Fast Video Coding Based on Gaussian Model of DCT Coefficients

Hanli Wang
Department of Computer Science
City University of Hong Kong
Hong Kong, China
Email: wanghl@cs.cityu.edu.hk

Sam Kwong
Department of Computer Science
City University of Hong Kong
Hong Kong, China
Email: cssamk@cityu.edu.hk

Chi-Wah Kok
Department of Electrical and Electronic Engineering
Hong Kong University of Science and Technology
Hong Kong, China
Email: eekok@ust.hk

Abstract—Discrete cosine transform (DCT), quantization (Q), inverse quantization (IQ), and inverse DCT (IDCT) are the building blocks in video coding 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 this paper to reduce the computational complexity of video encoding. Compared with other widely used models in the literature, the proposed model can achieve the best real time performance. Experimental results demonstrate that the proposed statistical model is superior to others in terms of processing speed at the expense of negligible degradation of video quality.

I. INTRODUCTION

The hybrid DCT motion-compensated approach has been the core of almost all recent digital video coding standards such as MPEG-4 [1], H.263 [2], and H.264 [3], where the discrete cosine transform (DCT), motion estimation (ME) and motion compensation (MC), quantization (Q), inverse quantization (IQ) and inverse DCT (IDCT) are the building blocks. Such an architecture, commonly used today, has a performance-critical feedback loop consisting of DCT, ME, Q, IQ, and IDCT stages. As a result, there is a significant interest and research in reducing these computations. Previously, the efforts to reduce the computations of video encoding are mainly focused on fast motion estimation algorithms. However, as the motion estimation becomes optimized, we also need to optimize other functions to further speed up video encoding.

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 to early detect Zero Quantized DCT (ZQDCT) coefficients, i.e., the DCT coefficients equal to zero after Q, before implementing DCT and Q. Yu et al. [4] propose to compare the sum of absolute difference $SAD$ available from motion estimation and the quantization parameter $Q_p$ with a predetermined threshold $T$. If $SAD < T \times Q_p$, then 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. [5] decrease the threshold value experimentally to detect all zero-DCT blocks. In [6], Zhou et al. perform simple theoretical analyses on the range of DCT coefficients and derive the same threshold value as [5] to skip redundant DCT and Q computations. Zhou’s model [6] is further refined by Sousa [7] and Kim [8] where tighter sufficient conditions are derived to obtain more reductions in the computational complexity without video quality degradation. Wang et al. [9] follow the model of [7] and apply it to the H.264 video standard.

However, the models mentioned above only consider the possibility that all the DCT coefficients within one block are predicted as zeros before quantization. In fact, 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 [10], Pao et al. propose a Laplacian distribution based statistical model for ZQDCT coefficients prediction. Based on this statistical model, an adaptive method with multiple thresholds is developed to reduce the computations of DCT, Q, IQ and IDCT. As a result, the computational complexity of video encoding is significantly reduced with very little degradation of video quality. It is generally believed that the motion compensated video frames are mainly composed of edge information of the original frames and could be well modelled by Laplacian distribution [11]. However, recent advances in motion estimation algorithms have shown that the motion compensated video frames are very close to random noise. Thus, such kind of motion compensated video frames are best to be modelled by generalized Gaussian distribution [12], [13].

In this paper, we extend Pao’s result to the case of Gaussian distributed motion compensated video frames, aiming to further reduce the computational complexity of video encoding without much video quality degradation. Both the theoretical analysis and experimental results demonstrate that the proposed model is superior to other models of [6], [7], [10] in terms of the real time performance of video encoding. Experimental results on several benchmark video sequences demonstrate that the proposed statistical model can achieve the best encoding efficiency with insignificant video quality degradation. The rest of this paper is organized as follows. The ZQDCT coefficients are analyzed in Section II. In Section III, the novel Gaussian distribution based statistical model is theoretically presented and compared with other models. Experimental results are presented in Section IV. Finally, Section V concludes this paper.

