AN ADAPTIVE DIVIDE-AND-PREDICT CODING FOR INTRA-FRAME OF H.264/AVC

Yinji Piao, HyunWook Park

Dept.of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea

ABSTRACT

A new intra-frame coding scheme is proposed for progressive video sequences in H.264/AVC. In the proposed method, every other pixel is sampled either in a horizontal or vertical direction without any anti-aliasing filtering, and two sub-frames of the even-pixel and odd-pixel are then generated. The even-pixel and odd-pixel sub-frames are coded by the intra- and inter-frame coding methods, respectively. Our proposed method can be expanded into a multi-level case in which the even-pixel sub-frame is again subdivided into an even-pixel sub-frame and an odd-pixel sub-frame, and so on to further improve the coding efficiency. The experimental results show that our proposed method reduces the bitrate by approximately 8.25% more than the conventional intra-frame coding in H.264/AVC.

Index Terms— Divide-and-predict, intra coding, intra frame prediction, motion estimation, H.264 video coding

1. INTRODUCTION

Unlike most of the intra-frame coding methods in previous video coding standards, the H.264/AVC [1] standard achieves a much higher compression performance as a result of a new directional spatial prediction technique for intra-frame coding. However, due to the progressive coding order, only the upper and left blocks that have already been reconstructed are utilized to derive the predictors for the current block; hence, the image blocks may be incorrectly extrapolated without knowing the right and bottom neighbors. Efforts to improve the intra-frame coding efficiency on the basis of H.264/AVC have focused on many aspects: for example, bi-directional intra-prediction [2], intra-prediction with multiple reference lines [3], template matching prediction [4], directional transforms [5], and adaptive coefficient scanning [6]. In spite of these efforts, there is still room for improvement.

In addition, there are several new techniques of inter-frame coding: namely, variable block-size motion compensation, quarter-pixel motion compensation, and context-adaptive entropy coding. These techniques make H.264/AVC superior to all other coding standards. In view of the shortcomings of intra-frame coding and the advantages of inter-frame coding, we propose a new intra-frame coding scheme in which inter-frame coding techniques are incorporated into intra-frame coding.

This paper is organized as follows: section 2 presents our proposed intra-frame coding scheme in detail; section 3 describes the corresponding multi-level version; section 4 presents the experimental results; and section 5 summarizes our conclusions.

2. PROPOSED METHOD

A. Adaptive divide-and-predict coding

Frame/field picture coding is supported in H.264/AVC for interlaced video sequences. Because the top and bottom fields are captured at different times in the interlaced video sequences, the field coding is more efficient than the frame coding of fast motion pictures. The concept of field picture coding can be adaptively applied to an intra-frame of a progressive video sequence. We propose a new intra-frame coding scheme based on the concept of field picture coding.

Fig. 1 shows the proposed intra-frame coding structure, which is called adaptive divide-and-predict coding. A frame is coded by a conventional intra-frame coding method. In addition, the input frame is divided into two sub-frames, which are coded by intra-frame and predicted-frame coding methods, respectively. Two modes of horizontal and vertical sampling are used to divide an input frame. The sampling modes are shown in Fig. 2.

In the horizontal sampling process, all odd-numbered columns are sampled and become a new sub-frame; the remaining even-numbered columns are also sampled and become another new sub-frame. As with the horizontal sampling process, the odd-numbered and even-numbered rows of the original frame are sampled respectively and become two sub-frames in the vertical sampling process. The vertical sampling is almost the same as the case of the top and bottom fields in the interlaced sequences. After the sampling process, one of the sub-frame is coded as an independent intra-frame while the other is coded as an inter-frame predicted from the first sub-frame because the first sub-frame has already been reconstructed. According to the picture type to be coded, we call the two sub-frames a sub-intra-frame (Ii) and sub-inter-frame (Ip), respectively. Thus, there are three modes of coding an intra-frame: the
conventional I-picture coding, horizontal sub-sampling coding, and vertical sub-sampling coding. The best coding mode is selected on the basis of comparison of the RD costs of the three different coding modes, and the corresponding bit stream is then transmitted.

Because the original frame is divided into two sub-frames, we can still utilize the nine directional intra-prediction modes in the sub-intra-frame coding to improve the coding efficiency. On the other hand, because the sub-inter-frame can be considered as a half-pixel shift of the sub-intra-frame, there should be high correlations between them; and these correlations definitely benefit the inter-frame coding.

