Improved Hybrid Coding Scheme for Intra 4x4 Residual Block Produced by H.264/AVC

Li-Li Wang and Wan-Chi Siu
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

Abstract—In this paper, the intra residual macroblock produced by H.264 is investigated. Based on its characteristics, an efficient two-layer coding scheme for the intra residual macroblock is developed. The rate-distortion performance for the proposed coding scheme is evaluated. Experimental results show that our algorithm can achieve better coding performance.

I. INTRODUCTION

Intra prediction technique plays an important role in modern block-based video coding. In 1997, spatial intra prediction was first reported by ITU-T and VCEG [1]. Subsequently, other proposals [2-4] involved in intra prediction techniques have been made. In the H.264, intra prediction is applied to both luminance component and chrominance component. Luminance component includes three block sizes corresponding to three intra prediction types: I4MB, I8MB and I16MB types in the high profile. Chrominance component only has one block size of intra prediction. For smooth and simple areas, intra prediction of H.264 indeed achieves better performance. However, with further development of video compression technology, researchers have found that the current intra prediction scheme is not enough to decorrelate the spatial blocks having complex structures.

In order to compensate for this lack of intra prediction in the current H.264, the usual idea is to improve the current intra prediction strategy [5-8] by adding extra intra prediction modes or adding more reconstructed pixels as reference at the present stage. For example, in [6], in order to improve the coding efficiency of intra prediction in H.264, the authors proposed a bidirection intra prediction (BIP) method, this method includes two schemes. One is to change the sub-block coding order in macroblock. The other is to add extra intra prediction modes through combining two existing intra prediction modes to form one new mode. In [7], template matching technique is used to search for matching blocks within the previously reconstructed pixels of the current frame. In [8], distant pixels instead of only the nearest ones are allowed as reference of intra prediction, which can improve the coding efficiency when the nearest reference pixels are contaminated with noise or occlusion.

In this paper, we propose a new idea to improve the coding efficiency of intra prediction in H.264. We firstly investigate the residue produced by the intra prediction of the current H.264, we found the values in it are still larger, so an efficiently pre-processing stage is introduced before the Integer Cosine Transform (ICT) is performed. In this paper, we mainly focus on the improvement of the coding scheme for the residual block produced by I4MB type. However, our method also can be applied to 8x8 and 16x16 luma components or 8x8 chroma component. Experimental results show that the scheme can achieve better rate-distortion performance.

The rest of the paper is organized as follows: In section 2, we investigate intra prediction residue produced by 4x4 intra prediction of the H.264. Then an efficient coding scheme for the residue is proposed in section 3. Experimental results are shown in section 4. Finally, we conclude the paper in section 5.

II. INVESTIGATION FOR THE INTRA RESIDUAL FRAME IN H.264/AVC

In H.264, intra prediction is performed by extrapolation of the neighboring reconstructed pixels[4,11]. Compared with previous intra coding techniques [1-3], spatial correlation based intra prediction of H.264 has made great compression efficiency for intra macroblock. However, we note that much redundancy still exists in the intra residual frame after intra prediction of H.264 is performed, which means that the residual frame still includes much information entropy. The residual frame is not like natural image, since the values in the residual frame range from -255 to 255. If DCT is directly performed on this kind of residual frame, the energy is, however, not compacted very well, and more high frequency contents appear. After the quantization is performed on these coefficients, more unexpected phenomena will appear by truncating the high frequency coefficients, such as ringing effect, and so on.

The inaccurate intra prediction makes the residual values larger. The possible reasons for increasing the intra prediction residual values could be one of the following cases or their combinations.

• Case 1. There are only a limited number of intra prediction modes.
• Case 2. When 4x4 blocks locate at the boundary of the frame, not all the prediction modes are available for this kind of blocks.
• Case 3. The fixed block segmentation results in that one 4x4 block may include more than one object.
• Case 4. Pixels which are far from the reconstructed reference block are usually difficult to predict.

