Fast Inter Mode Decision Algorithm Based on Residual Feature

Chen LIU, Tianruo ZHANG, Xin JIN, Minghui WANG, Satoshi GOTO (Member)
Graduate School of Information, Production and System, Waseda University

Summary
H.264/AVC introduces the variable block size motion estimation (VBSME), which brings huge computational cost of the encoder. In this paper, a novel fast inter mode decision algorithm for H.264/AVC has been proposed. The proposed algorithm evaluates the modes based on residual feature. The residual is obtained after the motion search of P16 × 16 mode or P8 × 8 mode. And then basing on the extracted residual feature, the complexity and similarity are evaluated for the inter mode decision. According to the evaluation of similarity between different sub-blocks and the complexity of each sub-block, the most possible inter modes for current block is chosen to be conducted. In the worst case, the proposed whole scheme of inter mode decision algorithm only conducts 4 modes, which is much more effective than conducting all the 8 modes in conventional approach. The simulation results show that, comparing to JM14.1, on average, the proposed algorithm achieves 57.98% and 55.72% time-saving on CIF and 720p sequences respectively, with equivalent 0.219 dB PSNR drop and 5.55% bit rate increase for CIF and 0.107 dB PSNR drop and 3.53% bit rate increase for 720p. Compared to existing inter mode decision algorithm, proposed algorithm achieves 10.68% and 13.26% timing-reduction on CIF and 720p sequences respectively with less performance loss.

Keywords: H.264/AVC, inter mode decision, residual feature

1. Introduction
H.264/AVC, also known as MPEG-4 Part 10 \(^1\) is the latest video compression standard. Compared to previous video coding standards, H.264/AVC achieves a significant bit-rate reduction, a small PSNR gain, while increasing the computational cost by more than one order of magnitude \(^2\). The standout compression efficiency of H.264/AVC is attributed by a series of advanced coding techniques and features, including the multiple reference frames and variable block size motion estimation (VBSME) with Lagrange rate-distortion optimization, integer discrete-cosine transform (DCT), quantization, context-based entropy coding, in-loop-deblocking filter, and so on. All of those features are somehow computation-hungry, while the main computational complexity still comes from the motion estimation (ME) part. Motion estimation costs about 60%–80% portion of computation in the whole encoder system \(^3\). The computation bottleneck in inter frame encoding is the rate-distortion optimization (RDO) based variable block size mode decision.

In H.264/AVC, the inter modes are defined as 8 different partition size as shown in Fig. 1. There are two levels in the tree-like hierarchical structure. The first level contains SKIP mode, P16 × 16 mode, P16 × 8 mode, P8 × 16 mode and P8 × 8 mode. The SKIP mode does not perform the motion compensation. It directly copies the content of the reference macroblock in the reference frame. Each 8 × 8 sub-macroblock in P8 × 8 mode can be further divided into four different types of partition, including P8 × 4 sub-macroblock mode, P4 × 8 sub-macroblock mode and P4 × 4 sub-macroblock mode. The second level of inter mode hierarchy is consisted by the four sub-
In conventional H.264/AVC mode decision process, the Lagrange optimization technique based rate-distortion (R-D) cost is computed for every valid mode to evaluate various coding modes. The mode with the minimum R-D cost is determined as the final mode. The Lagrange function of R-D cost is calculated by Eq. (1)

$$J = D + \lambda \cdot R$$

where $J$ is the R-D cost. $D$ denotes the distortion of current mode. $R$ denotes the number of bits including the macroblock header, motion vectors, and residual data. $\lambda$ is Lagrange multiplier which is defined as Eq. (2).

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

where $QP$ is the quantization parameter of macroblocks.

For getting the R-D cost, the reconstructed image and the actual bit account are necessary. Therefore, the encoder has to conduct the whole encoding-decoding cycle for one mode. Furthermore, the R-D cost has to be calculated for every possible mode. As a result, the mode decision process introduces significant computational burden.

