Research Article

Hybrid Modeling of Intra-DCT Coefficients for Real-Time Video Encoding

Jin Li, Moncef Gabbouj, and Jarmo Takala

Faculty of Computing and Electrical Engineering, Tampere University of Technology, 33720 Tampere, Finland

Correspondence should be addressed to Jin Li, jin.li@tut.fi

The two-dimensional discrete cosine transform (2-D DCT) and its subsequent quantization are widely used in standard video encoders. However, since most DCT coefficients become zeros after quantization, a number of redundant computations are performed. This paper proposes a hybrid statistical model used to predict the zeroquantized DCT (ZQDCT) coefficients for intratransform and to achieve better real-time performance. First, each pixel block at the input of DCT is decomposed into a series of mean values and a residual block. Subsequently, a statistical model based on Gaussian distribution is used to predict the ZQDCT coefficients of the residual block. Then, a sufficient condition under which each quantized coefficient becomes zero is derived from the mean values. Finally, a hybrid model to speed up the DCT and quantization calculations is proposed. Experimental results show that the proposed model can reduce more redundant computations and achieve better real-time performance than the reference in the literature at the cost of negligible video quality degradation. Experiments also show that the proposed model significantly reduces multiplications for DCT and quantization. This is particularly suitable for processors in portable devices where multiplications consume more power than additions. Computational reduction implies longer battery lifetime and energy economy.

1. INTRODUCTION

The discrete cosine transform (DCT) performs very close to the statistically optimum Karhunen-Loeve transform (KLT) in terms of compression performance [1], so it has been widely used in data compression. Traditionally, the objective in video coding has been high-compression efficiency, which is usually achieved at the expense of increasing computational complexity. As the newest video coding standard, H.264/AVC [2], significantly outperforms the others in terms of coding efficiency. However, the complexity is greatly increased. High-computational complexity limits real-time performance of the encoder as well as its application in digital portable devices such as mobile phones and digital cameras, as they are still suffering from lack of computational power. Thus, there is great significance and interest in reducing the redundant computations for fast encoding.

Many algorithms have been developed for fast DCT calculation. These algorithms can be classified into direct and indirect algorithms. Direct algorithms generally have a regular structure that reduces the implementation complexity. In 1991, Kou and Fjallbrant [3] proposed a direct computation method that slightly reduces the number of multiplications and additions. On the other hand, indirect algorithms exploit the relationship between DCT and other transforms. These algorithms include the calculations of DCT through the Hartley transform [4], the polynomial transform [5], and the Poisson equation [6]. Currently, many fast algorithms for multidimensional DCT computation are also emerging [7–10]. All these algorithms can speed up the calculations of DCT by utilizing more efficient structures. However, they cannot effectively reduce the computations of ZQDCT coefficients.

As the structure for DCT calculation is optimized, more efforts are focused on reducing the redundant computations for ZQDCT coefficients. Most of the effects are on motion-compensated DCT blocks [11–20], for video encoding and significant reductions have been obtained. Docef et al. [11] propose a quantized DCT method that embeds the quantization into the DCT operations. This method can

greatly reduce the required computations for DCT and quantization, but only for uniform quantization. In [12], Pao and Sun propose a Laplacian distribution-based statistical model for ZQDCT coefficients prediction. Based on this model, an adaptive method with multiple thresholds is mathematically derived to reduce the computations of DCT, quantization, inverse quantization, and inverse DCT in a video encoder. As a result, the computations are significantly reduced with negligible video quality degradation. It is believed that motion-compensated video frames are mainly composed of the edge information of the original frames and could be well modeled by the Laplacian distribution. In 2006, Wang et al. [13] further proposed a Gaussian-based model to predict the ZQDCT coefficients for fast video encoding. It has been proven that the Gaussian model has more effective thresholds than the Laplacian model. Significant complexity reduction is achieved when it is applied to XVID codec [14, 15] and H.264 [16–18]. A general method for detecting all-zero blocks prior to DCT and quantization [19] is proposed by Xie et al. in 2007. Using this model, a number of computations for all-zero DCT coefficient blocks are skipped. In addition, much less searching points are required for motion estimation. All the above proposals remarkably reduce the number of DCT calculations and speed up the encoding time. However, they are only applicable to motion-compensated blocks in video coding since these motion-compensated pixels naturally have a zero-mean value and an approximate Gaussian and Laplacian distribution.

Reducing the redundant calculations for intra-DCT and quantization has not been studied actively. Nishida et al. proposed a zero-value prediction for fast DCT calculation [21] in 2003. If consecutive zero elements are produced during the DCT operation, the remaining transform and quantization are skipped. This method can be directly applied to the intra-DCT and quantization. Although it reduces the total computations of DCT by 29% and quantization by 59% when applied to video coding, the video quality is degraded by 1.6 dB on the average.

In our previous work, a sufficient condition-based ZQDCT prediction method [22] is proposed for intra-DCT to speed up the encoding process without video quality degradation. However, the prediction efficiency need to be further improved. We also proposed a Laplacian modeling-based detection method for 3D DCT in [23]. Experimental results show a promising performance in terms of computational reduction, while the video quality is degraded by 0.03–0.8 dB.

In this paper, we extend Pao's [12] and Wang's [13] results to intra-DCT and quantization in video coding, aiming to simplify the encoding complexity and achieve better real-time performance with minimal video quality degradation. The pixel block at the input of DCT is decomposed into some mean values and a residual block. Subsequently, we prove that these residual pixels yield Gaussian distribution, and thus a statistical model with multiple thresholds is introduced. Then, a sufficient condition under which each quantized coefficient becomes zero is theoretically derived from the mean values. Finally, a hybrid model to reduce the computational complexity of DCT and quantization is developed. Although the proposed model is implemented based on the 8×8 DCT, it can be directly applied to other DCT-based image/video coding standards. As a result, high-prediction efficiency and good computational savings are achieved by the proposed model.

