LETTER

Fast Intra Mode Decision Using DCT Coefficient Distribution in H.264/AVC

Sung-Wook HONG¹, Nonmember and Yung-Lyul LEE(²), Member

SUMMARY The rate-distortion optimization (RDO) method in the H.264/AVC encoder is an informative technology that improves the coding efficiency, but increases the computational complexity. In this letter, a fast Intra mode decision algorithm using DCT (Discrete Cosine Transform) coefficients distribution is proposed to reduce the H.264 encoder complexity. The proposed method reduces the encoder complexity on average 63.44%, while the coding efficiency is slightly decreased compared with the H.264/AVC encoder.

key words: fast mode decision, DCT, intra prediction, H.264/AVC

1. Introduction

The latest video coding standard H.264/AVC, which is well-known to provide high coding efficiency, was developed by the joint work of ITU-T VCEG and ISO/IEC MPEG [1]. Its high coding efficiency is made possible by the new advanced coding tools such as 4 × 4 integer transform, variable block size motion estimation (ME), quarter-pixel accuracy motion compensation (MC), CABAC/CAVLC, spatial prediction for Intra coding, and so on. Usually, Inter coding is superior to Intra coding, but Intra coding is useful for various purposes such as random access, video editing, and scene extraction.

H.264/AVC supports several Intra MB (Macroblock, 16 × 16 block) predictions, i.e. Intra 16 × 16 Luma, Intra 8 × 8 Luma, Intra 4 × 4 Luma, and Intra 8 × 8 Chroma. In the case of Intra 8 × 8 and Intra 4 × 4 Luma predictions, those have the nine directional prediction modes for predicting the 8 × 8 and 4 × 4 blocks, respectively. These prediction methods are similar, but the sizes of the prediction blocks are different. Figure 1(a) shows the boundary pixels and the current block for Intra 4 × 4 prediction. The small letter pixels (a ~ p) are the current 4 × 4 block and the capital letter pixels (A ~ M) are the prediction pixels which are already decoded. Eight modes of the nine directional prediction modes are shown in Fig. 1(b), where the DC prediction (mode 2) that uses the average value of the eight boundary pixels (A ~ D and I ~ L) as the predictor is not shown. As an example of the horizontal prediction for Intra 4 × 4 which is marked as mode 1 in Fig. 1(b), the prediction of every sample is performed from the left block boundary samples I, J, K and L. The vertical prediction is performed in a similar way to the horizontal prediction in the vertical direction from the upper block boundary samples. Other remaining predictions are performed from three similarly chosen block boundary samples by considering the prediction direction in a similar way.

Compared with Intra 4 × 4 Luma prediction, Intra 16 × 16 Luma prediction uses four directional modes for predicting each 16 × 16 block: DC mode, vertical mode, horizontal mode, planar mode. Intra 8 × 8 Chroma prediction uses four directional modes: DC mode, horizontal mode, vertical mode, diagonal mode.

The rate-distortion optimization (RDO) is applied to select the best mode from Intra 4 × 4 prediction and Intra 16 × 16 prediction in the Baseline Profile. For Intra 4 × 4 prediction in the H.264/AVC encoder, the mode decision for each 4 × 4 block is performed by minimizing the following cost:

\[ J(s, c, IMODE|QP, \lambda_{MODE}) = D(s, c, IMODE|QP) + \lambda_{MODE} \cdot R(s, c, IMODE|QP) \]

\[ IMODE \in \{ \text{all nine Intra } 4 \times 4 \text{ modes} \} \]

where QP is the quantization parameter, \( \lambda_{MODE} \) is the Lagrangian multiplier for mode decision, and IMODE indicates one of the nine Intra 4 × 4 prediction modes. D is the distortion between the original 4 × 4 luminance block s, and its reconstruction c, which is computed by SSD (Sum of Squared Difference). R(s,c,IMODE|QP) represents the number of entropy coded bits associated with choosing IMODE.