A number of fast inter mode decision algorithms have been proposed these years. They are aimed at reducing the computational complexity and saving the encoding time. These algorithms can be generally categorized into three strategies. The first one is early detection of the SKIP mode. In all the modes defined in H.264/AVC, the encoded bits and computational complexity for SKIP mode MB is the simplest. All the residual, MVD and reference index of the SKIP mode macroblock are zero. Detecting SKIP mode at an early stage and judging some SKIP-like mode as SKIP mode can save the computational time. However, the speed-up of this strategy is limited, and strongly dependent on the contents of video sequences. So it is typically combined with other strategies to achieve higher time-saving rate. Another drawback is that the PSNR loss becomes large when many miss-judgments occur. The second one is macroblock classification. Some of the inter mode decision algorithms categorize all the modes into several classes. However, the speed-up rate is not high when many modes are defined in the same class. The third one is features based evaluation. Many inter mode decision algorithms use certain features as the criteria of mode evaluation. Choosing the effective features and performing reasonable analysis are the most important in these algorithms. Kim uses the temporal correlation between the related macroblocks to aid the current mode decision. Yu uses the texture of original macroblock to serve the complexity measurement. Wu evaluates the homogeneity and stationarity of regions. Wang and Wang use the residual texture to evaluate current mode. For these kinds of algorithms, choosing the proper features and defining the right criterions are very important as they are directly related to the performance and efficiency of algorithms.

In this paper, a fast inter mode decision algorithm based on the residual analysis has been proposed. P16×16 motion search and SKIP mode check is firstly performed to obtain the residual of P16×16 blocks. The characteristic of current macroblock and all the sub-blocks are extracted and the complexity and similarity are evaluated for choosing the next mode to be conducted. If P8×8 mode is chosen, the similar method is applied in the second level. Finally, the mode with the minimal R-D cost is determined as the final mode.

The paper is organized as follows. Section 2 analyzes the residual information and discusses its relationship with inter modes. Section 3 proposes a fast inter mode decision algorithm based on residual information. Section 4 shows the experimental results and performance comparison. Finally, the conclusion is given in Section 5.
2. Residual Based Motion Analysis

2.1 Residual Correlation Discussion

Some existing fast inter mode decision algorithms, such as Yu’s\cite{Yu}, evaluate each macroblock by the complexity of texture in the uncompressed one. The macroblocks with smooth texture are judged as big partition modes, while the ones with rough texture are judged as small-partition modes. However, this scheme is not very accurate. For the macroblocks on the still and complex background, many of those are judged as P8 × 8 mode under the above scheme; while in fact, most of their best mode are big partition modes in conventional full-mode-search inter mode decision. Moreover, according to the R-D cost, described in Eq. (1), the most impacting factor is the bits R. R is mainly decided by the value of residual matrix, but not the original image. Large residual often indicates an inappropriate mode prediction. In the case of big-partition modes, large residual often means that smaller partition mode should be adopted. Thus, in the proposed algorithm, the residual based criterions is used to do the inter mode prediction.

Literature 8) and 11) use residual to assist the prediction. Literature 8) focuses on getting the motion edge by residual. However, edges are not always related to the moving object. Besides, the evaluation methods in 8) cannot extract the edge information exactly. The algorithm proposed in 8) firstly sums up the residual in every 4 × 4 block. It causes performance loss, since the positive and negative numbers are counteracted. Through the summation, some partitions with both big positive and negative residual values are described as plain area. All of the criterions proposed in 8) are based on the sum of residual in 4 × 4 blocks. So the mode judging method of 8) cannot be accurate enough. In literature 11), the evaluation is adopted. Thus, in the proposed algorithm, the residual matrix, but not the original image. Large residual often indicates an inappropriate mode prediction. In the case of big-partition modes, large residual often means that smaller partition mode should be adopted. Thus, in the proposed algorithm, the residual based criterions is used to do the inter mode prediction.