The rest of the paper is organized as follows. The analysis and decomposition of DCT are performed in Section 2. In Section 3, a Gaussian distribution-based model is introduced to predict the ZQDCT coefficients for the residual block. Subsequently, a sufficient condition under which each quantized coefficient becomes zero is derived in Section 4. The experimental results are presented in Section 5. Finally, Section 6 concludes the paper.

2. MATHEMATICAL DECOMPOSITION OF DCT

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

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

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

Alternatively, the DCT in (1) can be expressed in matrix form as

$$\mathbf{F} = \mathbf{A} f \mathbf{A}^{\mathbf{T}}, \tag{2}$$

where the uth row of $\mathbf{A}$ is the basis vector $1/2 \, c(u) \cos((2x + 1) u\pi)/16$.

If $\overline{f}_x(y)$ is the mean value of the eight pixels in each row in an 8×8 pixel block and $f_r'(x, y)$ is the residual pixel, we define

$$\overline{f}_x(y) = \frac{1}{8} \sum_{x=0}^{7} f(x, y), \quad f_r'(x, y) = f(x, y) - \overline{f}_x(y). \tag{3}$$

Similarly, if $\overline{f}_y(x)$ is the mean value of the eight pixels in each column in the obtained residual block $f_r'(x, y)$, and $f'(x, y)$ is the residual value, we continue to decompose the pixel block $f_r'(x, y)$ as

$$\overline{f}_y(x) = \frac{1}{8} \sum_{y=0}^{7} f_r'(x, y), \quad f'(x, y) = f_r'(x, y) - \overline{f}_y(x). \tag{4}$$

Therefore, an 8×8 pixel block is decomposed into 16 mean values $\overline{f}_y(x), \overline{f}_x(y)$, and an 8×8 residual block $f'(x, y)$. They satisfy

$$f(x, y) = \overline{f}_y(x) + \overline{f}_x(y) + f'(x, y). \tag{5}$$

TABLE 1: Mathematical decomposition of the original pixel value block as (2).

100	99	102	100	100	103	101	100
104	100	101	103	100	99	102	98
99	103	100	99	103	99	99	102
100	100	102	115	114	111	108	107
113	109	109	113	112	110	111	126
128	125	129	128	129	126	126	129
120	123	125	116	118	122	120	119
122	118	118	123	120	120	123	117

The relationship among the original pixels, the mean pixel values and the residuals are shown in Tables 1 and 2.

$F(u, v)$

$$
= \begin{cases}
\sum_{y=0}^{7} \overline{f}_x(y), & \text{for } u=0,\ v=0, \\[2ex]
\sqrt{2} \sum_{y=0}^{7} \overline{f}_x(y) \cos \dfrac{(2y+1)v\pi}{16} + F'(0, v), & \text{for } u=0,\ v \neq 0, \\[2ex]
\sqrt{2} \sum_{x=0}^{7} \overline{f}_y(x) \cos \dfrac{(2x+1)u\pi}{16} + F'(u, 0), & \text{for } u \neq 0,\ v=0, \\[2ex]
F'(u, v), & \text{otherwise,}
\end{cases}
\tag{6}
$$

where

$$
\begin{aligned}
F'(u, v) = {} & \frac{c(u)c(v)}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} f'(x, y) \\
& \times \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}.
\end{aligned}
\tag{7}
$$

Then, each DCT coefficient can be, respectively, computed by $\overline{f}_y(x)$, $\overline{f}_x(y)$, and the residual pixel $f'(x, y)$ as (6). Equation (9) gives the deduction process.

Actually, $F'(0, v)$ and $F'(u, 0)$ in (6) are zero valued and can be ignored for the DCT calculation. From (6) and (7), 49 out of the 64 DCT coefficients can be directly calculated from the residual block $f'(x, y)$, and 14 DCT coefficients can be computed by the mean values. The DC coefficient is only relevant to the sum of $\overline{f}_x(y)$ in (3). Therefore, if we can efficiently predict the ZQDCT coefficients for the 63 AC coefficients, a lot of computations will be saved. The prediction algorithms for the 49 DCT coefficients and for the 14 DCT coefficients will be, respectively, presented in Sections 3 and 4.

3. GAUSSIAN MODELING OF THE DCT COEFFCIENTS

The experiments show that the distribution of the residual pixel $f'(x, y)$ can be modeled by a Gaussian distribution with a significant peak at zero. To investigate the distribution, we collected the residual pixels from four QCIF sequences (Akiyo, Miss America, Foreman, and Glasgow). The data

suggest that the residual pixels yield a Gaussian distribution. As an example, Figure 1 shows the distribution of the residual pixels of the four sequences and the ideal Gaussian distribution with zero mean. The Laplacian distribution [12] is also shown for reference.

Since the residual pixels $f'(x, y)$ have the same distribution property as the motion-compensated pixels, the Gaussian distribution-based model proposed by Wang et al. [13] for the motion-compensated residuals can be directly applied. In the following, we briefly introduce the proposed method by [13].

The residual pixel $f'(x, y)$ are approximated by a Gaussian distribution with zero mean and variance σ as

$$
p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}.
\tag{8}
$$

$F(u, v)$

$$
\begin{aligned}
= {} & \frac{c(u)c(v)}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} \left[\overline{f}_x(y) + \overline{f}_y(x) + f'(x, y) \right] \\
& \times \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \\
= {} & \frac{c(u)c(v)}{4} \left[\sum_{x=0}^{7} \sum_{y=0}^{7} \overline{f}_x(y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right. \\
& + \sum_{x=0}^{7} \sum_{y=0}^{7} \overline{f}_y(x) \cos \frac{(2x+1)u\pi}{16} \\
& \times \cos \frac{(2y+1)v\pi}{16} + \sum_{x=0}^{7} \sum_{y=0}^{7} f'(x, y) \\
& \left. \times \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right],
\end{aligned}
\tag{9}
$$

Since

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

$$
\text{for } u \neq 0,
$$

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

$$
\text{for } v \neq 0,
\tag{10}
$$

$$
\cos \frac{(2x+1)u\pi}{16} = 1, \text{ for } u = 0,
$$

$$
\cos \frac{(2y+1)v\pi}{16} = 1, \text{ for } v = 0,
$$

Thus, (6) is verified.

