]	1.09%	1.70%	0.63%	0.93%	0.30%	1.16%	0.92%	0.58%
(G)	2.77%	3.69%	1.46%	2.15%	0.94%	2.50%	2.01%	1.64%
(H)	2.95%	3.81%	1.71%	2.22%	0.96%	2.52%	2.05%	1.66%
FRR								
N_z	41,015,653	41,029,896	41,036,302	40,845,269	33,310,555	20,368,595	40,748,604	40,693,668
[7]	47.75%	38.89%	57.68%	45.03%	71.46%	62.41%	67.90%	83.07%
[8]	46.36%	37.80%	56.08%	43.01%	70.77%	61.40%	67.12%	82.37%
[12]	12.05%	9.51%	10.04%	13.21%	44.86%	27.07%	33.73%	45.63%
(G)	8.97%	6.78%	7.60%	10.10%	39.86%	22.68%	27.59%	38.39%
(H)	7.86%	5.97%	5.83%	9.08%	38.94%	21.50%	26.52%	37.09%

From the simulation results shown in Tables 6–10, some obvious conclusions can be drawn. Firstly, 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. This can be easily observed by comparing N_z with N_n. Take the video sequence "Foreman" as an example, the percentage for ZQDCT coefficients is 97.45% when $Q_p = 7$, and 99.46% when $Q_p = 14$. Thereby, an efficient predictive model of ZQDCT coefficients is desired to reduce redundant computations for fast video encoding. Secondly, the FRR of the proposed statistical model is smaller than those obtained in [7,8,12]. This indicates the proposed statistical model is more efficient to predict ZQDCT coefficients and explains why our statistical model is able to reduce more redundant computations as shown in the previous Section 4.1. Moreover, the FRR of the proposed hybrid model is smaller than that of the proposed statistical model demonstrating that the ZQDCT prediction capacity of our statistical model can be further boosted when combined with the comprehensive analysis of DCT coefficients. Finally, the FAR of the proposed statistical model is compatible to that of the model in [12], which indicates similar video quality degradations are observed by applying these two models. And since the value N_n is relatively small when compared with N_z, the improvement of prediction efficiency becomes more significant. Considering the video sequence"Silent" in the case of $Q_p = 14$ (see Table 8), when our Gaussian distribution based statistical model is applied, the number of non-ZQDCT coefficients being miss classified as ZQDCT coefficients N^g_{mn} and the number of ZQDCT coefficients being missed classified as non-ZQDCT coefficients N^g_{mz} are 715 and 8,478,025, respectively; if the Laplacian distribution based model [12] is applied, the corresponding N^l_{mn} and N^l_{mz} are 314 and 11,120,855. Therefore, compared with the Laplacian distribution based model [12], the proposed Gaussian distribution based model can correctly detect much more ZQDCT coefficients (e.g., $N^l_{mz} - N^g_{mz} = 2,642,830$) at the negligible expense of false acceptance of ZQDCT coefficients (e.g., $N^g_{mn} - N^l_{mn} = 401$). For other video sequences and Q_p values, the same facts are also observed. In truth,

the proposed Gaussian distribution based model is more efficient than the Laplacian distribution based model [12] to predict ZQDCT coefficients and hence improve the video encoder's performance.

5. Conclusion and future works

In this paper, a novel statistical model based on Gaussian distribution for ZQDCT coefficients prediction is proposed to reduce redundant DCT, Q, IQ and IDCT computations. Derived from the proposed statistical model, an adaptive scheme is provided to perform different types of DCT, Q, IQ and IDCT for computational complexity reduction. In addition, we also perform a comprehensive analysis on the quantized DCT coefficients and propose a hybrid model based on the proposed statistical model to further improve the real-time performance of video encoder. An MPEG-4 compliant video encoder is implemented to evaluate the performance of the proposed statistical model and hybrid model. Both the theoretical analysis and extensive simulation results have demonstrated that the proposed hybrid model outperforms other models for improving the encoder processing speed and can achieve almost the same rate-distortion performance as the original encoder.

In the future, we will apply the Gaussian distribution based model into the H.264 encoder [3] which adopts the integer 4×4 DCT [9]. Thus, four thresholds can be derived to determine five kinds of DCT, Q, IQ and IDCT implementations: *Skip*, 1×1 integer DCT, 2×2 integer DCT, 3×3 integer DCT and 4×4 integer DCT. This adaptive scheme of DCT implementation can be further optimized when considering our previous work [11], which presents an efficient sufficient condition to detect all-zero 4×4 DCT blocks in H.264 encoding.