Literature 8) and 11) use residual to assist the prediction. Literature 8) focuses on getting the motion edge by residual. However, edges are not always related to the moving object. Besides, the evaluation methods in 8) cannot extract the edge information exactly. The algorithm proposed in 8) firstly sums up the residual in every 4 × 4 block. It causes performance loss, since the positive and negative numbers are counteracted. Through the summation, some partitions with both big positive and negative residual values are described as plain area. All of the criterions proposed in 8) are based on the sum of residual in 4 × 4 blocks. So the mode judging method of 8) cannot be accurate enough. In literature 11), the evaluation is also based on the residual of 4 × 4 blocks. It evaluates the residual data by root mean square error (RMSE) without considering the average, which also causes inaccuracy. Moreover, in both 8) and 11), the first level (16 × 16 block) mode decision is performed based on the characteristics of four 8 × 8 blocks, which are calculated by the characteristics of four 4 × 4 blocks. Actually, in the first level mode decision, just to evaluate the residual of 8 × 8 blocks is enough. In my proposed algorithm, more effective criterions are defined to avoid problems mentioned above and reduce the inaccuracy and redundancy. The proposed algorithm utilizes the more effective criterions in 12) to avoid the aforementioned problems and reduce the inaccuracy and redundancy.

2.2 Evaluation Method Based on Residual Analysis

The residual produced by the motion estimation is the difference of current block and the reference block, which means the best matching block of current block in the reference frames. Both the sign and the absolute value of residual are important information to get the relationship between current block and its reference block. In the residual, the positive and negative values denote different directions of the difference between current pixels and reference pixels, which means the sign of residual indicates the motion direction. And for the value of residual, the larger absolute value indicates the bigger mismatch.

Consequently, for mode decision, both the sign and the absolute value of residual are the key attributes. First, the sign is used to distinguish the motion trend in a group of pixels. Secondly, within a specific sign, the absolute value reflects the degree of mismatch. Our residual feature evaluation methods work as the following steps.

Assuming the size of current block is n × n. To extract its characteristics, let \( r = \{ r_{ij}, i,j \in [1,n]\} \), each pair of \((i,j)\) represents a pixel in the \( n \times n \) residual block \( Q \). \( r \in R_{n \times n}, r_{ij} \in R \) is the residual value of the \( i\)-th row, \( j\)-th column \((i,j \in [1,n])\). Then, the function \( f \) is performed on the residual block \( Q \) to get the characteristic of current block, which is defined as:

\[
\begin{align*}
    f : R_{n \times n} & \rightarrow R_{2 \times 1} \\
    q & = f(r) = (q_p, q_n)^T \\
    q_p & = \frac{1}{2} \sum_i \sum_j (r_{ij} + |r_{ij}|) \\
    q_n & = \frac{1}{2} \sum_i \sum_j (r_{ij} - |r_{ij}|)
\end{align*}
\]

The complexity of current block \( q \) is defined like:

\[
CPX[q] = \sqrt{q_p^2 + q_n^2}
\]

The similarity of two \( n \times n \) residual blocks \( Q_1 \) and \( Q_2 \) is defined as distance between \( q_1 \) and \( q_2 \):
### 3. Proposed Fast Inter Mode Decision Algorithm Based on Residual Information

#### 3.1 Mode Judgment Process

Assuming the current processing block \( M \) consists of \( 2n \times 2n \) pixel, and \( M \) is divided into four \( n \times n \) sub-blocks named \( A, B, C, D \), as shown in Fig. 2. The characteristics \( q_A, q_B, q_C, q_D \) of block \( A, B, C, D \) are calculated by Eq. (3) respectively. The characteristics of \( M \) are obtained by Eq. (6).

\[
q_M = q_A + q_B + q_C + q_D \tag{6}
\]

The judgment of the modes in one level is shown in step 1 to 6. Here \( n = 8 \) for the first level mode decision and \( 4 \) for the second level mode decision. During this process, at most two inter modes are conducted. Note that the \( 8 \times 8 \) mode decided during step 2 of the first level mode decision \( (n = 8) \) is further conducted in step 1 of the second level mode decision \( (n = 4) \).