To resolve these problems, a new method to improve the coding efficiency of the residual block produced by the intra
prediction in the current video coding standard H.264/AVC is addressed in this paper.

III. THE PROPOSED ENCODING SCHEME FOR INTRA PREDICTION RESIDUAL

At present, the scheme of intra prediction used in H.264 limits the efficiency of video coding. Note that some spatial redundancies remain in the residual frame. Making use of the redundancies, we propose a two-layer coding scheme for the residual block. A related coding scheme has been verified efficient for the coding of arbitrary shape video objects based on MPEG-4 in our early work[9].

A. The proposed coding scheme of the residual block produced by 14MB intra prediction in H.264

In our algorithm, we separate a 4x4 residual block produced by an intra 14MB type block into two layers, modulus layer (ML) and basic layer (BL), which can be denoted by (1).

\[ R(j, i) = ML(j, i) + BL(j, i) \quad j, i = 0, 1, ..., 3 \]  

where \((j, i)\) is the position of sample \(R(j, i)\) in the current 4x4 residual block; \(ML(j, i)\) and \(BL(j, i)\) represent the sample values in the modulus layer and basic layer, which can be calculated by (2) and (3), respectively.

\[ ML(j, i) = \langle R(j, i) \rangle_{Th} \]  
\[ BL(j, i) = R(j, i) - ML(j, i) \]

where \(<n>_{th}\) denotes the residue of the number \(n\) modulo \(Th\). Due to this operation, we can see that \(BL(j, i)\) is a multiple of \(Th\). Therefore, we just need to code \(RBL(j, i)\) which is the quotient of dividing \(BL(j, i)\) by the threshold, as in (4).

\[ RBL(j, i) = BL(j, i) / Th \]

In lossy coding part of H.264/AVC, the residual block must be transformed and quantized before entropy coding is performed. H.264 uses a scalar quantizer and a total of 52 values of Qstep are supported by the standard and these are indexed by a Quantization Parameter, QP[4]. Note that Qstep increases with the increment of QP, which means that the precision of transformed residual data further reduces when QP is larger, or vice versa. Based on this observation, the threshold \(Th\) should be dependent on Qstep. Table I shows the thresholds Th for some commonly used QPs, which have been determined by our experimental work.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>QP</td>
<td>22, 27, 32, 37</td>
</tr>
<tr>
<td>Qstep</td>
<td>8, 14, 26, 44</td>
</tr>
<tr>
<td>Th</td>
<td>16, 16, 26, 44</td>
</tr>
</tbody>
</table>

Table I. Thresholds for Different QPs

Fig.1 shows an example about the separation process of one 4x4 residual block, which is from an intra prediction block of the foreman sequence when Th is 44 (QP=37). After deriving ML and RBL, we can see that the values in ML range between -43 and 43 when QP is equal to 37. We also plot the intra residual values (IntraResidual), ML (Modulus), RBL (Quotient) and inter residual values (InterResidual) produced by H.264 video coding in fig.2. We can see that this separation may reduce abrupt peak errors from residual block, and thus the traditional DCT becomes more efficient to code the resultant ML.

We suggest to use the improved rate-distortion function as shown in (5). The scheme which results in minimum rate-distortion cost \(J\) will be chosen.

\[ J = D(s, c, scheme\{QP\}) + \lambda \cdot R(s, c, scheme\{QP\}) \]  

where \(D\) is the distortion between the original video signal \(s\) and its reconstructed video signal \(c\), which can be measured by the sum of squared differences (SSD) as shown in (6). The scheme represents the encoding method of the current residual block, which can either be the conventional ICT based scheme (scheme-I) or the two-layer coding scheme (scheme-II). \(R\) is the number of bits which are associated with the chosen scheme and QP. For scheme-I, the bits just include one part, which is from the encoding of the residual block produced by one of the nine intra prediction modes. For scheme-II, the bits are composed of two parts for encoding of ML and RBL data. \(\lambda\) is the Lagrange multiplier. After extensive work, we found when \(\lambda\) is calculated according to (9), which is identical to that...
used in JM12.2[13], a better trade-off between distortion and rate can be obtained.