The computational complexity of RDO needs huge amounts of computations in the encoder so that several fast algorithms have been proposed to reduce the computational complexity [3], [4]. In the paper [3], it presents an Intra prediction based on edge detection histogram and local edge information in the pixel domain. A fast mode decision
algorithm, called the motion activity-based mode decision (MAMD), is also proposed to speed up the encoding process in the pixel domain [4]. However, the proposed method performs the fast Intra mode decision in DCT domain.

In this letter, a fast intra mode decision making use of DCT coefficients distribution in DCT domain is proposed. In Sect. 2, the relationship between DCT coefficients and pixel data will be analyzed to predict the Intra block modes. The fast mode decision algorithm using the distribution of DCT coefficients is introduced in Sect. 3. The experimental results are shown in Sect. 4. Finally, Sect. 5 concludes this letter.

2. Analysis of DCT Coefficient Distribution

H.264/AVC applies a 4×4 DCT-based integer transform and quantization to a 4×4 residual block. The 4×4 integer DCT transform is computed as follows:

$$Y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & -1 & -2 & 1 \\ -1 & 1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \otimes \begin{bmatrix} a^2 & ab & ab \\ ab & \frac{b^2}{2} & \frac{b^2}{2} \\ ab & \frac{b^2}{2} & \frac{b^2}{2} \\ ab & \frac{b^2}{2} & \frac{b^2}{2} \end{bmatrix}$$ (2)

where $X$ denotes a 4×4 matrix representing the residual 4×4 pixel block, the symbol $\otimes$ indicates element by element multiplication, $a = 1/2$, and $b = \sqrt{2}/5$. After calculation of Eq. (2), the quantized coefficients divided by $Q_{\text{step}}$, which have the floating point values, are obtained as follows:

$$Z_{ij} = Y_{ij}/Q_{\text{step}}$$ (3)

$Y_{ij}$ denotes $(i,j)$-th element of matrix $Y$ in Eq. (2). $Q_{\text{step}}$ denotes the quantization step size.

In the proposed method, $X$ in Eq. (2) is not the 4×4 residual block, but the original 4×4 pixel block in order to estimate the block mode of the 4×4 original pixel block.

When only the quantized DCT coefficient $Z_{ij}$ in position of Fig. 2(a) has a nonzero value, the 16 pixels of the 4×4 decoded block have the same value in the spatial domain; therefore, a block having only a DC component can be predicted from the DC, horizontal and vertical prediction modes in Fig. 1(a). When only the coefficients in the top row of the 4×4 inverse quantized block have nonzero values as shown in Fig. 2(b), the four pixels in each column have the same value in the spatial domain. This block can predicted from the vertical and DC prediction modes. When only the coefficients in the far left column have nonzero values as shown in Fig. 2(c), the four pixels in each row have the same value in the spatial domain. This block can predicted from the horizontal and DC prediction modes. When the coefficients are symmetric to ±45° in Fig. 2(d), this block can predicted from the mode 3 and mode 4 diagonal predictions, and DC prediction.

However, the 4×4 original block does not show the definite horizontal and vertical directions, DC components, and diagonal direction even in homogeneous region due to the delicate pixel value changes in real image. Therefore, the information of DCT coefficient distribution is utilized in grouping the nine directional prediction modes for the proposed method.

3. Fast Mode Decision Using the DCT Coefficient Distribution

Based on the DCT coefficient distribution, the proposed fast mode decision avoids the several RDO process on the unnecessary directional modes by investigating the quantized DCT coefficients.

H.264/AVC supports several Intra MB predictions, i.e. Intra 4×4 and Intra 16×16 predictions in Luma, Intra 8×8 prediction in Chroma in the Baseline Profile.

3.1 Intra 4×4

The coefficient $Z_{ij}$ of Eq. (3) is modified into $\hat{Z}_{ij}$ using the threshold $\omega$ as follows:

$$\begin{cases} \hat{Z}_{ij} = Z_{ij}, & \text{if } |Z_{ij}| > |\omega| \\ \hat{Z}_{ij} = 0, & \text{otherwise} \end{cases}$$ (4)

where the threshold $\omega$ is set to 0.5.

