An Improved Detection Method for Zero Quantized Blocks on H.264/AVC

Bo-Jhih Chen and Shen-Chuan Tai
Institute of Computer and Communication Engineering,
Department of Electrical Engineering, National Cheng Kung University,
Tainan City 701, Taiwan (R.O.C.)
Email: cbj@lily.ee.ncku.edu.tw and sctai@mail.ncku.edu.tw

Abstract—An improved detection method for observing the zero quantized block (ZQB) is proposed. The additional computational cost would be reduced due to $4 \times 4$ ZQBs being detected prior to the $4 \times 4$ DCT forward transform and the quantization processes on H.264/AVC video coder, we report a new criterion based on the statistical analysis by considering the energy conservation theorem. Experiments are also carried out to validate the present method. The results indicate that the present method has both a better detection rate with the negligible PSNR degradation and a reasonable error and/or false detection comparing to the prevalent methods. Particularly, computation savings are obtained as well.

Keywords—zero quantized block; DCT; energy conservation; H.264/AVC;

I. INTRODUCTION

H.264/MPEG4 Part 10 advance video coding (H.264/AVC) provides a number of advances in video coding techniques. H.264 can achieve better performance in both the coding efficiency and the visual quality than previous standards such as MPEG-1/2/4 and H.261/263 [1], [2], and [3]. As far as the low bit-rate coding is concerned, i.e., a higher quantization parameter (QP) for encoding, a large amount of $4 \times 4$ residual blocks after transform (DCT) / quantization (Q) will be considered as zero-quantized blocks (ZQBs) with sixteen zero quantized coefficients. Consequently, if ZQBs can be predicted, the computation of DCT/Q is skipped and the computational complexity is reduced as well.

Several efforts focused on an issue of detecting ZQBs [4], [5], and [6]. Sousa [5] proposed the early detection algorithms used for the $8 \times 8$ DCT-based video encoder such as H.263. In [4], an improved detection algorithm has shown the higher detection rate and more computation savings than [5] based on theoretical analyses. Wang et al. [6] have analyzed the dynamic range of DCT coefficients and further have derived the condition for detecting the ZQBs.

Particularly, other approaches used the characteristic of the input residual data of DCT with statistical model to predict ZQBs. Pao et al. [7] addressed that the correlation of the residual pixel values after the motion-compensated prediction could be obtained approximately by use of a Laplacian distribution. Similarly, Wang et al. [8] extended Pao’s results to present a hybrid method applying in H.264. In [9], a ZQB detection method was presented based on the energy conservation theorem and the rate-distortion of Gaussian distribution. Therefore, Xie’s method was superior to the existing methods in terms of the detection accuracy and the complexity of motion estimation.

In this paper, our detection method for ZQB is derived with respect to the analyses of DCT coefficients. Experimental results show that the average detection rate obtained by using our method is superior to other detection algorithms with negligible visual degradations. The remainder of this paper is organized as follows. Both the background of the integer discrete cosine transform of H.264/AVC and the overview of the ZQB early detection methods are reviewed in Section II. The preliminary work for DCT coefficients and the proposed method are presented in Section III. Experimental results are shown in Section IV to compare the proposed method with the comparative methods. Conclusions are given in Section V.

II. BACKGROUND AND OVERVIEW

A. Integer-DCT Transform and Quantization

H.264 adopts the $4 \times 4$ integer DCT transform to avoid the mismatch problem caused by the inverse transform [10] and [11]. Given a $4 \times 4$ residual block $X$, i.e., $X = \{ f(m,n) \mid 0 \leq m, n \leq 3 \}$, the transform coefficients, $F = \{ F(u,v) \mid 0 \leq u, v \leq 3 \}$, can be calculated by the following form:

$$F = AXA^T = (CXC^T) \otimes PF = W \otimes PF$$ (1)

where

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & 1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}, \quad PF = \begin{bmatrix} a^2 & ab & a^2 & ab \\ \frac{ab}{2} & \frac{a^2}{2} & \frac{ab}{2} & \frac{a^2}{2} \\ \frac{ab}{2} & \frac{a^2}{2} & \frac{ab}{2} & \frac{a^2}{2} \\ \frac{ab}{2} & \frac{a^2}{2} & \frac{ab}{2} & \frac{a^2}{2} \end{bmatrix}$$

and

$$a = \frac{a}{2}, \quad b = \sqrt{\frac{a}{2}}, \quad c = \frac{a}{2} \sqrt{\frac{a}{2}}$$