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

$$
E[|x|] = \sqrt{\frac{2}{\pi}}\sigma.
\tag{11}
$$

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

$$E[|x|] \approx \frac{SAD}{N}, \tag{12}$$

where N is the number of coefficients (i.e., 64 for 8×8 block) and

$$SAD = \sum_{x=0}^{7} \sum_{y=0}^{7} |f'(x,y)|. \tag{13}$$

Hence, we get

$$\sigma \approx \sqrt{\frac{\pi}{2}} \frac{SAD}{N}. \tag{14}$$

We define

$$\sigma_F^2(u,v) = \sigma^2 [\mathbf{ARA^T}]_{u,u} [\mathbf{ARA^T}]_{v,v}, \tag{15}$$

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

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho^7 \\ \rho & 1 & \cdots & \rho^6 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^7 & \rho^6 & \cdots & 1 \end{bmatrix}, \tag{16}$$

where ρ is the correlation coefficient. In this work, we set $\rho = 0.6$ in accordance with [12, 13].

By the central limit theorem, the DCT coefficient $F'(u,v)$ will be quantized to zero with a probability controlled by the confidence parameter γ as

$$\gamma \sigma_F(u,v) < Q(u,v), \tag{17}$$

where $Q(u,v)$ is the quantization parameter at pixel locations (u,v) in the DCT block.

Therefore, $F'(u,v)$ will be truncated to zero with very high probability if the quantization $Q(u,v) > \gamma \sigma_F(u,v)$, for all $u,v \in \{0,\dots,7\}$. For instance, if $\gamma = 3$ and $Q(u,v) > \gamma \sigma_F(u,v)$, the probability of $F'(u,v)$ to be quantized to zero is 99.73%. Derived from (12), (15), and (17), a criterion for the ZQDCT coefficient $F'(u,v)$ with high probabilitiy is

$$SAD < \beta_G(u,v) \times Q(u,v), \tag{18}$$

where

$$\beta_G(u,v) = \frac{\sqrt{2}N}{\gamma \sqrt{\pi [\mathbf{ARA^T}]_{u,u} [\mathbf{ARA^T}]_{v,v}}}. \tag{19}$$

Given $N = 64$, $\gamma = 3$, the thresholds $\beta_G(u,v)$ for ZQDCT coefficients are shown in Table 3.

Based on the above analysis, the Gaussian distribution-based model with multiple thresholds is proposed to reduce the intra-DCT and quantization computations. If $SAD < \beta_G(u_1,v_1) \times Q(u_1,v_1)$, for all $u_1,v_1 \in \{1,\dots,7\}$, only DCT coefficient $F(u,v)$ at those pixel locations (u,v) at which $\beta_G(u,v) \times Q(u,v) < \beta_G(u_1,v_1) \times Q(u_1,v_1)$ are computed,

Table 2: The decomposed residual block with the mean values as (3).

	0	−2	0	1	1	0	0	1
101	−1	0	1	−2	−2	2	0	−2
101	3	1	0	1	−2	−2	1	−4
100	−1	5	0	−2	2	−1	−1	1
107	−7	−5	−5	7	6	4	1	0
113	0	−2	−4	−1	−2	−3	−2	12
127	1	0	2	0	1	−1	−1	1
120	0	5	5	−5	−3	2	0	−2
120	2	0	−2	2	−1	0	3	−4

Table 3: Threshold matrix $\beta_G(u,v)$.

5.54	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	35.52	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

the others will be predicted as zeros. For instance, if $SAD < 9.44 \times Q(1,1)$, we skip all the DCT calculations; otherwise, if $SAD \geq 65.73 \times Q(7,7)$, all the 49 DCT coefficients of the residual pixels, for all $u,v \in \{1,\dots,7\}$, have to be computed using the traditional method.

In practice, since the quantization for intra-DCT coefficients is usually fixed before video processing, the thresholds only need to be calculated once and can be constructed prior to intra-DCT. In this way, only comparisons are operated for prediciton purpose for each intra-DCT block. In addition, the Gaussian-based statistical model is only used for the calculations of DCT coeffcients at $u \neq 0$, $v \neq 0$, thus only 49 AC coefficients need to be compared or computed.

4. FURTHER COMPUTATIONAL REDUCTION FOR OTHER AC COEFFCIENTS

4.1. Case of DCT coefficients at $u = 0$, $v \neq 0$

First considering $F'(u,v)$ at $u = 0$, $v \neq 0$, $F'(0,v)$ in (7) can be expressed as

$$F'(0,v) = \frac{1}{4\sqrt{2}} \sum_{y=0}^{7} \left[\cos \frac{(2y+1)v\pi}{16} \times \sum_{x=0}^{7} f'(x,y) \right]. \tag{20}$$

Together, with (3) and (4), it is easy to prove that

$$\sum_{x=0}^{7} f'(x,y) = 0. \tag{21}$$

Therefore, the DCT coefficients at $u = 0,\ v \neq 0$ can be easily calculated from (6) and (20) by

$$F(u, v) = \sqrt{2} \sum_{y=0}^{7} \overline{f}_x(y) \cos \frac{(2y+1)v\pi}{16}. \quad (22)$$

Moreover, the DCT coefficient $F(u, v)$ will be quantized to zero if the following condition holds:

$$F(u, v) < Q(u, v). \quad (23)$$

Similar to the decomposition in Section 2, we continue to decompose the mean values $\overline{f}_x(y)$ into a mean value $\overline{\overline{f}_x}$ and eight residual pixel values $f'_x(y)$ as

$$\overline{\overline{f}_x} = \frac{1}{8} \sum_{y=0}^{7} \overline{f}_x(y), \qquad f'_x(y) = \overline{f}_x(y) - \overline{\overline{f}_x}. \quad (24)$$

Then, each DCT coefficient $F(u, v)$ at $u = 0, v \neq 0$ can be computed by the eight residual pixel values $f'_x(y)$ as

$$F(u, v) = \sqrt{2} \sum_{y=0}^{7} f'_x(y) \cos \frac{(2y+1)v\pi}{16}. \quad (25)$$

Since this can be easily proved following (9) and (10) in Section 2, we just skip the deduction process.