**Group $G_0$:**

If only $\hat{Z}_{ij}$ in the DC component as shown in Fig. 2(a) has nonzero values, the original 4×4 block is considered as the vertical mode (mode 0), horizontal mode (mode 1) and DC mode (mode 2).

**Group $G_1$:**

If only $\hat{Z}_{ij}$ in the top row as shown in Fig. 2(b) has nonzero values, the original 4×4 block is considered as the vertical mode and DC mode.

**Group $G_2$:**

If only $\hat{Z}_{ij}$ in the far left column as shown in Fig. 2(c) has nonzero values, the original 4×4 block is considered as the horizontal mode and DC mode.

**Group $G_3$:**

If $\hat{Z}_{ij}$ in the ±45° diagonal directions as shown in Fig. 2(d) has the same value, the original 4×4 block is considered as mode 3, mode 4 and DC mode.
Table 1  Group of fast Intra prediction.

<table>
<thead>
<tr>
<th>Prediction Modes</th>
<th>Group</th>
<th>Candidate Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra4×4</td>
<td>$G_0$</td>
<td>0 (Vertical), 1 (Horizontal), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_1$</td>
<td>0 (Vertical), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_2$</td>
<td>1 (Horizontal), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_3$</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>$G_4$</td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8</td>
</tr>
<tr>
<td>Intra16×16</td>
<td>$G_0$</td>
<td>0 (Vertical), 1 (Horizontal), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_1$</td>
<td>0 (Vertical), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_2$</td>
<td>1 (Horizontal), 2 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_3$</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>Chroma8×8</td>
<td>$G_0$</td>
<td>0 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_1$</td>
<td>2 (Vertical), 0 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_2$</td>
<td>1 (Horizontal), 0 (DC)</td>
</tr>
<tr>
<td></td>
<td>$G_3$</td>
<td>0, 1, 2, 3</td>
</tr>
</tbody>
</table>

Group $G_3$:
If $\hat{Z}_{ij}$ coefficients distributions are not included in $G_0$, $G_1$, $G_2$, and $G_3$, the RDO processes on all nine directional prediction modes are performed.

### 3.2 Intra 16 × 16

The original 16 × 16 samples are sub-sampled to a 4 × 4 samples by extracting one sample from each 4 × 4 samples, and Eqs. (2) and (3) are applied to the sub-sampled 4 × 4 block. The grouping of $G_0$, $G_1$, $G_2$, and $G_3$ for Intra 16 × 16 prediction is shown in Table 1. The grouping process of Intra 16 × 16 is the same as that of Intra 4 × 4 except that the diagonal direction is not taken into account for grouping since Intra 16 × 16 prediction uses the planar prediction mode instead of the diagonal prediction mode in Intra 4 × 4 prediction. The threshold $\omega$ is set to 0.5.

### 3.3 Chroma 8 × 8

The original 8 × 8 samples are sub-sampled to a 4 × 4 samples by extracting one sample from each 2 × 2 samples, and Eqs. (2) and (3) are applied to the sub-sampled 4 × 4 block. The grouping of $G_0$, $G_1$, $G_2$, and $G_3$ for Chroma 8 × 8 prediction is shown in Table 1. The grouping process of Intra 8 × 8 in Chroma is the same as that of Intra 16 × 16 in Luma. The threshold $\omega$ is set to 1.0.

### 4. Experimental Results

The proposed fast intra mode decision is implemented in JM11.0 [5]. The experiments are performed on a Pentium IV 2.4 GHz computer with 2 GB RAM. The proposed method is applied to all Intra frames in the Baseline Profile of H.264/AVC when the QP values are 28, 32, 36 and 40. Several test sequences, each of which has 150 frames, are used under the experimental conditions shown in Table 2.

In Table 3, BD-PSNR and BD-Rate mean the PSNR gain and average bit-rate increase that are calculated based on the BD-PSNR and BD-Rate in VCEG-M33 [6], respectively. Total encoding time saving is defined as follows:

\[
\text{Total encoding time saving(\%)} = \frac{T_{\text{method}} - T_{JM}}{T_{JM}} \times 100 \tag{5}
\]

where $T_{JM}$ denotes the total encoding time processed by the JM encoder and $T_{\text{method}}$ the total encoding time processed by the proposed method or Pan’s method, respectively.