$A$ is the floating transform marix and $C$ is the core transform matrix. $W$ is a forward transform matrix obtained by $CXC^T$. $PF$ is a post-scaling factor. The symbol $\otimes$ represents that each element of $W$ is multiplied by a factor in the same position in $PF$. Given a quantization parameter,
QP, which varies from 0 to 51. The quantization of H.264 is defined as
\[
|F_Q(u, v)| = |(W(u, v) \cdot MF(QP/6, r) + k) >> qbits
\] (2)
where \(r = 2 - w/2 - v/2\), \(qbits = 15 + floor(QP/6)\), and \(sign(F_Q(u, v)) = sign(W(u, v))\). The constant \(k\) is \(2^{qbits}/3\) for the intra-coded block or \(2^{qbits}/6\) for the inter-coded block. The symbol >> denotes the right-shift operator. The quantization matrix \(MF\) can be defined as:
\[
MF(QP/6, r) = \begin{bmatrix}
5243 & 8066 & 13107 \\
4660 & 7490 & 11916 \\
4194 & 6554 & 10082 \\
3647 & 5825 & 9362 \\
3355 & 5243 & 8192 \\
2893 & 4559 & 7283
\end{bmatrix}
\] (3)

From (2), the quantized coefficient \(F_Q(u, v)\) is equal to zero while the following sufficient criterion holds true:
\[
|W(u, v)| < T(r)
\] (4)
where
\[
T(r) = \frac{2^{qbits} - k}{MF(QP/6, r)}
\] (5)

B. Related work

Sousa [5] provided a sufficient condition to find the 8 × 8 zero-quantized blocks. Wang [8] rewrote it for 4 × 4 blocks,
\[
SAD < TH_{Sousa} = T(0)/4
\] (6)

Based on the theoretical analysis, Moon [4] derived a new threshold which could detect more ZQBs by comparing to Sousa’s method:
\[
SAD < TH_{Wang}
\] (7)
where
\[
TH_{Wang} = \min \left\{ \frac{T(0)}{4} + \min \{|hs(0,3), hs(1,2)|, T(1)|, \frac{T(1)}{2} \right\}
\]
and
\[
hs(i, j) = \sum_{y=0}^{3} (|f(i, y)| + |f(j, y)|)
\]

In [6], Wang proposed an early detection algorithm according to the basis of DCT transform
\[
SAD < TH_{Wang}
\] (8)
where the threshold \(TH_{Wang}\) is
\[
TH_{Wang} = \min \left\{ \frac{(T(0)+m_0)}{2}, \frac{T(1)+m_1)}{2} , T(2) \right\}
\] (9)
\(m_0\) and \(m_1\) are defined as
\[
m_0 = \min(S_3 - 2S_1, S_1 - 2S_3, S_4 - 2S_2, S_2 - 2S_4)
\]
\[
m_1 = \min(S_1 + S_2, S_1 + S_4, S_2 + S_3, S_3 + S_4)
\]
and \(S_i\) is calculated by
\[
S_1 = |f(0,0)| + |f(0,3)| + |f(3,0)| + |f(3,3)|
\]
\[
S_2 = |f(0,1)| + |f(0,2)| + |f(3,1)| + |f(3,2)|
\]
\[
S_3 = |f(1,1)| + |f(1,2)| + |f(2,1)| + |f(2,2)|
\]
\[
S_4 = |f(1,0)| + |f(1,3)| + |f(2,0)| + |f(2,3)|
\]