- Step 1: Do the \( 2n \times 2n \) block motion search.
- Step 2: If \( CPX[q_M] \) is larger than \( T_1 \), or in \( CPX[q_A], CPX[q_B], CPX[q_C] \) and \( CPX[q_D] \) if there are more than one which is larger than \( T_2 \), then conducts the \( Pn \times n \) mode and go to step 6.
- Step 3: If \( CPX[q_M] \) is less than \( T_3 \), go to Step 6.
- Step 4: If both \( SIM[q_A, q_C] \) and \( SIM[q_B, q_D] \) are less than \( T_4 \), then conducts the \( Pn \times 2n \) mode and go to Step 6.
- Step 5: If both \( SIM[q_A, q_B] \) and \( SIM[q_C, q_D] \) are less than \( T_4 \), then conducts the \( P2n \times n \) mode and go to Step 6.
- Step 6: Processing of this level ends.

\[
SIM[q_1, q_2] = \sqrt{(q_{1p} - q_{2p})^2 + (q_{1n} - q_{2n})^2} \tag{5}
\]
section are executed with \( n = 8 \). And then, if the \( P8 \times 8 \) mode is enabled, the second level mode decision for \( 8 \times 8 \) sub-block is performed using Step 1 to 6 with \( n = 4 \). Finally the mode with the smallest R-D cost in all the conducted modes is set as the best mode for current macroblock.

The proposed scheme is effective. Even in the worst cases, the proposed fast inter mode decision algorithm conducts only 4 modes for one macroblock, which is much more effective than conducting all the 8 modes in conventional JM.

4. Results and Discussions

The proposed fast inter mode decision algorithm has been implemented in reference software JM14.1 \(^{14}\) provided by JVT. CIF and 720p sequences with different characteristics are used for experiments. The encoding parameters are shown in Table 3. Existing fast motion estimation method EPZS \(^{15}\) is adopted as the search mode for all test cases. For comparison, Wang’s \(^{8}\) algorithm, which also utilize residual information in the optimization, is implemented and tested under the same environment. The simulation is accomplished on a PC with Intel(R) Core\(TM\)2 Duo CPU E8500@3.16 GHz and 3.25 GB RAM.

\[
TRR = \left( 1 - \frac{EncodingTime[\text{fast}]}{EncodingTime[\text{JM}]} \right) \times 100\% \quad (7)
\]

Table 4 and Table 5 show the performance of proposed algorithm and Wang’s method \(^8\). For both algorithms, nine CIF sequences and nine 720p sequences are tested respectively. All of the results are compared with conventional full-mode-search JM code and use the same encoding parameters referred above. The time-reduction ratio (TRR) of the whole encoding time in percentage between fast algorithm and conventional JM is defined in Eq. (7). BDPSNR in dB and BDBR in percentage represent the equivalent difference in PSNR and bit rate respectively. They are calculated by the method proposed in 16). Also, the differences of the items between proposed and Wang’s algorithm are given.

Fig. 4 (a)–(d) and Fig. 5 (a)–(d) show the TRR and R-D performance comparison of sequences (CIF: Akiyo and Paris, 720p: ShuttleStart and Sailorman) with QP 20, 24, 28, 32 respectively.

In Table 4 for CIF sequences, it shows that on average, the time saving of proposed algorithm achieves 57.98% which is 10.68% higher compared to Wang’s algorithm \(^8\). On average, the BDPSNR of proposed algorithm is \(-0.219\) dB which is 0.134 dB higher than...
Table 4 Performance comparison of proposed algorithm and Wang’s on CIF sequences

<table>
<thead>
<tr>
<th>CIF Sequences</th>
<th>TRR (%)</th>
<th>BDPSNR (dB)</th>
<th>BDBR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td>46.27</td>
<td>61.31</td>
<td>15.04</td>
</tr>
<tr>
<td>Container</td>
<td>49.17</td>
<td>59.81</td>
<td>10.64</td>
</tr>
<tr>
<td>Foreman</td>
<td>47.16</td>
<td>54.08</td>
<td>6.92</td>
</tr>
<tr>
<td>Hall Monitor</td>
<td>46.42</td>
<td>58.34</td>
<td>11.93</td>
</tr>
<tr>
<td>Headwithglasses</td>
<td>46.64</td>
<td>55.70</td>
<td>9.06</td>
</tr>
<tr>
<td>Highway</td>
<td>46.68</td>
<td>59.16</td>
<td>12.47</td>
</tr>
<tr>
<td>News</td>
<td>48.36</td>
<td>61.60</td>
<td>13.24</td>
</tr>
<tr>
<td>Paris</td>
<td>46.90</td>
<td>53.77</td>
<td>6.87</td>
</tr>
<tr>
<td>Silent</td>
<td>48.09</td>
<td>58.06</td>
<td>9.97</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>47.30</td>
<td>57.98</td>
<td>10.68</td>
</tr>
</tbody>
</table>