B. Multi-level divide-and-predict coding

Because one of the two sub-frames in the proposed method is coded as an intra-frame, we can easily expand the algorithm into multi-level divide-and-predict coding. That means the sub-intra-frame can be subdivided into two further sub-frames.

Let a pixel in the sub-intra-frame at the $k$th level be $I^s_{i}(x, y)$ while $I^p_{j}(x, y)$ denotes a pixel in sub-inter-frame at the same level. The multi-level divide-and-predict coding can then be described as follows:

(i) For $nth$-level intra-frame coding mode, no further sampling process is needed.
(ii) For $nth$-level horizontal sampling mode, the sub-intra frame of the $(n-1)th$-level is subdivided into the following two sub-frames:

\begin{align*}
I^s_{i}(x, y) &= I^{n-1}_{i}(2x-1, y) \\
I^p_{j}(x, y) &= I^{n-1}_{j}(2x, y)
\end{align*}

(iii) For $nth$-level vertical sampling mode, the sub-intra frame of the $(n-1)th$-level is subdivided into the following two sub-frames:

\begin{align*}
I^s_{i}(x, y) &= I^{n-1}_{i}(x, 2y-1) \\
I^p_{j}(x, y) &= I^{n-1}_{j}(x, 2y)
\end{align*}

where $I^0 = I$ is the original intra-frame to be coded.

From the above equations, only the $(n-1)th$-level sub-intra-frame of the best sampling mode is considered for $nth$-level divide-and-predict coding, and the RD cost comparison is performed on the three coding modes at the $nth$ level. Fig. 3 shows an example of two-level divide-and-predict coding. Because the proposed coding scheme provides two new coding modes, it requires additional bits to flag the coding mode in the slice level. Fig. 4 describes how the coding mode is assigned in the multi-level divide-and-predict coding.

Fig. 1. Blockdiagram of the proposed divide-and-predict coding

Fig. 2. Two sampling modes of the divide-and-predict coding: (a) horizontal sampling (HS), and (b) vertical sampling (VS)

Fig. 3. Example of two-level divide-and-predict coding: first level: HS, second level: VS
3. EXPERIMENTAL RESULTS

The proposed coding scheme, which was implemented on JM11.0 [7], the reference encoder software of H.264, was applied to both the luminance and chrominance components. The common test conditions of the VCEG [8] were applied to the experiment so that the coding performance of the proposed method could be evaluated. In addition to the video sequences recommended in [8], four more 4CIF video were included in the experiments.

When one frame is subdivided into two or more sub-frames, the assignment of an appropriate quantization parameter (QP) to each sub-frame becomes a challenging problem because one of the sub-frames is coded as an intra-frame while the others are coded as inter-frames having different sizes. The best method of assigning QPs is to assign them frame by frame or sequence by sequence. This method balances the distortions of different types and levels of sub-frames and achieves an optimum performance with a cost of heavy computation overhead. However, we introduce a simple method of QP assignment and the feasibility of the method is confirmed experimentally. Let \( QP^0_n \) and \( QP^s_n \) be \( n \)-th level QPs for the sub-intra-frame and sub-inter-frame, respectively. The QP for the \( n \)-th level sub-frames are then assigned as follows:

\[
QF^0_n = QP^0_n - 2 \]
\[
QF^s_n = QP^s_n - 3
\]

where \( QP^0_n = QP(I) \), and \( QP(I) \) is given by user for each I slice, which, in the experiment, is one of 22, 27, 32 and 37.

