Novel Fast Mode Decision Algorithm for P-Slices in H.264/AVC

Kan Chang, Bo Yang and Wenhao Zhang
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, China
Pandack0619@163.com

Abstract—H.264/AVC encoder complexity is remarkable mainly due to variable block size motion estimation (ME) and exhaustive Rate Distortion Optimization (RDO). This makes real-time video coding very difficult. In this paper, a novel fast mode decision algorithm for P-Slices is presented to reduce the computational load. Based on the temporal and spatial correlation between macroblocks, a four paths prediction structure is proposed with flexible early termination strategy. Experiment results show that with this algorithm, 42%-65% encoding time can be saved with a negligible loss in Peak Signal-to-noise Ratio (PSNR) and very little increment in bit rate. Combined with Choi’s method [2], our algorithm can achieve a time reduction up to 74.31% with little loss in coding efficiency.

Keywords—H.264/AVC; fast mode decision; early termination; temporal-spatial correlation

I. INTRODUCTION

H.264/AVC is a recent international video coding standard jointly developed by the ITU-T VCEG and the ISO/IEC MPEG standards committees. Compared to previous video coding standards, H.264/AVC has adopted many new techniques to achieve much higher compression efficiency [1]. The design highlights include variable block size and quarter sample accurate motion compensation for improving the rate distortion (RD) performance and decoupling of extra delay associated with bi-predictive coding. The standard allows bi-predictive pictures to be used as references for better motion compensation, weighted offsetting of prediction signal for coding efficiency in scenes and improved “skipped” and “direct” mode inference for better RD performance in video sequences containing neighboring macroblocks moving in a common direction. H.264 further allows directional edge extrapolation in intra coded areas for improving the quality of the prediction signal and allowing prediction from neighboring areas that are inter coded, in loop deblocking filter for removing compression artifacts/proving better quality reconstructed signal for subsequent motion compressions. To achieve high coding efficiency, advanced entropy coding techniques Context-based adaptive variable length coding (CAVLC) and context-based adaptive binary arithmetic coding (CABAC) are used.

H.264/AVC applies rate distortion optimization (RDO) technique to get the best coding result in terms of maximizing coding quality and minimizing resulting data bits. The basic process of RDO can be briefly summarized as follows: the encoder encodes a macroblock (MB) using all the possible modes, such as different block sizes intra predication, different block sizes motion estimation, multiple reference frames. The coding mode with the least RD cost will be used in the final coding. As the RD cost is obtained only after a sequence of operations such as motion estimation, integer transform, quantization, inverse quantization, inverse integer transform and entropy coding, it results in extremely high complexity in the encoder. Therefore, algorithms to reduce the complexity without compromising the coding efficiency are indispensable for real-time implementation of H.264/AVC.

Fast mode decision algorithms have been proposed in the literature to obtain low complexity video coding. Choi et al. used early SKIP detection method and a selective intra-mode search for inter-frames to speed-up the encoding system [2]. This method has been adopted as a fast mode decision option in JM reference software. In [3] Wu et al. make use of the spatial homogeneity of a video object’s textures and the temporal stationary characteristics inherent in video sequences. However, the method suffers from a drawback to obtain the edge image for textual information. Crecos and Yang [4] used the prediction and thresholding scheme to obtain the edge image for textual information. Crecos and Yang [4] used the prediction and thresholding scheme to accelerate the procedure of mode decision. [5], [6] make use of the temporal information between the consecutive pictures. For the current macroblock, a high correlated macroblock is found in the reference frame, due to the high correlation between the successive frames, the mode of the found macroblock can give a good prediction for the mode of the current macroblock. [7]-[9] exploit the spatial correlation between neighboring macroblocks and get the predicted mode from the spatial information.

In this paper, we propose a novel fast mode selection algorithm with an early termination strategy by making full use of both temporal information and spatial information. The rest of this paper is organized as follows. In Section II, a brief overview of mode decision for P-Slices in H.264/AVC is given. In Section III, our proposed algorithm is described. Experiment results are presented in Section IV and conclusions are given in Section V.

II. OVERVIEW OF MODE DECISION FOR P-SLICES IN H.264/AVC

H.264/AVC supports more flexibility in the selection of motion compensation block sizes and shapes, in which the
motion compensation block size may vary from 16x16 to as small as 4x4. There are 4 kinds of size for luminance component of an inter macroblock—16x16, 16x8, 8x16, and 8x8. The 8x8 partitions may be further divided into sub-partitions, such as 8x4, 4x8, and 4x4.