In addition, the sum of absolute difference SAD_x of the eight residual pixels is defined as

$$SAD_x = \sum_{y=0}^{7} |f'_x(y)|. \quad (26)$$

From (25) and (26), the DCT coefficient $F(u, v)$ at $u = 0,\ v \neq 0$ is bounded by

$$F(u, v) \leq \sqrt{2} \max \left\{ \left| \cos \frac{(2y+1)v\pi}{16} \right| \right\} \times SAD_x. \quad (27)$$

So $F(u, v)$ can be predicted as zero if

$$SAD_x \leq \frac{Q(u, v)}{\sqrt{2} \max \left\{ \left| \cos \dfrac{(2y+1)v\pi}{16} \right| \right\}}. \quad (28)$$

Therefore, we can predict $F(u, v)$ as zero by comparing SAD_x with the threshold in (28). Each DCT coefficient is bounded relying on the frequency position that affects the maximum value of the cosine function. As a result, the thresholds to determine ZQDCT coefficients are listed in Table 4.

4.2. Case of AC coefficients at $u \neq 0,\ v = 0$

Since $f'(x, y)$ at $u \neq 0,\ v = 0$ has a zero mean value as shown in Table 6, we need not further decompose it as in Case A. Therefore, the DCT coefficient $F(u, v)$, $for\ u \neq 0,\ v = 0$, will be predicted as zero if the following condition holds:

$$SAD_y \leq \frac{Q(u, v)}{\sqrt{2} \max \left\{ \left| \cos \dfrac{(2x+1)u\pi}{16} \right| \right\}}, \quad (29)$$

TABLE 4: Thresholds of ZQDCT coefficients ($u = 0,\ 1 \leq v \leq 7$).

Threshold	DCT coefficient (u, v)
$T_{u1} = \dfrac{2Q(0, v)}{\cos(\pi/16)}$	$v = 1, 3, 5, 7$
$T_{u2} = \dfrac{2Q(0, v)}{\cos(\pi/8)}$	$v = 2\ , 6$
$T_{u3} = 2\sqrt{2}Q(0, v)$	$v = 4$

TABLE 5: Thresholds of ZQDCT coefficients $1 \leq u \leq 7,\ v = 0$).

Threshold	DCT coefficient (u, v)
$T_{v1} = \dfrac{2Q(u, 0)}{\cos(\pi/16)}$	$u = 1, 3, 5, 7$
$T_{v2} = \dfrac{2Q(u, 0)}{\cos(\pi/8)}$	$u = 2, 6$
$T_{v3} = 2\sqrt{2}Q(u, 0)$	$u = 4$

TABLE 6: Further decomposition of the mean values in a column.

	0	−2	0	1	1	0	0	1
0	0	−2	0	1	1	0	0	1

TABLE 7: Further decomposition of the mean values in a row.

	101	101	100	107	113	127	120	120
111	−10	−10	−11	−4	2	16	9	9

where

$$SAD_y = \sum_{y=0}^{7} |f'(0, y)|. \quad (30)$$

The threshold for each DCT coefficient $F(u, v)$ to be quantized to zero is listed in Table 5.

Theoretically, the DCT coefficient $F(u, v)$ can be most likely predicted as zero in the following two cases: the first is when all the eight pixel values are very close to zero; the second is when the pixel values are large, but the variation is small enough. Tables 6 and 7 give examples based on the 16 mean values $\overline{f}_x(x)$ and $\overline{f}_y(y)$ in Table 2. Table 7 shows the eight residual pixel values $f'_x(y)$ and the mean value $\overline{\overline{f}_x}$ after the decomposition. Although these pixels $\overline{f}_x(y)$ are large, the residuals are very small. The residual pixel values $f'_y(x)$ in Table 6 are close to zero, which means that they contain little energy and have a very high probability to be quantized to zero. Therefore, all DCT coefficients will be predicted as zero without taking the discrete cosine transform and quantization.

4.3. Implementation of the proposed statistical model

Given the above analysis, we propose a hybrid statistical model to predict the ZQDCT coefficients for intra-DCT and
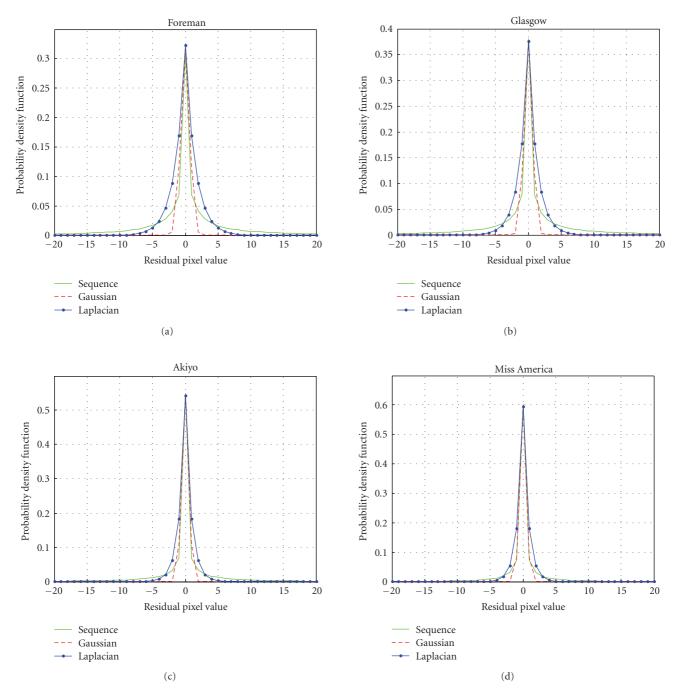
(a)

(b)

(c)

(d)

FIGURE 1: Distribution of the residual pix $f'(x, y)$ of the four sequences: (a) Foreman, (b) Glasgow, (c) Akiyo, and (d) Miss America. The dashed red line shows the ideal Gaussian distribution having a zero mean and a variance approximate to that of the collected data.

quantization calculations by combining the Gaussian-based model in Section 3 and the sufficient condition prediction algorithm in Section 4. Generally, the proposed hybrid model is summarized as follows.