To show the effectiveness of the proposed method, we compared it with H.264. The comparison was based solely on the intra-frame coding in the baseline profile with respect to the Bjontegaard Delta (BD) method [9] of bitrate reduction and PSNR gain, which were recommended in the VCEG [8]. Table I lists improvements of the proposed method for the single-level versions and the two-level versions with respect to the BD bitrate and the BD PSNR. The proposed method outperforms the conventional intra-frame coding with an average of 0.61dB for 4CIF sequences, 0.40dB for the 720p60 sequences and 0.12dB for the 1080p24 sequence. For the Soccer sequence, a maximum bitrate saving of 15.91% can be achieved. Unfortunately, there is no improvement for the CIF sequences. This lack of improvement occurs because of the following principle: the smaller the frame size is, the weaker the correlation between sub-frames is; as a result, the inter-coding is inefficient. When a sub-frame is subdivided into smaller sub-frames, the correlations between the sub-frames weakens so that the improvement can be limited in the multiple levels. Therefore, the ratio to be coded in any sampling modes in the second level is much lower than the corresponding ratio for the first level. Fig. 5 illustrates the rate-distortion performance and the distribution of the selected modes at various QPs. It also shows that the proposed method is very efficient at a low bitrate and that the horizontal sampling mode is predominant except for the Harbour sequence. Because the Harbour sequence has much more vertical details than horizontal ones, the vertical sampling mode is more efficient than the horizontal sampling mode. The complexity of proposed method is increased because of the three-pass encoding for the best sampling mode. The relative computational complexity compared to the conventional intra frame coding is shown in Table II.

![Fig. 4. Flags for the coding modes of the proposed divide-and-predict coding](image)

### Table I. Improvement of the proposed method in comparison with the conventional intra-frame coding

<table>
<thead>
<tr>
<th>Test sequence</th>
<th>Video Format</th>
<th>Frame Rate</th>
<th>Proposed (one-level)</th>
<th>Proposed (two-level)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bitrate reduction(%)</td>
<td>PSNR gain(dB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreman</td>
<td>CIF</td>
<td>30fps</td>
<td>0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>Mobile</td>
<td>4CIF</td>
<td>30fps</td>
<td>11.78 0.86</td>
<td>12.20 0.87</td>
</tr>
<tr>
<td>Tempete</td>
<td>4CIF</td>
<td>30fps</td>
<td>4.26 0.28</td>
<td>4.41 0.29</td>
</tr>
<tr>
<td>Paris</td>
<td>15fps</td>
<td>0 0</td>
<td>6.00 0.28</td>
<td>8.14 0.36</td>
</tr>
<tr>
<td>Average for CIF</td>
<td></td>
<td></td>
<td>15.0 0.85</td>
<td>15.91 0.92</td>
</tr>
<tr>
<td>Average for 4CIF</td>
<td></td>
<td></td>
<td>9.26 0.57</td>
<td>10.17 0.61</td>
</tr>
<tr>
<td>Average for 720p</td>
<td></td>
<td>7.32 0.39</td>
<td>7.54 0.40</td>
<td></td>
</tr>
<tr>
<td>Average for 1080p</td>
<td></td>
<td>2.85 0.09</td>
<td>4.15 0.12</td>
<td></td>
</tr>
<tr>
<td>Average for 4CIF &amp; 720p &amp; 1080p</td>
<td>7.65 0.43</td>
<td>8.25 0.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II. Computational complexity of proposed method

<table>
<thead>
<tr>
<th>Test sequence</th>
<th>Encoder</th>
<th>Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Proposed</td>
</tr>
<tr>
<td>City (4CIF)</td>
<td>1 3.71</td>
<td>1 0.95</td>
</tr>
<tr>
<td>Harbour (4CIF)</td>
<td>1 3.72</td>
<td>1 0.72</td>
</tr>
<tr>
<td>Crew (720p)</td>
<td>1 3.62</td>
<td>1 0.99</td>
</tr>
<tr>
<td>Night (720p)</td>
<td>1 3.82</td>
<td>1 0.71</td>
</tr>
</tbody>
</table>
4. CONCLUSION

In this paper, we proposed a new intra-frame coding scheme. In contrast to the conventional intra-frame coding, a frame was subdivided into two sub-frames and coded as two successive frames. To guarantee a high correlation between the two sub-frames and to achieve a high compression performance, we incorporated the two sampling methods of horizontal sampling and vertical sampling in the proposed divide-and-predict coding. We also proposed a multi-level divide-and-predict coding method for further improvement of the coding efficiency. The experimental results demonstrated that our proposed method outperformed the conventional intra-frame coding scheme in H.264 with an average PSNR improvement of 0.45dB and reduces the maximum bit rate by 15.91%.

To achieve the optimal coding performance, we compared the RD costs of three different coding modes, which resulted in a high degree of computational complexity. In future work, we intend to conduct a pre-analysis of picture characteristics so that the proposed method can be implemented in a single pass.

5. ACKNOWLEDGMENT

This work was partly supported by LG Electronics.

6. REFERENCES