Intra prediction technique for intra macroblock coding, which allows intra blocks to be predicted from spatially neighboring samples, is adopted in H.264/AVC even for P-Slices. There are two types of intra prediction blocks—Intra-16x16 and Intra-4x4 blocks. For Intra-16x16 blocks, there are 4 prediction modes—horizontal, vertical, DC and plane prediction. For Intra-4x4 blocks, there are as many as 9 prediction modes.

As described above, for P-slices, best mode is chosen from the set of candidate modes: \{ SKIP, MODE_16x16, MODE_16x8, MODE_8x16, MODE_8x8 \} (including all 4 partitions). Intra_16x16 (including all 4 prediction modes), Intra_4*4 (including all 9 prediction modes)

In order to select the best mode from all the candidate modes, mode decision is done by minimizing the Lagrangian functional

\[
J(s,c, \text{MODE} | QP, \lambda_{\text{MODE}}) = \text{SSD}(s,c, \text{MODE} | QP) + \lambda_{\text{MODE}} \times R(s,c, \text{MODE} | QP)
\]

where \text{MODE} is one of candidate modes, \( QP \) is the quantization parameter. For Intra or P-slices, the Lagrangian multiplier, \( \lambda_{\text{MODE}} \), is defined in (2). SSD gives the sum of the squared differences between \( s \) (original block) and \( c \) (reconstructed block) according to the formula in (3). \( R(s,c,\text{MODE}|QP) \) presents the number of bits based on \text{MODE} and \( QP \).

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

\[
\text{SSD}(s,c, \text{MODE} | QP) = \sum_{x=1}^{16} \sum_{y=1}^{16} (S_{x,y} - C_{x,y, \text{MODE} | QP})^2 + \sum_{x=1}^{8} \sum_{y=1}^{8} (S_{x,y} - C_{x,y, \text{MODE} | QP})^2 + \sum_{x=1}^{8} \sum_{y=1}^{8} (S_{x,y} - C_{x,y, \text{MODE} | QP})^2
\]

III. PROPOSED MODE DECISION ALGORITHM FOR P-SLICES

A. Stationary Determination

Many natural video sequences contain stationary regions. These regions are mostly encoded in bigger block sizes such as 16x16 or SKIP mode. Therefore, if we can decide that a macroblock is temporal stationary, we can safely skip all the other modes, and encode the macroblock with large block size.

Generally, the sum of absolute difference (SAD) contains lots of information about the motions in successive frames. Large amplitudes will appear on the moving edges or boundaries of moving objects while small amplitudes in homogeneous areas. So if the amplitudes in a macroblock are small, it is most likely that using only larger block sizes in motion estimation will be accurate. The SAD is calculated as follows:

\[
SAD = \sum_{i=1}^{16} \sum_{j=1}^{16} \text{abs}(M[i,j] - N[i,j])
\]

where \( M[i,j] \) and \( N[i,j] \) are the pixel intensities in the current macroblock and the previous macroblock, respectively.

Wu et al. [3] used the SAD between current coding macroblock and its co-located macroblock to determine the stationary condition, setting a predefined threshold \( T_s \) to 200. However, calculation of SAD over every macroblock before RDO for stationary determination is time consuming. Although SAD could be done on a pixel basis, decimated pattern can also be used to reduce time complexity. We use 1:4 decimation pattern in both horizontal resolution and vertical resolution for the calculation of SAD since it achieves a balance between speed and accuracy. The threshold \( T_1 \) for stationary determination is set to 25 in our algorithm.

Note that, in calculating SAD the previous original frame is used as the reference rather than the previous reconstructed one. This is because coding errors in the reconstructed frame may affect the accuracy.

B. Spatial Correlation Based Mode Prediction

As pointed out by many previous works [7]-[9], macroblocks with the same mode tend to cluster together. Based on such spatial correlation, the mode of the neighboring macroblocks can be used to predict the mode of the current one. The positions of the current macroblock and its neighbors are illustrated in Fig.1.

![Fig.1. Positions of spatial neighboring macroblocks](image)

In our method, only the two directly adjacent neighbors—A and B, are used. The reason is that the two macroblocks have closer relation to the current macroblock than others, and can provide enough prediction information. In [7], statistical analysis of the spatial correlation between the neighboring macroblocks is made. According to the neighboring condition, mode of current macroblock is predicted. Different from method in [7], we adopt the modes of macroblock in position A and B directly as the candidate modes. If the RD cost value of the candidate modes of current macroblock is smaller than the threshold \( T_2 \), early termination is triggered.