III. PROPOSED METHOD FOR ZQBS DETECTION

A. Analysis of the Input Residual Block

Pao [7] and Wang [8] assumed that the input data \(f\) of DCT may be approximately by the Gaussian and the Laplacian distribution, respectively. Hence, the variance of DCT coefficients \(\sigma_F(u, v)\) can be defined as
\[
\sigma_F^2(u, v) = \sigma^2 [ARA]^T_{u,u} [ARA]^T_{v,v}
\] (10)
where \(A\) is the transform matrix in (1) and \([\cdot]_{u,u}\) denotes the \((u, u)^{th}\) component of the matrix \(ARA^T\); \(R\) is a covariance matrix given as
\[
R = \begin{bmatrix}
\rho & \rho^2 & \rho^3 \\
\rho & 1 & \rho^2 \\
\rho^3 & \rho^2 & \rho & 1
\end{bmatrix}
\] (11)
where \(\rho\) is the correlation coefficient between vertical and horizontal directions from two pixels of a residual block. Empirically, \(\rho = 0.6\) was set as same as [7] and [8]. From (10) and (11), we have
\[
\sigma_F^2(u, v) = \sigma^2 [ARA]^T_{u,u} [ARA]^T_{v,v}
\] (12)

As shown in (12), the variance of the DCT coefficients can be estimated from the variance of the input residual data \(f\). It also shows that the variance of the \((0, 0)^{th}\) (i.e., the DC coefficient) is larger than that of AC coefficients.

In [9], the condition of each DCT coefficients which are less than one, \(|F(u, v)| < 1\), was derived as follows.
\[
F(u, v) < \frac{5}{6}Q_{step}
\] (13)
where \(Q_{step} = 0.625 \times 2^{Q/6}\).

Based on the property of normally distributed data, the probability of the DCT coefficients within \([-3\sigma_F, 3\sigma_F]\) is about 99.73%. Considering the (13), the probability of \(F(u, v)\) equal to zero is over 99% if the following condition is true
\[
3\sigma_F(u, v) < \frac{5}{6}Q_{step}
\] (14)
From (10) and (14), we have
\[
\sigma_F^2 < \frac{(\frac{5}{6}Q_{step})^2}{3^2[ARA]^T_{u,u}[ARA]^T_{v,v}}
\] (15)
Table I

<table>
<thead>
<tr>
<th>Loc(i)</th>
<th>Position (u, v)th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loc(0)</td>
<td>DC</td>
</tr>
<tr>
<td>Loc(1)</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>Loc(2)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td>Loc(3)</td>
<td>(1, 1)</td>
</tr>
</tbody>
</table>

**B. Proposed Method for ZQB Detection**

Determining the largest magnitude of DCT coefficients, $F_{MAX}$, of a $4 \times 4$ block is an important part for early detecting the ZQBs. In order to investigate the distribution of DCT coefficients, we first analyzed six CIF (352 x 288) video sequences, ('Akiyo', 'Coastguard', 'Foreman', 'News', 'Silent', and 'Table Tennis'), with different QPs (18, 24, 30, 36, 40, and 46). As shown in Table I, a 4 x 4 DCT block is divided into the four locations Loc(i) separately and the DC coefficient locates in Loc(0). We define the probability of the largest of DCT coefficients occurred in Loc(i) as follows

$$P(i) = \frac{N_{Loc(i)}}{N_{All}}$$  \hfill (16)

where $N_{All}$ is the total number of the 4 x 4 blocks and the $N_{Loc(i)}$ denotes the number of the $F_{MAX}$ occurred in Loc(i). Figure 1 shows the occurrence probability of Loc(i) with various quantization parameters QPs. It shows that the distribution of Loc(0) has larger variation with various QPs, while Loc(1) is the dominant location with the largest DCT coefficients, over 50%, especially at lower QPs (higher bit-rates). On average, over 70% and 80% of the largest DCT coefficients fall within top-left $2 \times 2$ and $3 \times 3$, respectively.