II. ZQDCT COEFFICIENTS ANALYSIS

First, we will analyze the sufficient condition for the quantized DCT coefficients to be zeros. To simplify our discussion, and without loss of generality, we focus on the $8 \times 8$ DCT which is widely used in MPEG-4 [1] and H.263 [2] standards. We define $f(x,y)$, $0 \leq x, y \leq 7$, as the $8 \times 8$ motion-compensated pixel block, such that

$$f(x,y) = I(x,y) - I_m(x,y), \quad 0 \leq x, y \leq 7 \quad (1)$$

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

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

The two-dimensional $8 \times 8$ DCT coefficients $F(u,v)$, $0 \leq u, v \leq 7$, are computed by

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

$$f(x,y) = I(x,y) - I_m(x,y), \quad 0 \leq x, y \leq 7 \quad (1)$$

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

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

The two-dimensional $8 \times 8$ DCT coefficients $F(u,v)$, $0 \leq u, v \leq 7$, are computed by

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos \left(\frac{(2x+1)u\pi}{16}\right) \cos \left(\frac{(2y+1)v\pi}{16}\right) \quad (3)$$
where \( C(u), C(v) = 1/\sqrt{2} \), for \( u, v = 0 \), and \( C(u), C(v) = 1 \), otherwise. \( F(u, v) \) are quantized for compression and will be quantized to zero if the following condition holds true

\[
F(u, v) < \alpha Q_p
\]

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 \( \alpha \) is related to the quantization method applied. For example, the quantization performed in H.263 and MPEG-4 inter mode follows

\[
L(u, v) = \left| \frac{F(u, v) - Q_p}{2Q_p} \right|
\]

where \( L(u, v) \) is the quantized DCT coefficient. The DCT coefficients are quantized to zeros if \( L(u, v) < 1 \). As a result, \( \alpha \) should be chosen as \( \alpha = 2.5 \) such that the prediction of ZQDCT coefficients in (4) will not result in video quality degradation.

III. PROPOSED GAUSSIAN DISTRIBUTION BASED MODEL

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

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

The expected value of \( |x| \) can be calculated as

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

Since \( E[|x|] \) can be approximated by

\[
E[|x|] \approx \frac{SAD}{N}
\]

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

\[
\sigma \approx \sqrt{\frac{\pi SAD}{2N}}
\]

Note that the variance of the \((u,v)\)th DCT coefficient \( \sigma_{F}^2(u,v) \) [11] can be written as

\[
\sigma_{F}^2(u,v) = \sigma^2[A_{RAR}A]_{u,v},\sigma^2
\]

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

\[
R = \begin{bmatrix}
1 & \rho & \cdots & \rho^7 \\
\rho & 1 & \cdots & \rho^7 \\
\vdots & \vdots & \ddots & \vdots \\
\rho^7 & \cdots & \cdots & 1 \\
\end{bmatrix}
\]

where \( \rho \) is the correlation coefficient. In this work, we set \( \rho \) equal to 0.6 which is the same as in [10]. By 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 the confidence parameter \( \gamma \) in the following form:

\[
\gamma \sigma_{F} < \alpha Q_p
\]

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 prediction with high probabilities is

\[
SAD < \beta_{Z}(u,v) \times \alpha Q_p
\]

where

\[
\beta_{Z}(u,v) = \frac{\sqrt{2N}}{\gamma \sqrt{\pi A_{RAR}A} |A_{RAR}A|_{u,v}}
\]

Given \( \rho = 0.6, N = 64, \) and \( \gamma = 3 \), \( \beta_{Z} \) are shown in Table I.

Based on the above analysis, we propose the following adaptive scheme to reduce the DCT, IQ, and IDCT computations. If \( SAD < 5.54 \times \alpha Q_p \), the DCT is not performed and all the coefficients are set to zeros. Else if \( 5.54 \times \alpha Q_p \leq SAD < 7.22 \times \alpha Q_p \), only the upper left coefficient (DC coefficient) is computed and all other coefficients (AC coefficients) are set to zeros. Else if \( 7.22 \times \alpha Q_p \leq SAD \leq 14.34 \times \alpha Q_p \), the 4 x 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.

In the current work, we mainly compare the proposed model with models in [6], [7], and [10], because the model in [8] is derived for H.264 codec (not 8 x 8 but 4 x 4 DCT). Compared with [6] and [7], which only consider the case of detecting all-zero DCT blocks, our proposed threshold \( 5.54 \times \alpha Q_p \) is larger than the threshold \( 4\alpha Q_p \) [6], and \( \frac{4\alpha Q_p}{\cos(\pi/16)} \pi/16 )\) [7], 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 [10], each of the threshold \( \beta_{C}(u,v) \) derived by our Gaussian distribution based model is larger than the corresponding threshold \( \beta_{C}(u,v) \) proposed in [10], and the relationship can be established as

\[
\beta_{C}(u,v) = \frac{2}{\sqrt{\pi}} \beta_{Z}(u,v), 0 \leq u, v \leq 7
\]

Therefore, the proposed statistical model is able to predict more ZQDCT coefficients than the model in [10], and consequently more efficient to improve the real-time performance of video encoding.