C. Temporal Correlation Based Mode Prediction

Temporal correlation between successive frames can provide good information for an efficient mode search. If we can accurately use temporal information, speed of the mode decision procedure will be increased. Therefore, to locate the region that has the highest correlation in the previous frame
is crucial. In order to find the most correlated macroblock, Kim [5] processes 16x16 block size ME before coding a macroblock. After the search procedure, the macroblock that has the largest overlapping segment with the displaced block is chosen to be the most correlated one. In order to reduce the complexity in finding the most correlated macroblock, a similar but much simpler algorithm is proposed here.

The threshold RD cost value of this mode is smaller than the threshold T3 in the previous original frame. The chosen position x is determined, the mode of macroblock at position k in the previous original frame.

After position x is determined, the mode of macroblock at this position is selected as the candidate mode of current macroblock. If RD cost value of this mode is smaller than the threshold T3, early termination is triggered.

**D. Early Termination Strategy**

Early termination strategy is developed to speed up the coding procedure. In this part, the threshold T2 and T3 mentioned before are discussed respectively.

The threshold T2 for spatial early termination depends on the modes of position A and B shown in fig. 1. In order to illustrate, we separate the modes available for P-Slices into two classes: class 1 and class 2. Modes {SKIP, MODE_16x16, MODE_16x8, MODE_8x16} are included in class 1 while the others belong to class 2. The modes in class 1 are suitable for coding background, temporal stationary objects while modes in class 2 are used for details or objects with high motion. If modes of position A and B are in the same class, conclusion can be made that the content type of current macroblock are likely to be the same as them. In that case, a loose threshold is suitable. On the other hand, the prediction mode of current macroblock from neighboring macroblock is fallibility and a strict threshold should be used. To sum up, T2 is defined as follows:

$$T_2 = \begin{cases} \min(RD_A, RD_B) \times (1 + \alpha), & \text{If Case 1 or Case 2} \\ \frac{(RD_A + RD_B)}{2} \times (1 + \alpha), & \text{If Case 3 or Case 4} \end{cases}$$

where RD_A, RD_B denote the RD cost of neighboring macroblocks A and B, Mode_A and Mode_B are the modes of spatial neighboring macroblocks A and B, \(\alpha\) is a parameter set to 0.2.

The threshold T3 for temporal early termination is computed by

$$T_3 = \min(T_2, RD_x)$$

where RD_x is the RD cost value of chosen position x defined in (5).

**E. Overall Algorithm**

The proposed fast mode decision is illustrated in Fig. 3, and is described in details as follows.

1) Path A: Stationary condition is firstly checked. If stationary condition is satisfied, only SKIP and P-16x16 modes are set as candidate modes. Otherwise, spatial correlation based mode prediction is needed.

2) Path B: Mode_A, Mode_B, P-16x16 and SKIP are set as candidate modes, and the one that has the minimum R-D cost is recorded. If the minimum R-D cost is less than threshold T2, the recorded mode is chosen to be the best one. Note that, Mode_A, Mode_B may be the same type as P-16x16 or SKIP. Therefore the number of candidate modes may vary from 2 to 4.

3) Path C: If minimum RD cost among candidate modes {Mode_A, Mode_B, P-16x16, SKIP} is larger than T2, temporal correlation based mode prediction is carried out. The process will be terminated for a next macroblock if the RD cost of position x is less than threshold T3. Otherwise the procedure will switch to path D.

4) Path D: If all the early termination condition are not satisfied, final refinement is launched. The rest modes are checked one by one and the RD cost of every checked mode is compared with threshold T3. If the value is below T3, early termination is enabled. Flow chart of this path is shown in fig. 4.
IV. EXPERIMENT RESULTS AND DISCUSSION

The algorithm has been implemented and incorporated into the JM 14.1 reference software. Its outcome is compared with Choi’s mode decision algorithm [2] to verify its effectiveness. Note that, our proposed algorithm does not supply any strategy to early detect the SKIP mode or to stop unnecessary Intra mode checking. Therefore, the proposed algorithm can be combined with mode decision algorithm in [2] to achieve even higher time reduction.