In addition, the total energy of DCT satisfies the following equality based on the energy conservation property,

$$\sum m \sum n |f(m, n)|^2 = \sum u \sum v |F(u, v)|^2$$  \hfill (17)

where $f(m, n)$ and $F(u, v)$ are the residual pixel values and the DCT coefficients, respectively. The energy of AC coefficients of a $4 \times 4$ DCT block, $E_{ACS}$, can be derived as the variance of the residual data,

$$E_{ACS} = \sum u \sum v |F(u, v)|^2 - |F(0, 0)|^2$$

$$= \sum m \sum n |f(m, n)|^2 - \left| \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m, n) \right|^2$$

$$= N^2 \sigma_{fi}^2$$  \hfill (18)

In (18), it shows that $N^2$ times of the variance of the input residual data $\sigma_{fi}$ is approximately the energy of AC coefficients $E_{ACS}$. From (15) and (18), we have the threshold for judging the ACs’ energy $E_{ACS}$ as follows,

$$E_{ACS} < TH_{ACS}(u, v)$$  \hfill (19)

where

$$TH_{ACS}(u, v) = \frac{N^2 (\frac{\sigma_{fi}}{N})^2}{3^2 [ARA]^2}$$

Therefore, we first use the (13) to check the DC coefficient $F(0, 0)$ and then employ (19) to judge the ACs’ energy. An improved method is derived from [8] and [9] for detecting ZQBs as follows.

$$\left\{ \begin{array}{c}
|F(0, 0)| < \frac{\sigma_{fi}}{N} Q_{step} \\
E_{ACS} < TH_{ACS}(0, 2)
\end{array} \right.$$  \hfill (20)

where the threshold $TH_{ACS}(u, v)$ is a symmetrical matrix, i.e., $TH_{ACS}(u, v) = TH_{ACS}(v, u)$. Based on an above statistical analysis, it has shown that the occurrence probability that the largest DCT coefficients fall within top-left $3 \times 3$ is over 80%. Therefore, only use $TH_{ACS}(0, 2)$ as the threshold for judging the ACs’ energy and further reduce the number of comparisons compared with [8]. Moreover, by comparing the second threshold of judging ACs’ energy, a higher threshold of proposed method is that of [9]. Hence, proposed method have more capability to early detecting more ZQBs.

**IV. Experimental Results**

We use the reference software JM11.0 [12] to implement the proposed method. Six video sequences (‘Akiyo’, ‘Coastguard’, ‘Foreman’, ‘News’, ‘Silent’, and ‘Table Tennis’) with different motion activities are used. They are in CIF format (352 x 288) and have 150 frames to be encoded by IPPP coding structure. The fast motion estimation with search range 32 is enable and the number of reference frame is set to 1. Rate distortion optimization is enabled for mode selection and use six QP values in our experiments to examine the performance at different bit rates. Note that if the residual block is determined as ZQB, DCT/Q/DQ/IDCT are skipped and residual block is assigned to zeros.

The objective performance in terms of the video quality degradation ($\Delta P$) is given as

$$\Delta P = P_{Org.} - P$$  \hfill (21)

where $P_{Org.}$ and $P$ are the peak signal-to-noise ratio (PSNR, dB) obtained from the original encoder and the encoder with each method, respectively. In addition, the detection rate (DR, %) is defined as

$$DR(\%) = \frac{N}{N_2} \times 100\%$$  \hfill (22)

where $N$ is the number of ZQBs being determined by early detection algorithms prior to DCT/Q and $N_2$ is the total number of actual ZQBs. Table II-VII shows the results of the PSNR degradation and detection rate by comparing to the original encoder. Note that a negative $\Delta P$ means the PSNR gain. On average, the better performance in terms of detection rate is obtained by using our method with insignificant PSNR degradation.
As mentioned, the detection rate of the proposed algorithm is higher than that of other algorithms and the detection quality rate is 90% in average. However, the PSNR of the proposed algorithm is lower than the other algorithms. This is due to the threshold proposed by our method is higher than others (i.e., the number of NZQBs being wrong predicted as ZQBs is more than others).