IV. EXPERIMENTAL RESULTS

In this work, the XVID codec [14] is implemented for experiments, which is an MPEG-4 compliant video codec. \( \alpha \) is equal to 2.5 as stated in Section II. To avoid introducing any biasing factors, the encoder fixes the quantization parameter \( Q_p \) during video encoding. Four benchmark video sequences are used, each of which is of CIF format (352 x 288). All the simulations are running on a PC with Intel Pentium 3.0 GHz CPU and 512 Mbytes of RAM.

A. Video Quality and Encoding Time

Firstly, we will study the trade-off between the computational complexity and video quality for the proposed Gaussian distribution based statistical model. For notational simplicity, we let \( \{O\} \) indicate the performance of the original MPEG-4 encoder and \( \{G\} \) indicate the performance of the proposed statistical model. The video quality is objectively measured by the Peak Signal to Noise Ratio (PSNR, dB). The comparative results about PSNR are shown in Table II, where PSNR performance is presented as the difference between the test model and the original encoder. The positive values mean increments whereas negative values indicate decrements compared with the test model and the original encoder. From Table II, it is observed that the video quality degradation in terms of PSNR drop resulted from the proposed

<table>
<thead>
<tr>
<th>( \alpha Q_p )</th>
<th>( \pi \beta_{Z} )</th>
<th>( \beta_{Z} )</th>
<th>( \rho )</th>
<th>( N )</th>
<th>( \gamma )</th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.54</td>
<td>7.22</td>
<td>9.26</td>
<td>11.86</td>
<td>14.34</td>
<td>16.46</td>
<td>18.07</td>
</tr>
<tr>
<td>7.22</td>
<td>9.44</td>
<td>12.10</td>
<td>15.50</td>
<td>18.73</td>
<td>21.51</td>
<td>23.61</td>
</tr>
<tr>
<td>9.26</td>
<td>12.10</td>
<td>15.51</td>
<td>18.87</td>
<td>22.92</td>
<td>27.88</td>
<td>30.26</td>
</tr>
<tr>
<td>11.86</td>
<td>15.50</td>
<td>19.87</td>
<td>23.85</td>
<td>30.76</td>
<td>35.33</td>
<td>38.76</td>
</tr>
<tr>
<td>14.34</td>
<td>18.73</td>
<td>22.92</td>
<td>27.88</td>
<td>30.76</td>
<td>35.33</td>
<td>38.76</td>
</tr>
<tr>
<td>16.46</td>
<td>21.51</td>
<td>27.88</td>
<td>32.92</td>
<td>40.89</td>
<td>49.02</td>
<td>53.79</td>
</tr>
<tr>
<td>18.07</td>
<td>23.61</td>
<td>30.26</td>
<td>38.76</td>
<td>40.89</td>
<td>49.02</td>
<td>53.79</td>
</tr>
<tr>
<td>19.07</td>
<td>24.91</td>
<td>31.93</td>
<td>40.90</td>
<td>49.44</td>
<td>56.76</td>
<td>62.29</td>
</tr>
</tbody>
</table>

| TABLE I | THRESHOLD MATRIX \( \beta_{Z} \), \( \rho = 0.6, N = 64, \) AND \( \gamma = 3 \) |
TABLE II