(1) An 8×8 pixel block $f(x, y)$ is decomposed into an 8×8 residual block $f'(x, y)$ and mean values $\overline{f_x}(y)$ and $\overline{f_y}(x)$.

(2) The DC coefficient $F(0, 0)$ is directly computed by $\overline{f_x}(y)$ as (6).

(3) The Gaussian distribution-based model with multiple thresholds is constructed prior to the video processing relying on the quantization $Q(u, v)$ and γ in (18).

(4) During the intratransform, for residual $f'(x, y)$, if $SAD < \beta_G(u_1, v_1) \times Q(u_1, v_1)$, for all $u_1, v_1 \in \{1, \dots, 7\}$, only $F(u, v)$ at pixel locations (u, v) for which $\beta_G(u, v) \times Q(u, v) < \beta_G(u_1, v_1) \times Q(u_1, v_1)$ are computed; the others will be predicted as zeros.
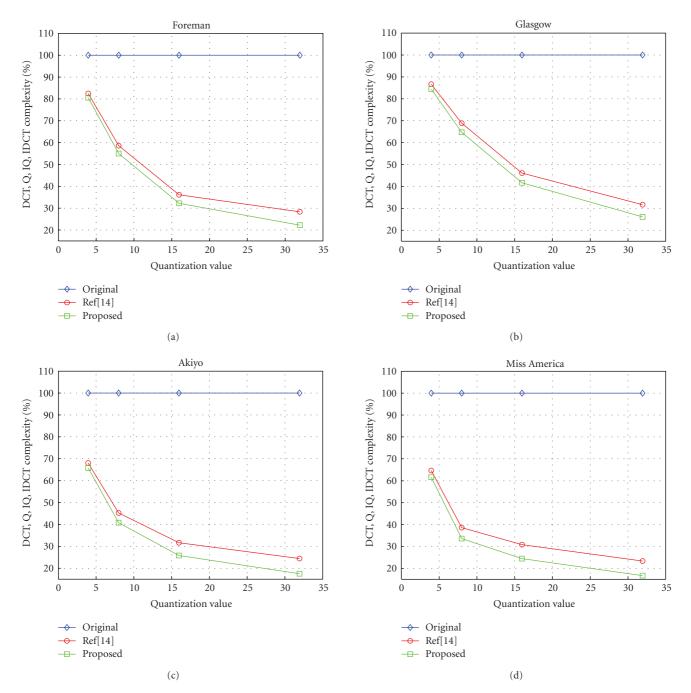
Figure 2: Computational reduction of the proposed model and the reference encoder [14] compared to original XVID codec for different sequences, (a) Foreman, (b) Glasgow, (c) Akiyo, and (d) Miss America.

(5) For the mean values $\overline{f_y}(x)$ and $\overline{f_x}(y)$, if SAD_y (or SAD_x, resp.) is smaller than the first threshold in Table 4 (or Table 5), we directly set $F(u,v)$ as zero. If SAD_y (or SAD_x) is larger than the first threshold and smaller than the second threshold, we only need to compute the coefficients at odd positions, that is, 1, 3, 5, 7. Otherwise, we calculate all the DCT coefficients as usual.

(6) Combine the DC coefficient in step 2, the 49 AC coefficients at $u, v \neq 0$ in step 4, and the 14 AC

coefficients at $u = 0$ or $v = 0$ in step 5 together, the 64 DCT coefficients are calculated.

In order to follow the butterfly row-column transform commonly used in image and video standards, the proposed model is implemented as follows: (1) for the residual block, if the proposed model detects an all-zero block, all the coefficients are directly set to zeros without calculation. Otherwise, if only one coefficient is predicted as a nonzero value, we only need to do eight rowwise transforms and one columnwise transform. If the coefficients in two columns are
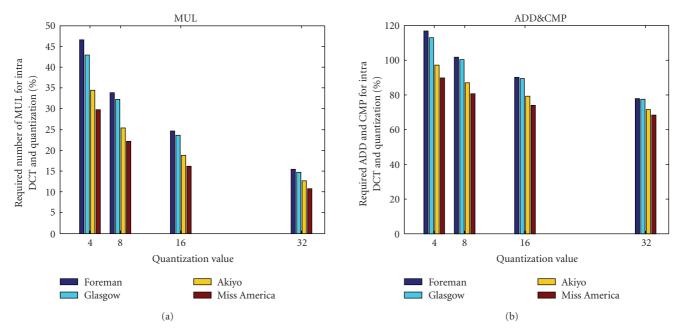
FIGURE 3: Complexity reduction for intra-DCT and quantization including the overhead computations, (a) multiplications (MULs), (b) additions (ADDs), and comparisons (CMPs).

found to be nonzero valued only eight rowwise transforms and two columnwise transforms are required, and so on. (2) For the coefficients computed from the mean values, if *SAD* is smaller than the first threshold in Table 2 or Table 3, all the coefficients are directly set to zeros. If *SAD* is bigger than the first threshold and smaller than the second threshold, only coefficients at odd position, that is, 1, 3, 5, 7 need to be computed, thus the other coefficients at even position, that is, 0, 2, 4, 6 need not to be calculated. Based on the butterfly structure, some operations are saved. Otherwise, do the transform as usual.