Table 5 Performance comparison of proposed algorithm and Wang’s on 720p sequences

<table>
<thead>
<tr>
<th>720p Sequences</th>
<th>TRR (%)</th>
<th>BDPSNR (dB)</th>
<th>BDBR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>42.97</td>
<td>54.91</td>
<td>11.94</td>
</tr>
<tr>
<td>Cyclists</td>
<td>41.61</td>
<td>58.42</td>
<td>16.81</td>
</tr>
<tr>
<td>Night</td>
<td>42.18</td>
<td>51.21</td>
<td>9.02</td>
</tr>
<tr>
<td>Optis</td>
<td>40.80</td>
<td>55.22</td>
<td>14.42</td>
</tr>
<tr>
<td>Raven</td>
<td>42.47</td>
<td>59.98</td>
<td>17.51</td>
</tr>
<tr>
<td>Sailermon</td>
<td>42.43</td>
<td>50.76</td>
<td>8.33</td>
</tr>
<tr>
<td>Sheriff</td>
<td>41.43</td>
<td>53.65</td>
<td>12.22</td>
</tr>
<tr>
<td>Shields ter</td>
<td>43.46</td>
<td>52.33</td>
<td>8.86</td>
</tr>
<tr>
<td>ShuttleStart</td>
<td>44.82</td>
<td>65.05</td>
<td>20.23</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>42.46</td>
<td>55.72</td>
<td>13.26</td>
</tr>
</tbody>
</table>

Wang’s, and the BDBR of proposed algorithm is about 5.55% which is 3.18% less than Wang’s. For the sequence “Akiyo” with the highest average TRR difference 15.04%, at all tested QPs, the TRR of proposed algorithm are much higher than Wang’s (see Fig. 4 (a)), and the difference is up to about 19% on QP 20. And the PSNR drop is small, the R-D curve of proposed one and Wang’s are nearly the same (see Fig. 5 (a)). For the sequence “Paris” with the minimal TRR, proposed algorithm achieves higher TRR on four different QPs than Wang’s (see Fig. 4 (b)), and the R-D performance of proposed one is obviously much better than Wang’s (see Fig. 5 (b)).

In Table 5 for 720p sequences, it shows that on average, the time saving of proposed algorithm achieves about 55.72% which is 13.26% higher than Wang’s. On average, the BDPSNR of proposed one is −0.107 dB which is 0.021 dB higher than Wang’s, and the BDBR is also 0.67% less than the compared one. For sequence “ShuttleStart” with the highest average TRR difference 20.23%, at all tested QPs, the TRR of proposed algorithm are much higher than Wang’s (see Fig. 4 (c)), and the difference is up to about 23% on QP 20. And the PSNR drop is small, the R-D curve of proposed one and Wang’s are nearly the same, and very close to the curve of JM (see Fig. 5 (c)). For the sequence “Paris” with the minimal TRR, proposed algorithm achieves higher TRR on four different QPs than Wang’s (see Fig. 4 (d)), and the R-D performance of proposed one is a little better than Wang’s, and very close to the curve of JM (see Fig. 5 (d)).