To evaluate the efficiency of detection algorithms, detection quality rate (DQ) are employed to compare the detection capacity of ZQBs. The $DQ$ is defined as

$$DQ(\%) = \frac{(N_z + N_n) - (N_m + N_f)}{N_z + N_n} \times 100\%$$  \hspace{1cm} (23)

where $N_z$ and $N_n$ are the total number of ZQBs and non-ZQBs (NZQB), respectively. $N_m$ is the number of ZQBs being miss predicted as NZQBs and $N_f$ is the number of NZQBs being false determined to ZQBs. Therefore, the higher $DQ$ means that detection algorithms can more efficiently detect the ZQBs. As shown in Figure 2, the results of DQ curves obtained by our method are better than other comparative algorithms. Our method have average $DQ$ of 90% with negligible video quality degradation at different QP.
Table V

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>0.055</td>
<td>43.8</td>
<td>0.046</td>
</tr>
<tr>
<td>24</td>
<td>0.072</td>
<td>78.6</td>
<td>0.064</td>
</tr>
<tr>
<td>30</td>
<td>0.074</td>
<td>87.2</td>
<td>0.068</td>
</tr>
<tr>
<td>36</td>
<td>0.067</td>
<td>91.7</td>
<td>0.058</td>
</tr>
<tr>
<td>40</td>
<td>0.025</td>
<td>94.7</td>
<td>0.047</td>
</tr>
<tr>
<td>46</td>
<td>-0.002</td>
<td>96.8</td>
<td>-0.029</td>
</tr>
</tbody>
</table>

Table VI

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>0.015</td>
<td>26.1</td>
<td>0.015</td>
</tr>
<tr>
<td>24</td>
<td>0.039</td>
<td>56.8</td>
<td>0.036</td>
</tr>
<tr>
<td>30</td>
<td>0.038</td>
<td>66.0</td>
<td>0.046</td>
</tr>
<tr>
<td>36</td>
<td>0.018</td>
<td>85.8</td>
<td>0.021</td>
</tr>
<tr>
<td>40</td>
<td>0.028</td>
<td>93.2</td>
<td>0.021</td>
</tr>
<tr>
<td>46</td>
<td>0.032</td>
<td>97.2</td>
<td>0.051</td>
</tr>
</tbody>
</table>

CSR) is calculated by

$$CSR(\%) = \left(1 - \frac{OP}{OP_{\text{Org.}}} \right) \times 100\% \quad (24)$$

where $OP_{\text{Org.}}$ is the total number of computation required by DCT/Q/DQ/IDCT in original encoder, and $OP$ is the number of operations of early detection approaches. Figure IV shows that the average $CSR$ of each approaches at different QPs. As we can see, our method can save up to 49% of computation in DCT/Q/DQ/IDCT.

V. CONCLUSION

This paper proposes an improved detection method to predict ZQBs by accounting for the experimental distribution of largest DCT coefficients. The results exhibit insignificant PSNR degradation with 0.08dB and a better detection quality of 90%. The computational results also show that the computation of DCT/Q/DQ/IDCT are saved up to 54.6%.
### Table VIII
THE NUMBER OF REQUIRED OPERATIONS PER 4 × 4 BLOCK

<table>
<thead>
<tr>
<th>Operator</th>
<th>Original</th>
<th>Additional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DC/T/Q</td>
<td>DQ/IDCT</td>
</tr>
<tr>
<td>ADD</td>
<td>80</td>
<td>64</td>
</tr>
<tr>
<td>MUL</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>SFT</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>CMP</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

![XIE2007 WANG2007 Proposed](image)

Figure 3. Comparisons of computational saving rate (CSR, %).

### REFERENCES