\[ SSD = \sum_{x=0}^{3} \sum_{y=0}^{3} (s(x,y) - c(x,y \mid \text{scheme}))^2 \]  

(6)

\[ R(\text{scheme-I}) = R(\text{intra mode}) + R(\text{residual}) \]  

(7)

\[ R(\text{scheme-II}) = R(\text{intra mode}) + R(\text{ML}) + R(\text{RBL}) \]  

(8)

\[ \lambda = 0.85 \times 2^{(QP-12)/3} \]  

(9)

In (6), c can be calculated according to (10) or (11) for different schemes

\[ c(x,y \mid \text{scheme-I}) = p(x,y) + \text{Dn}(x,y) \]  

(10)

\[ c(x,y \mid \text{scheme-II}) = p(x,y) + \text{Th} \times \text{Dn}_{11}(x,y) + \text{Dn}_{22}(x,y) \]  

(11)

where \( p(x,y) \) is the predicted value. \( \text{Dn}(x,y) \) is the reconstructed value of residual sample produced by scheme-I, \( \text{Dn}_{11}(x,y) \) and \( \text{Dn}_{22}(x,y) \) are the reconstructed values of quotient and modulus produced by scheme-II, respectively.

C. Entropy coding of the scheme

After the residual block has been separated into two layers ML and RBL, each layer will be characterized by its own spatial correlation and with different importance. Therefore, we can apply different encoding techniques to these two layers based on their own characteristics.

\[ \text{SSD} = s(x,y) - c(x,y \mid \text{scheme}) \]  

(6)

\[ R(\text{scheme-II}) = R(\text{ML}) + R(\text{RBL}) \]  

(8)

\[ \lambda = 0.85 \times 2^{(QP-12)/3} \]  

(9)

In (6), c can be calculated according to (10) or (11) for different schemes

\[ c(x,y \mid \text{scheme-I}) = p(x,y) + \text{Dn}(x,y) \]  

(10)

\[ c(x,y \mid \text{scheme-II}) = p(x,y) + \text{Th} \times \text{Dn}_{11}(x,y) + \text{Dn}_{22}(x,y) \]  

(11)

where \( p(x,y) \) is the predicted value. \( \text{Dn}(x,y) \) is the reconstructed value of residual sample produced by scheme-I, \( \text{Dn}_{11}(x,y) \) and \( \text{Dn}_{22}(x,y) \) are the reconstructed values of quotient and modulus produced by scheme-II, respectively.

On the one hand, we can see that this separation can decorrelate RBL as shown in fig.3(d) and fig.4. It is obtained through dividing BL by Th, which is the basic component of the residual block. As a result, the accuracy of it is extremely important. Therefore, we employ lossless compression coding method, which means after this layer is produced, it will be directly reordered and then performed entropy coding. However, the lack of the ICT operation will result in one problem that the samples in RBL will distribute randomly instead of having the energy compaction. If only the traditional zigzag scan is applied into RBL, the efficiency of entropy coding will not be high. Therefore, in order to rearrange the samples in RBL into the order of decreasing energy, we suggest to use multiple scanning modes [9] as depicted in fig.5 to improve the entropy coding efficiency of RBL component. The mode which results in the minimum bits encoded RBL will be selected.