Fig. 6 shows the saving factor comparison of proposed algorithm and Wang’s. The saving factor
Fig. 4 TRR of Akiyo (CIF), Paris (CIF), ShuttleStart (720p) and Sailormen (720p) at QP 20, 24, 28 and 32

Fig. 5 R-D performance comparison of Akiyo (CIF), Paris (CIF), ShuttleStart (720p) and Sailormen (720p)
(SF), which means the percentage of saved mode-search amount, is defined in Eq. (8). For all the tested sequences, the proposed algorithm achieves higher saving factor than Wang’s [8]. The saving factor of proposed algorithm is at least 60%, and it’s up to 72% for the best case.

\[
SF = \left(1 - \frac{\#RD_{search[fast]}}{\#RD_{search[JMV]}}\right) \times 100\% \tag{8}
\]

5. Conclusion

A fast inter mode decision algorithm based on residual feature in H.264/AVC has been proposed. The defined similarity and complexity are calculated by the residual feature that obtained from P16 × 16 or P8 × 8 motion search. By evaluating the complexity of current block and the similarity of two subblocks, the candidate modes are selected. Finally the best mode is chosen as the one with the minimal R-D cost. For one macroblock, in the worst case, the proposed algorithm only conducts 3 modes which are much more efficient than conducting all the 9 modes in conventional JM code. The experimental results show that, compare to conventional JM, on average, the proposed algorithm achieves 57.98% and 55.72% time-saving for CIF and 720p sequences respectively, with equivalent 0.219 dB PSNR drop and 5.55% bit rate increase for CIF and 0.107 dB PSNR drop and 3.53% bit rate increase for 720p. In the case of 720p sequences, proposed algorithm keeps the time-reduction ratio and gets better performance on the PSNR and bit rate than the case of CIF sequences. Compared to existing inter mode decision algorithm which also relate the residual data, proposed algorithm achieves 10.68% and 13.26% timing-reduction on CIF and 720p sequences respectively with less the performance loss. And for all the tested results with different QPs and sequences, the proposed algorithm has more timing-reduction ratio than the compared one.

Acknowledgments This research was supported by a grant of Knowledge Cluster Initiative 2nd stage implemented by Ministry of Education, Culture, Sports, Science and Technology (MEXT), CREST JST and “Ambient SoC Global COE Program of Waseda University” of the MEXT, Japan.

References


(Received April 1, 2010) (Revised May 31, 2010)

Chen Liu She received the B.S. degrees in the department of computer science and technology from Nanjing University, China in 2007 and M.E. degree in Graduate School of Information, Production and Systems from Waseda University in 2009. She is currently a Ph.D. candidate of Graduate School of Information, Production and Systems, Waseda University, Japan. Her research interests are about smart video codec.

Tianruo Zhang He was born on October 4th, 1984 in Beijing, China. He received the B.E. degree in Department of Electronic Engineering from Tsinghua University in 2004. He received the M.E. degree in Graduate School of Information, Production and Systems from Waseda University in 2008. He is now a doctor candidate in Waseda University. His research interest includes video coding systems and VLSI architecture.

Xin Jin She received the M.S. degree in Communication and Information System and Ph.D. degree in Information and Communication Engineering both from the Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2005 respectively. She is a visiting lecturer of Information Technology Research Organization, Waseda University. From 2006 to 2008, she was a Postdoctoral Fellow in the Department of Electronic Engineering at The Chinese University of Hong Kong. During 2004 to 2005, she was an Intern in Internet Multimedia Group at Microsoft Research Asia, Beijing, China. She has published over 30 technical papers in refereed conferences and journals. Her current research interests include power-rate-distortion theory, power-constrained video processing and compression, and multimedia communications and applications.

Minghui Wang He was born on June 16th, 1985 in Beijing, China. He received the B.E. degree in Department of Electronic Engineering from Tsinghua University in 2007. He is now a master candidate in Graduate School of Information, Production and Systems of Waseda University. His research interests are about algorithms of image and video processing.

Satoshi Goto (Member) He was born on January 3rd, 1945 in Hiroshima, Japan. He received the B.E. degree and the M.E. degree in Electronics and Communication Engineering from Waseda University in 1968 and 1970, respectively. He also received the Dr. of Engineering from the same university in 1981. He is IEEE fellow, Member of Academy Engineering Society of Japan and professor of Waseda University. His research interests include LSI System and Multimedia System.