The proposed method and Pan’s one are compared with JM11.0 as JM vs. Pro (Proposed) and JM vs. Pan, respectively, in terms of BD-PSNR, BD-Rate, and total encoding time saving. The proposed method always shows slightly better BD-PSNR and BD-Rate than Pan’s one on average.

Comparing Pan’s algorithm with JM (JM vs. Pan), the former reduces the average 53.30% encoding time, but results in the average BD-PSNR loss of 0.36 and average BD-Rate increase of 7.18, as shown in Table 4. Comparing the proposed method with JM (JM vs. Pro), the proposed method results in the average BD-PSNR loss of 0.27 and average BD-Rate increase of 5.08, which are less loss than those of Pan’s algorithm. The proposed method also reduces the average 63.44% encoding time.

From the RD (Rate-Distortion) results shown in Fig. 3, it is noted that the proposed algorithm has better performances in terms of the coding efficiency as well as encoding complexity than Pan’s algorithm.

We analyzed the selected percentage of each group in
Table 4  Selected percent (%) of each group in Intra4x4 prediction.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Test Sequences</th>
<th>G0</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIF</td>
<td>Foreman</td>
<td>58.51</td>
<td>0.63</td>
<td>0.93</td>
<td>0.73</td>
<td>39.20</td>
</tr>
<tr>
<td></td>
<td>Mobile</td>
<td>28.97</td>
<td>0.88</td>
<td>1.23</td>
<td>0.74</td>
<td>68.18</td>
</tr>
<tr>
<td></td>
<td>Paris</td>
<td>41.79</td>
<td>1.38</td>
<td>1.55</td>
<td>0.37</td>
<td>54.91</td>
</tr>
<tr>
<td>VGA</td>
<td>Ballroom</td>
<td>53.38</td>
<td>1.25</td>
<td>2.07</td>
<td>0.42</td>
<td>42.88</td>
</tr>
<tr>
<td></td>
<td>Exit</td>
<td>75.26</td>
<td>1.55</td>
<td>0.99</td>
<td>0.28</td>
<td>21.93</td>
</tr>
<tr>
<td></td>
<td>Flamenco</td>
<td>71.88</td>
<td>0.90</td>
<td>1.05</td>
<td>0.26</td>
<td>25.91</td>
</tr>
<tr>
<td>HD</td>
<td>Bigships</td>
<td>59.13</td>
<td>0.57</td>
<td>2.37</td>
<td>0.52</td>
<td>37.41</td>
</tr>
<tr>
<td></td>
<td>Shuttlestart</td>
<td>89.22</td>
<td>0.15</td>
<td>0.67</td>
<td>0.18</td>
<td>9.78</td>
</tr>
<tr>
<td></td>
<td>Crew</td>
<td>80.38</td>
<td>0.88</td>
<td>0.74</td>
<td>0.42</td>
<td>17.58</td>
</tr>
</tbody>
</table>

![Fig 3 RD-curve for Shuttlestart with/without the proposed fast algorithm and with Pan’s algorithms in JM 11.0.]

Intra 4 × 4 prediction in Table 4. It shows that the most selected groups fall into groups G0 and G4. It is coincident with the original prediction result falling into mode 0, mode 1, mode 2, and other modes.

5. Conclusion

In this paper, we present a fast intra prediction algorithm based on the DCT coefficients distribution to reduce the encoding complexity. The proposed algorithm selects a few prediction modes among all intra prediction modes from the distribution of DCT coefficients, and only the selected modes are used for the RDO process. The experimental results show that the proposed algorithm achieves an encoding time saving of 63.44% with negligible loss of PSNR and average bite-rate increase.

Acknowledgments

This work was supported by the IT R&D program of MKE/KEIT [2009-F-031-01, Development of Advanced Video Encoder with Low Complexity for Mobile Video Service].

References