COMPARISON OF PSNR

<table>
<thead>
<tr>
<th>Qp</th>
<th>Foreman</th>
<th>Silent</th>
<th>News</th>
<th>Table Tennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(O) 41.219</td>
<td>44.609</td>
<td>43.110</td>
<td>41.334</td>
</tr>
<tr>
<td></td>
<td>[6] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[10] -0.003</td>
<td>-0.002</td>
<td>-0.005</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(G) 0.007</td>
<td>0.004</td>
<td>0.015</td>
<td>0.007</td>
</tr>
<tr>
<td>7</td>
<td>(O) 38.829</td>
<td>38.816</td>
<td>37.886</td>
<td>38.872</td>
</tr>
<tr>
<td></td>
<td>[6] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[10] -0.003</td>
<td>-0.001</td>
<td>-0.005</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(G) 0.009</td>
<td>0.003</td>
<td>0.011</td>
<td>0.006</td>
</tr>
<tr>
<td>14</td>
<td>(O) 32.665</td>
<td>32.558</td>
<td>33.815</td>
<td>32.272</td>
</tr>
<tr>
<td></td>
<td>[6] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[10] -0.013</td>
<td>-0.010</td>
<td>-0.015</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(G) 0.011</td>
<td>0.002</td>
<td>0.013</td>
<td>0.006</td>
</tr>
<tr>
<td>21</td>
<td>(O) 30.645</td>
<td>30.745</td>
<td>31.614</td>
<td>30.374</td>
</tr>
<tr>
<td></td>
<td>[6] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[10] -0.005</td>
<td>-0.006</td>
<td>-0.006</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(G) 0.010</td>
<td>0.002</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>28</td>
<td>(O) 29.301</td>
<td>29.749</td>
<td>30.159</td>
<td>29.058</td>
</tr>
<tr>
<td></td>
<td>[6] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7] 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[10] -0.005</td>
<td>-0.001</td>
<td>-0.005</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(G) 0.011</td>
<td>0.003</td>
<td>0.009</td>
<td>0.010</td>
</tr>
</tbody>
</table>

As far as the encoding time is concerned, we plot the encoding time of all the evaluated models via different Qp values in Figs. 1-4, where we can clearly observe that the proposed statistical model can achieve the best real-time performance compared with other models.

B. Computation Reduction of DCT, Q, IQ and IDCT

To further demonstrate the proposed model can greatly reduce the computational complexity of video encoding, the comparisons of computational complexity about DCT, Q, IQ and IDCT between the test predictive models and the original encoder are illustrated in Figs. 5-8. In these figures, the required computational complexity of DCT, Q, IQ and IDCT for the test model is defined as

\[
C = \frac{T_d}{T_d^O} \times 100\%
\]

where \( T_d \) is the encoding time of DCT, Q, IQ and IDCT for the test model, and \( T_d^O \) is the encoding time of these four stages in the original encoder. From these figures, it is obvious that the proposed model can obtain better performance in reducing the computational complexity of DCT, Q, IQ and IDCT than the other models in [6, 7] and [10]. It reveals that the proposed model can effectively eliminate redundant computations which are impossible to detect in [6, 7] and [10]. In general, for different Qp values and different video sequences, the average computations of DCT, Q, IQ and IDCT have been decreased by about 50 percent as compared with the original encoder when the proposed model is applied.

C. False Acceptance Rate and False Rejection Rate

Finally, the false acceptance rate (FAR) and false rejection rate (FRR) are provided to compare the ZQDCT prediction capacity of the comparative models. The results are shown in Table III, where the FAR and FRR are defined as

\[
FAR = \frac{N_{mn}}{N_e} \times 100\%, \quad FRR = \frac{N_{mz}}{N_e} \times 100\%
\]

where \( N_{mn} \) is the number of non-ZQDCT coefficients being miss classified as ZQDCT coefficients, \( N_e \) is the total number of non-ZQDCT coefficients, \( N_{mz} \) is the number of ZQDCT coefficients being miss classified as non-ZQDCT coefficients, and \( N_e \) is the total number of ZQDCT coefficients. The smaller the FAR is, the less the video quality degrades. The smaller the FRR is, the more efficiently the model can detect ZQDCT coefficients. Therefore, it is more desirable to have small FAR and FRR for an efficient predictive model of ZQDCT coefficients.
From the simulation results shown in Table III, 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 $N_x$. Take Foreman as an example, the percentage of 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 very desired to avoid redundant computations for fast video encoding. Secondly, the FRR of the proposed statistical model is smaller than those of [6], [7], [10]. This indicates our model is more efficient to predict ZQDCT coefficients and explains why our model is more efficient than the Laplacian distribution based model [10] to predict ZQDCT coefficients and hence improve the video encoder’s performance.

V. CONCLUSION

In this paper, a novel statistical model based on Gaussian distribution for ZQDCT coefficients prediction is proposed to avoid 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. A real time MPEG-4 video encoder is implemented to evaluate the performance of the proposed statistical model. Both the theoretical analysis and experimental results have demonstrated that the proposed statistical model outperforms other ZQDCT predictive models in the literature to improve the encoder processing speed at the cost of negligible video quality degradation.

ACKNOWLEDGMENTS

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

REFERENCES