In the proposed hybrid model, a total of 63 AC coefficients are compared or calculated, among which 49 are computed by the Gaussian-based model and 14 by the sufficient condition algorithm. Since the Gaussian-based model is derived by approximating the ideal Gaussian distribution with a peak at zero, falsely classifying the non-ZQDCT coefficients into the ZQDCT coefficients is possible, hence resulting in video quality degradation. However, the sufficient condition algorithm for ZQDCT coefficients is mathematically derived from the ideal maximum value of the DCT coefficients, thus it does not cause any information loss compared to the traditional transform method.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed model, a series of experiments was carried out using the XVID codec [24] and compared against the hybrid model proposed in [14] in the literature. In this reference, the hybrid model detects the ZQDCT coefficients for the motion-compensated residual pixels, that is, inter-DCT coefficients.

In the experiments, we apply the proposed model to intra-DCT calculations based on the reference encoder [14] to further reduce the redundant computations for intra-ZQDCT blocks. Four benchmark QCIF video sequences are tested. All the simulations are running on a PC with Intel Pentium 3.2 GHz and 1.5 Gbytes of RAM. Four quantization values Q: 4, 8, 16, 32 are used to assess the performance at different bit rates.

5.1. Computational reduction of DCT and quantization

Firstly, we will study the computational complexity of the proposed hybrid model. The complexities of the DCT and the quantization are illustrated in Figure 2.

In this figure, the computational reduction for the proposed model and the reference encoder is defined as

$$C = \frac{T_d}{T_d^o} \times 100\%, \tag{31}$$

where T_d is the encoding time of the DCT, the quantization (Q), the inverse quantization (IQ), and the inverse discrete cosine transform (IDCT) in the test model and T_d^o denotes the encoding time for DCT, Q, IQ, and IDCT in the original XVID encoder. According to the experiments, significant complexity reduction is obtained by the proposed model. Compared to the reference encoder, since the proposed model is able to not only reduce the redundant computations for intertransform and quantization, but also for intraoperations, the overall DCT, Q, IQ, and IDCT complexity is further simplified by 1.82–5.63%.
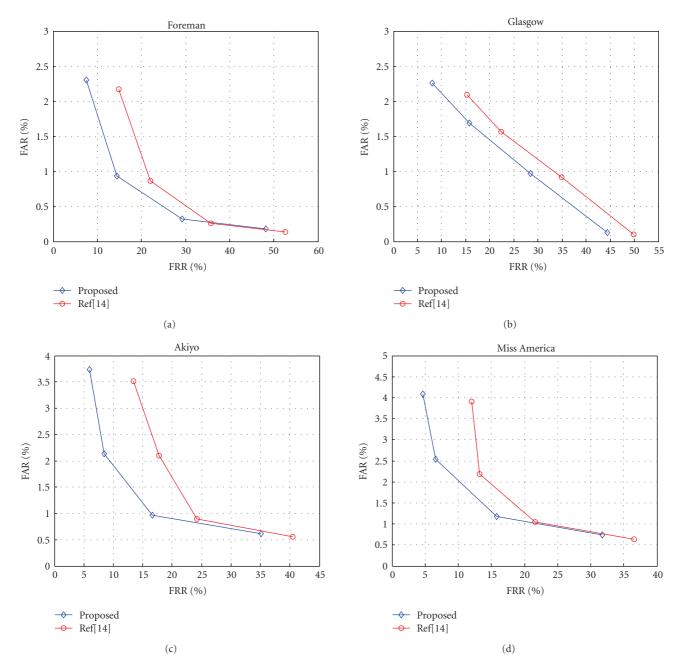
Figure 4: Receiver operating characteristics of the proposed method and the reference codec [14] for (a) Foreman, (b) Glasgow, (c) Akiyo, and (d) Miss America.

Additional operations are performed for the calculation of the residual pixels. The overhead computations (multiplications, additions, and comparisons) have been taken into account in the encoding time in Figure 3. In addition, the overall required number of multiplications (MULs) and additions (ADDs) including the overhead is compared with the original XVID encoder for the calculations of intra-DCT and Q as shown in Figure 3. Since the number of comparisons (CMPs) is very small in the experiments, they are included into the ADD operations. Although the number of ADD operations remains high, the required MUL operations are reduced approximately by 53–90%. Therefore,

the overall processing time for intra-DCT and Q is reduced compared to the reference encoder and the original XVID encoder. In addition, MUL reduction can benefit processors in portable devices such as mobile phones, since MUL consumes more time and power than ADD and CMP due to the implementation structure.

5.2. False acceptance rate and false rejection rate

As two important evaluation parameters, the false acceptance rate (FAR) and the false rejection rate (FRR) are provided to
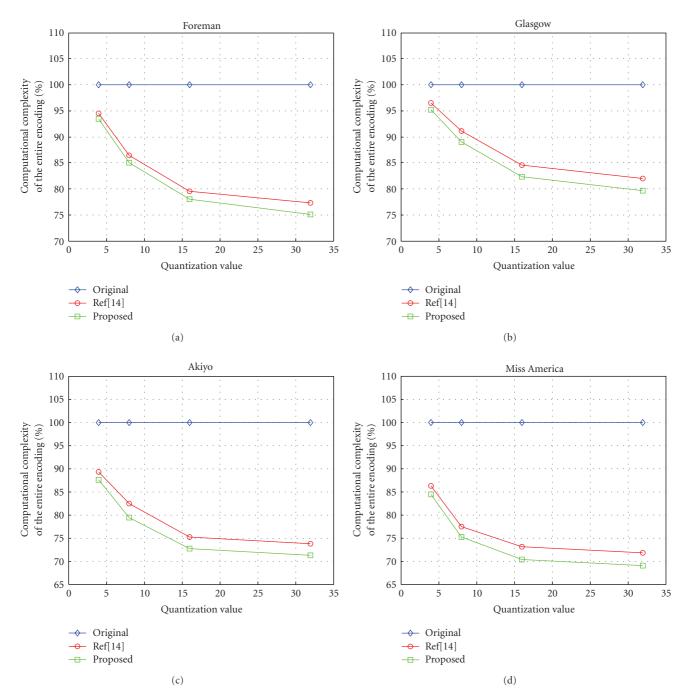
FIGURE 5: Comparison of required encoding time among the proposed hybrid model, the XIVD encoder, and the reference encoder [14] for different sequences: (a) Foreman, (b) Glasgow, (c) Akiyo, and (d) Miss America.

evaluate the proposed hybrid model [15]. The FAR and FRR are defined as

$$\text{FAR} = \frac{N_{mn}}{N_n} \times 100\%, \qquad \text{FRR} = \frac{N_{mz}}{N_z} \times 100\%, \qquad (32)$$

where N_{mn} is the number of non-ZQDCT coefficients being falsely classified as ZQDCT coefficients and N_n is the total number of nonzero-quantized coefficients. While N_{mz} is the number of zero-quantized coefficients being miss classified and N_z is the total number of zero-quantized coefficients.