Acknowledgement

The authors acknowledge the City University of Hong Kong Strategic Grant 7001697 for the financial support.

References

[1] MPEG-4 video verification model version 18.0, ISO/IEC JTC1/SC29/WG11 N3908, Pisa, Italy, January 2001.
[2] Video coding for low bit rate communication, ITU-T Rec. H.263, Febuary 1998.
[3] Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264, March 2005.
[4] H.T. Chen, P.C. Wu, Y.K. Lai, L.G. Chen, A multimedia video conference system: using region base hybrid coding, IEEE Trans. Consum. Electron. 42 (1996) 781–786.
[5] A. Yu, R. Lee, M. Flynn, Early Detection of All-Zero Coefficients in H.263, in: Proceedings of Coding Symposium, 1997 pp. 159–164.
[6] A. Yu, R. Lee, M. Flynn, Performance enhancement of H.263 encoder based on zero coefficient prediction, in: Proceedings of the Fifth ACM International Multimedia Conference, November 1997, pp. 21–29.
[7] X. Zhou, Z. Yu, S. Yu, Method for detecting all-zero DCT coefficients ahead of discrete cosine transformation and quantization, Electron. Lett. 34 (19) (1998) 1839–1840.
[8] L.A. Sousa, General method for eliminating redundant computations in video coding, Electron. Lett. 36 (4) (2000) 306–307.
[9] H.S. Malvar, A. Hallapuro, M. Karczewicz, L. Kerofsky, Low-complexity transform and quantization in H.264/AVC, IEEE Trans. Circuits Syst. Video Technol. 13 (7) (2003) 598–603.
[10] G.Y. Kim, Y.H. Moon, J.H. Kim, An early detection of all-zero DCT blocks in H.264, in: Proceedings of IEEE International Conference on Image Processing, October 2004, pp. 453–456.
[11] H. Wang, S. Kwong, C.W. Kok, Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization, IEEE Trans. Circuits Syst. Video Technol. 16 (4) (2006) 547–552.
[12] I.M. Pao, M.T. Sun, Modeling DCT coefficients for fast video encoding, IEEE Trans. Circuits Syst. Video Technol. 9 (4) (1999) 608–616.
[13] M. Jindal, R. Prasad, K. Ramkishor, Fast video coding at low bit-rates for mobile devices, Int. Conf. Info. Commun. Signal Proc. Pacific Rim Multimedia 3 (2003) 483–487.
[14] R.C. Reininger, J.D. Gibson, Distributions of the Two-Dimensional DCT Coefficients for Images, IEEE Trans. Commun., vol. COM-31, no. 6, Junuary 1983, pp. 835–839.
[15] J.D. Eggerton, M.D. Srinath, Statistical distributions of image DCT coefficients, Comput. Elect. Eng. 12 (3/4) (1986) 137–145.
[16] F. Müller, Distribution shape of two-dimensional DCT coefficients of natural images, Electron. Lett. 29 (22) (1993) 1935–1936.
[17] T. Eude, R. Grisel, H. Cherifi, R. Debrie, On the Distribution of the DCT Coefficients, in: Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 5, April 1994, pp. 365–368.
[18] K.A. Birney, T.R. Fischer, On the modeling of DCT and subband image data for compression, IEEE Trans. Image Process. 4 (2) (1995) 186–193.
[19] S.R. Smoot, Study of DCT coefficients distributions, Proceedings of SPIE (1996) 403–411.
[20] G.S. Yovanof, S. Liu, Statistical analysis of the DCT coefficients and their quantization error, Conf. Rec. 30th Asilomar Conf. Signals, Syst. Comput. 1 (1997) 601–605.
[21] E.Y. Lam, J.W. Goodman, A Mathematical Analysis of the DCT Coefficient Distribution for Images, IEEE Trans. Image Processing 9 (10) (2000) 1661–1666.
[22] E.Y. Lam, Analysis of the DCT Coefficient Distributions for Document Coding, IEEE Signal Processing Lett. 11 (2) (2004) 97–100.
[23] A.K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1989.
[24] XVID Team, [Online] Available: http://www.xvid.org.