Eight scanning modes for RBL component

Table II lists a typical coding result of one macroblock in the foreman sequence, which are the numbers of the bits of different components for two coding schemes, also gives the rate-distortion cost (J). This table clearly shows that our coding scheme may work more efficiently for some macroblocks. As a result, the coding bits can be reduced through an evaluation of the two schemes.
IV. EXPERIMENTAL RESULTS

The improved encoding algorithm has been integrated into H.264 reference software JM 12.2[13] encoder provided by JVT for performance evaluation. The test was based on the High profile, but only 4x4 ICT was used. All frames were coded as I slice. The quantization parameters (QP) are 22, 27, 32 and 37. A total of nine image sequences were used for tests, including five CIF sequences: Hall, Mobile, Paris, Foreman and Tempete; and four HDTV sequences (720p format): Sunflower, Intotree, Ducktakeoff and Crowdrun. In all cases a total of 150 frames were used. The frame rate was set to 30. The average differences of PSNR (ΔPSNR) and bitrate (ΔBR) were computed using the RD-curve fitting method of 4 data points in [12].

Tables III lists the average coding gain achieved by our proposed coding scheme. In this table, positive values indicate increase, and negative values indicate decrease. The preliminary results do not consider the bits used in coding the extra information (indices of scheme and scan mode). From this table, we can see that the average bitrate reduction using the proposed scheme is 2.632% with an average PSNR increase of 0.208dB for CIF sequences. For the HDTV sequences, a slight improvement also can be obtained, which is 0.846% reduction in bitrate and with an average PSNR increase of 0.052dB. Table III also lists the average coding gain for low bitrate and high bitrate parts. All these can show that our algorithm may reduce the bitrate and improve the PSNR to some good extent. One example rate distortion curve is shown in fig.6, which obviously shows that an improvement has been achieved.

TABLE III. EVALUATION RESULTS OF THE PROPOSED ALGORITHM

<table>
<thead>
<tr>
<th>Format</th>
<th>Sequence</th>
<th>PSNR (dB)</th>
<th>ΔPSNR (%)</th>
<th>ΔBR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>CIF</td>
<td>Hall</td>
<td>0.103</td>
<td>-1.622</td>
<td>-1.902</td>
</tr>
<tr>
<td></td>
<td>Mobile</td>
<td>0.192</td>
<td>-1.759</td>
<td>-2.108</td>
</tr>
<tr>
<td></td>
<td>Paris</td>
<td>0.321</td>
<td>-3.408</td>
<td>-4.889</td>
</tr>
<tr>
<td></td>
<td>Tempete</td>
<td>0.135</td>
<td>-1.491</td>
<td>-1.754</td>
</tr>
<tr>
<td></td>
<td>Foreman</td>
<td>0.288</td>
<td>-4.879</td>
<td>-5.176</td>
</tr>
<tr>
<td>Average</td>
<td>0.208</td>
<td>-2.632</td>
<td>-2.222</td>
<td>-2.364</td>
</tr>
<tr>
<td>HDTV</td>
<td>Sunflower</td>
<td>0.013</td>
<td>-0.217</td>
<td>-0.335</td>
</tr>
<tr>
<td>(720p)</td>
<td>Intotree</td>
<td>0.041</td>
<td>-1.128</td>
<td>-0.871</td>
</tr>
<tr>
<td></td>
<td>Duck</td>
<td>0.048</td>
<td>-0.696</td>
<td>-0.637</td>
</tr>
<tr>
<td></td>
<td>Crowdrun</td>
<td>0.104</td>
<td>-1.341</td>
<td>-0.850</td>
</tr>
<tr>
<td>Average</td>
<td>0.052</td>
<td>-0.846</td>
<td>-0.623</td>
<td>-1.018</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, we have proposed an efficient algorithm for the coding of intra prediction residual blocks. The experimental results show that the preprocessing strategy of residual samples can remove some spatial redundancies produced by inaccurate intra prediction in the current H.264 video coding. It is always desirable to code the extra information and RBL efficiently, which forms one of our future works. We will also apply this idea into the luminance block with the size of 16x16 and 8x8, also into the chroma block with the size of 8x8.

ACKNOWLEDGMENT

This work is supported by the Centre for Multimedia Signal Processing, Hong Kong Polytechnic University and the Research Grant Council of the Hong Kong SAR Government (PolyU 5267/07E).

Reference