Normally, the smaller the FAR, the less the video quality degrades and the smaller the FRR, the more efficient is the prediction model. Therefore, it is desirable to have both small FAR and FRR for an efficient prediction model and a low video quality degradation.

Table 8 shows the FRR comparisons between the proposed model and the reference encoder during the calculations for both intra- and inter-DCTs, Q, IQ, and IDCT. Based on the experimental results, some obvious conclusions can be drawn. Firstly, the proposed hybrid model can efficiently

Table 8: Comparison of FRR (%).

	Q	Proposed	Ref [14]
	4	48.23	52.54
Foreman	8	29.24	35.67
	16	14.39	22.11
	32	7.59	14.80
	4	44.47	49.81
Glasgow	8	28.45	34.95
	16	15.78	22.37
	32	8.03	15.30
	4	35.11	40.52
Akiyo	8	16.67	24.25
	16	8.45	17.82
	32	5.96	13.43
	4	31.73	36.54
Miss America	8	15.78	21.62
	16	6.63	13.23
	32	4.69	12.02

Table 9: Comparison of FAR (%).

Image	Q	Proposed	Ref [14]
	4	0.18	0.14
Foreman	8	0.32	0.26
	16	0.93	0.86
	32	2.30	2.17
	4	0.13	0.10
Glasgow	8	0.97	0.92
	16	1.69	1.57
	32	2.26	2.09
	4	0.61	0.55
Akiyo	8	0.97	0.89
	16	2.13	2.10
	32	3.74	3.52
	4	0.74	0.63
Miss America	8	1.17	1.05
	16	2.54	2.18
	32	4.08	3.91

predict the ZQDCT coefficients. Compared to the reference, the proposed model can detect more ZQDCT coefficients and thereby it is more desirable to avoid redundant computations for fast encoding process. The reason is that the proposed model is able to detect the ZQDCT coefficients for both inter- and intra-transforms and quantizations. Secondly, the proposed hybrid model becomes more efficient with increasing quantization Q. Take Glasgow as an example, the FRR is 44.47% when $Q = 4$ and then decreases to 8.03% at $Q = 32$. This means that the proposed model is especially suitable for low bit rates.

The proposed hybrid model has a small FAR ranging from 0.18–4.08% according to the results in Table 9, which indicates that some video quality degradations will occur. Usually, the closer the distribution of the residual pixels is to the ideal Gaussian model, the smaller the FAR will be. Together with Figure 1 and Table 9, Miss America has the closest distribution to the ideal Gaussian, therefore, it has the smallest FAR as can be seen in Table 9. Compared to the reference, the proposed model has a little higher FAR, which indicates that more video quality degradation will occur.

Figure 4 plots the receiver operating characteristics, that is, FAR versus FRR, of the proposed method and the reference. Based on the experimental results, the proposed method has a smaller FAR at the same condition of FRR compared to the reference. Thus, the proposed method is expected to result in lower video quality degradation than the reference at the same prediction efficiency.

5.3. Video quality and encoding time comparison

Finally, we will study the video quality and the encoding time of the proposed model. The objective video quality is measured by the peak signal-to-noise ratio (PSNR). In Table 10, a negative value actually means a PSNR degradation. Experiments show that the falsely classified nonzero

Table 10: Comparison of PSNR (dB) and bit rate (ΔR%).

Image	Q	ΔPSNR (dB)		$\Delta R(\%)$	
		Proposed	Ref [14]	Proposed	Ref [14]
	4	−0.027	−0.015	−0.10	−0.09
Foreman	8	−0.021	−0.017	−0.17	−0.14
	16	−0.016	−0.013	−0.33	−0.27
	32	−0.020	−0.012	−0.50	−0.42
	4	−0.019	−0.011	−0.08	−0.07
Glasgow	8	−0.022	−0.010	−0.21	−0.16
	16	−0.025	−0.013	−0.19	−0.17
	32	−0.022	−0.009	−0.27	−0.22
	4	−0.017	−0.006	−0.16	−0.13
Akiyo	8	−0.019	−0.009	−0.25	−0.19
	16	−0.025	−0.014	−0.47	−0.36
	32	−0.036	−0.019	−0.52	−0.41
	4	−0.013	−0.008	−0.21	−0.17
Miss America	8	−0.018	−0.011	−0.35	−0.28
	16	−0.025	−0.016	−0.34	−0.30
	32	−0.032	−0.027	−0.48	−0.39

coefficients are usually the high-frequency coefficients, thus they do not result in obvious PSNR degradation as shown in Table 10. Based on the results, although the proposed hybrid model has a slightly higher PSNR deterioration than the reference encoder, the degradation is still tolerable. Moreover, along with a nonzero FAR, the skipped calculations for DCT not only reduce the computations but also reduce the bits required to code these coefficients. Therefore, the compression efficiency of the proposed model is even slightly higher than the reference and the original encoder as shown in Table 10.

Figure 5 shows the comparisons of the entire encoding time, where the encoding time reduction ΔT is presented as

$$\Delta T = \frac{T}{T_{\mathrm{org}}} \times 100\%, \tag{33}$$

where T_{org} and T are the entire encoding time of the original XVID encoder and the proposed model, respectively.