8 sequences are chosen to complete the experiment. They range from sequences of low motion activity such as “hall_monitor” and “container” to sequences with very high motion such as “Mobile” and “Bus”. Some details associated with the experiments are listed below:

1. frames to be coded = 100
2. frame rate = 30.0
3. motion search range = 16
4. intra period = 0 (only first)
5. MV resolution = 1/4 pel;
6. symbol mode = 1 (CABAC)
7. number reference frames = 1
8. number B frames = 0 (IPPP)
9. QP = 28, 32, 36, 40

As for the average PSNR differences and average bit rate differences, calculation follows the specification in [10]. Computational efficiency is measured by the amount of time reduction (TR), which is computed as follows:

\[
\text{Time reduction} = \frac{\text{Time}_{264} - \text{Time}_{\text{proposed}}}{\text{Time}_{264}} \times 100\% \quad (9)
\]

<table>
<thead>
<tr>
<th>Contents</th>
<th>TR[%]</th>
<th>BDP[PSNR[\dB]]</th>
<th>BDBR[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hall_monitor (QCIF)</td>
<td>Choi’s method 45.77</td>
<td>0.001</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Proposed 64.68</td>
<td>-0.023</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 74.31</td>
<td>-0.012</td>
<td>0.21</td>
</tr>
<tr>
<td>Container (QCIF)</td>
<td>Choi’s method 43.99</td>
<td>0.040</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Proposed 61.22</td>
<td>-0.081</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 70.16</td>
<td>-0.095</td>
<td>1.12</td>
</tr>
<tr>
<td>News (QCIF)</td>
<td>Choi’s method 42.43</td>
<td>-0.082</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>Proposed 58.82</td>
<td>-0.067</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 67.18</td>
<td>-0.124</td>
<td>1.34</td>
</tr>
<tr>
<td>Carphone (QCIF)</td>
<td>Choi’s method 29.61</td>
<td>-0.078</td>
<td>-0.66</td>
</tr>
<tr>
<td></td>
<td>Proposed 51.37</td>
<td>-0.092</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 59.19</td>
<td>-0.113</td>
<td>0.98</td>
</tr>
<tr>
<td>Mobile (QCIF)</td>
<td>Choi’s method 17.31</td>
<td>-0.153</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>Proposed 42.86</td>
<td>-0.201</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 46.73</td>
<td>-0.214</td>
<td>0.58</td>
</tr>
<tr>
<td>Foreman (CIF)</td>
<td>Choi’s method 27.12</td>
<td>-0.093</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Proposed 49.93</td>
<td>-0.126</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 56.96</td>
<td>-0.145</td>
<td>2.11</td>
</tr>
<tr>
<td>Flower (CIF)</td>
<td>Choi’s method 23.76</td>
<td>-0.085</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>Proposed 47.34</td>
<td>-0.077</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 51.12</td>
<td>-0.127</td>
<td>1.12</td>
</tr>
<tr>
<td>Bus (CIF)</td>
<td>Choi’s method 21.12</td>
<td>-0.067</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>Proposed 44.03</td>
<td>-0.082</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Proposed+Choi’s 50.42</td>
<td>-0.119</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Note that, if current P frame is the first one in a GOP, path C is not available. It should also be pointed out that macroblocks on the top, left, right, and bottom boundaries of images are always checked using exhaustive mode selection process.
Table I shows a summary of the comparison outcome among proposed algorithm, mode decision method in [2] and the combination of both. As can be observed from this table, our algorithm provides a time reduction in the range of 42%-65% depending on the video content, while generating the same video quality and very little increment in bitrate. Combined with Choi’s method, the proposed algorithm can provide a time reduction in the range of 46%-75% with negligible loss in coding efficiency. It should also be noticed that, Choi’s algorithm achieves a great time reduction for sequences that hold low spatial detail or contain slow motion. When it comes to sequences with high motion or plentiful detail (e.g. “mobile” and “bus”), his algorithm can’t offer satisfied outcome. On the contrary, our algorithm performs well for different kinds of sequences (even for “mobile”, a TR of 42.8% is guaranteed with little loss in coding efficiency).

The graphical representation of the R-D outcome is provided in figs. 5 and 6 for two sample video sequences. Observed from the figures, both of the proposed algorithm and Choi’s method give R-D performance similar to the exhaustive mode search.

V. CONCLUSION

In this paper we proposed a novel fast mode decision algorithm for P-Slices in H.264/AVC. In our algorithm, stationary determination is first carried out to check whether large block size coding is suitable for current macroblock; spatial correlation based mode prediction and temporal correlation based prediction are applied to make full use of temporal and spatial information; if all the early termination conditions are not satisfied, final refinement with early termination is performed over all the modes except the predicted ones. Extensive experiments show that the proposed algorithm can achieve 42%-65% encoding time reduction compared to JM 14.1, and only yield a negligible PSNR loss and little increment in bit rates. Combined with the algorithm proposed by Choi[2], our algorithm can achieve a time reduction up to 74.31% with negligible loss in coding efficiency.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support by the National Science Foundation of China under grantumber 60672087.

REFERENCES