From Figure 5, it is shown that the proposed hybrid model achieves the best real-time performance compared to the reference and the original codec. This validates that the proposed model can reduce the computational complexity of the encoder and is superior to the original XVID encoder and the reference encoder. In addition, the proposed method brings about additional benefits at low bit rates, since large quantization result in more DCT coefficients to be treated as ZQDCT coefficients and thus more computational reduction is obtained. Take Akiyo, for example, when it is coded at $Q = 4$, the running time is reduced to 87.58% of the original encoder, while the entire encoding time decreases to 71.26% when Q is increased to 32.

Overall, the proposed hybrid model can significantly reduce the required computations of intra-DCT and quantization and speed up the encoding process. Compared to the reference encoder, the proposed method is able to further reduce the ZQDCT coefficients for intratransform and quantization and thus has better real-time performance for video encoding. Although the video quality degradation is slightly worse than the reference, the deterioration is still negligible. Moreover, the experiments show that the proposed model is more suited for low bit rates.

6. CONCLUSIONS

In this paper, a hybrid model using a Gaussian-based statistical model and a sufficient condition algorithm is proposed to predict the ZQDCT coefficients for intrablocks, aiming to achieve better real-time performance. Experimental results show that compared to the original XVID codec, the proposed model can significantly reduce the redundant computations and speed up the whole encoding process at the expense of negligible video quality degradation. In addition, since the proposed model is implemented based on the reference encoder [14], both inter- and intra-DCTs and quantization are predicted, thus it is able to further reduce the redundant calculations and achieve better real-time encoding. Furthermore, since the proposed method mainly reduces the number of multiplications, it may improve power efficiency in an implementation for low-power processors. Computational reduction also implies longer battery lifetime and energy economy for portable devices.

ACKNOWLEDGMENTS

REFERENCES

[1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 23, no. 1, pp. 90–93, 1974.

[2] H. Kalva, "The H.264 video coding standard," *IEEE Multimedia*, vol. 13, no. 4, pp. 86–90, 2006.

[3] W. Kou and T. Fjallbrant, "A direct computation of DCT coefficients for a signal block takenfrom two adjacent blocks," *IEEE Transactions on Signal Processing*, vol. 39, no. 7, pp. 1692–1695, 1991.

[4] H. Malvar, "Fast computation of discrete cosine transform through fast Hartley transform," *Electronics Letters*, vol. 22, no. 7, pp. 352–353, 1986.

[5] P. Duhamel and C. Guillemot, "Polynomial transform computation of the 2-D DCT," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '90)*, vol. 3, pp. 1515–1518, Albuquerque, NM, USA, April 1990.

[6] K. Yamatani and N. Saito, "Improvement of DCT-based compression algorithms using poisson's equation," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3672–3689, 2006.

[7] Q. Dai, X. Chen, and C. Lin, "Fast algorithms for multidimensional DCT-to-DCT computation between a block and its associated subblocks," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3219–3225, 2005.

[8] A. Elnaggar and H. M. Alnuweiri, "A new multidimensional recursive architecture for computing the discrete cosine transform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 1, pp. 113–119, 2000.

[9] X. Chen, Q. Dai, and C. Li, "A fast algorithm for computing multidimensional DCT on certain small sizes," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 213–220, 2003.

[10] Z. Zhao, H. Chen, X. Du, and A. Sang, "Real-time video compression based-on three-dimensional matrix DCT," in *Proceedings of the 7th International Conference on Signal Processing (ICSP '04)*, vol. 2, pp. 1155–1158, Beijing, China, August 2004.

[11] A. Docef, F. Kossentini, K. Nguuyen-Phi, and I. R. Ismaeil, "The quantized DCT and its application to DCT-based video coding," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 177–187, 2002.

[12] I.-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 608–616, 1999.

[13] H. Wang, S. Kwong, and C.-W. Kok, "Fast video coding based on gaussian model of DCT coefficients," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '06)*, pp. 1703–1706, Island of Kos, Greece, May 2006.

[14] H. Wang, S. Kwong, and C.-W. Kok, "Efficient predictive model of zero quantized DCT coefficients for fast video encoding," *Image and Vision Computing*, vol. 25, no. 6, pp. 922–933, 2007.

[15] H. Wang, S. Kwong, and C.-W. Kok, "Analytical model of zero quantized DCT coefficients for video encoder optimization," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME ')*, pp. 801–804, Toronto, Canada, July 2006.

[16] H. Wang and S. Kwong, "Hybrid model to detect zero quantized DCT coefficients in H.264," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 728–735, 2007.

[17] G. Y. Kim, Y. H. Moon, and J. H. Kim, "Early detection method of all-zero DCT blocks in H.264," in *Proceedings of IEEE International Conference on Image Processing (ICIP '04)*, vol. 1, pp. 453–456, Singapore, October 2004.

[18] H. Wang, S. Kwong, and C.-W. Kok, "Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 4, pp. 547–552, 2006.

[19] Z. Xie, Y. Liu, J. Liu, and T. Yang, "A general method for detecting all-zero blocks prior to DCT and quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 2, pp. 237–241, 2007.

[20] 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 Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 1053–1057, 2005.

[21] Y. Nishida, K. Inoue, and V. G. Moshnyaga, "A zero-value prediction technique for fast DCT computation," in *Proceedings of IEEE Workshop on Signal Processing Systems (SIPS '03)*, pp. 165–170, Seoul, Korea, August 2003.

[22] J. Li, J. Takala, M. Gabbouf, and H. Chen, "A detection algorithm for zero-quantized DCT coefficients in JPEG," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 1189–1192, Las Vegas, Nev, USA, March 2008.

[23] J. Li, M. Gabbouj, J. Takala, and H. Chen, "Laplacian modeling of DCT coefficients for real-time encoding," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 797–800, Hannover, Germany, June 2008.

[24] http://www.xvid.org/.