Fast MBAFF/PAFF Motion Estimation and Mode Decision Scheme for H.264

Ju-Heon Seo, Sang-Mi Kim, Jong-Ki Han, Nonmember

Abstract—In the H.264, MBAFF (Macroblock adaptive frame/field) and PAFF (Picture adaptive frame/field) is used to enhance the coding efficiency for interlaced video sequences. But the computational complexity of the ME module using the MBAFF and PAFF techniques is huge. Therefore, reducing the speed of MBAFF and PAFF module is one of the important issues to construct an efficient H.264 encoder. In this paper, we proposed three efficient algorithms to reduce the complexity of ME (Motion Estimation) and MD (Mode Decision) modules using MBAFF and PAFF. The simulation results show that the proposed scheme can reduce the computational complexity while the generated bit rate and the image qualities are unchanged.

Index Terms—H.264, Motion estimation, Mode decision, MBAFF coding, PAFF coding.

I. INTRODUCTION

H.264/AVC is an international video coding standard approved by ITU-T as Recommendation H.264 and by ISO/IEC as MPEG-4 part 10 AVC. The scheme has been designed to provide a technical solution appropriate for broadcast, storage device, and conversational service over wireless networks, VOD or multimedia streaming service [1]. In H.264, new coding tools, such as rate constrained coder control, variable block size motion estimation with small block size, quarter sample accurate motion compensation, context adaptive entropy coding, multiple reference frame motion estimation, directional spatial prediction for intra coding, weighted prediction, picture adaptive frame/field (PAFF) [1][2] coding, macroblock adaptive frame/field (MBAFF) [1][3] coding, have been adapted [1]. Due to these efficient technical coding tools, H.264 has a significant performance benefit from the view point of compression efficiency when compared with previous standards.

Among the techniques adapted in H.264, an efficient coding method using MBAFF and PAFF is very important to have high picture quality and compression ratio for interlaced video sequence. Interlaced video sequence with regions of moving objects or camera motion consists of two fields, scanning at different time instants. The two fields of a frame can be coded jointly (e.g. frame coding) or separately (e.g. field coding). Traditionally, pictures with high and low motions are preferable coded by field and frame coding respectively. When a sequence consists of pictures high motion or low motion in the picture level, using PAFF technique can increase the coding performance. On the other hand, in the macroblock level, a picture contains some MBs with high motion and others with low motion, MBAFF coding technique will enhance the coding performance. The coding efficiency can be improved by adaptive coding modes instead of fixed coding mode.

Since the computational complexity of the MBAFF and PAFF coding is one of the biggest modules in the H.264 encoding system, increasing the speed of MBAFF and PAFF is a very important issue to construct an efficient H.264 encoder. Various algorithms have been proposed to reduce the computational complexity of MBAFF and PAFF coding [4][5][6]. In [4], a fast PAFF and MBAFF mode prediction which uses the motion detection and statistic of motion vectors (MVs) has been studied. In [5], a new data structure called MBG (Macro Block Group) for progressive and interlaced video has been proposed for MBAFF coding of H.264. A fast decision scheme [6] using motion activity decides a coding mode of the picture between frame and field types for PAFF coding in H.264 encoder.

In the conventional H.264 [7], MBAFF coding is performed before PAFF process. At first, a current picture is encoded with a frame type, where each MB is compressed by frame MB mode or field MB mode according to the property of pixels in the MB. If the RDcost of the current MB with frame MB mode is smaller than that with field MB mode, the MB is decided to be coded as a frame mode. This process is called as MBAFF. After a current picture is encoded by frame with MBAFF coding technique, the frame is compressed with field picture coding mode where the frame is split into a top field and a bottom field. If the RDcost of the current picture encoded by frame coding with MBAFF is smaller than that resulted from the field picture coding mode with top field and bottom field coding, the picture is decided to be coded as a frame coding mode using MBAFF. Otherwise, the frame is separated into the top/bottom fields, and each field is compressed by field coding mode. This process is called as PAFF. In this paper, we propose a modified MBAFF and PAFF scheme to reduce the complexity of motion estimation (ME) and mode decision (MD) module.

This paper is organized as follows. Section 2 describes the ME and MD scheme using MBAFF and PAFF coding adapted in H.264 codec [7]. In Section 3, we propose an efficient
algorithm for MBAFF and PAFF coding. Computer simulation results for the proposed algorithm are presented in Section 4. The conclusion is given in Section 5.

II. MBAFF AND PAFF IN H.264

The H.264 [7] standard allows three picture coding modes which are frame MB coding, field MB coding, field picture coding. Flowchart of MBAFF and PAFF is represented in Fig.1 [3]. In MBAFF coding, the selection of frame or field coding is at MB level, where an input super MB (whose size is 32x16) can be coded as two frame MBs or two field MBs based on RDcost. In PAFF, an input frame can be coded by the frame picture coding mode incorporating the MBAFF or field picture coding mode.

Fig. 1. Flowchart of MBAFF and PAFF coding for a frame picture

In MBAFF coding, a super MB can be encoded after it is split to a top frame MB and a bottom frame MB as shown in Fig. 2 (a). On the other hand, it can be processed after splitting into two field MBs (a top field MB and a bottom field MB) as in Fig. 2 (b). Note that the sizes of all kind MB (top frame MB, bottom frame MB, top field MB, bottom field MB) are 16x16, except for a super MB (32x16). Coding order is modified due to using super MB which is described in Fig. 3.

Fig. 2. Structure of a super MB in MBAFF coding. (a) A super MB can be split to two frame MBs, (b) A super MB can be split to two field MBs.

In the H.264 [7], a multi-pass approach is used to estimate MVs and optimal modes for a frame picture. Figure 4 shows the flowchart of the conventional process using MBAFF and PAFF coding where mcost and RDcost are defined as in (1) and (2).

\[
\text{mcost} = SAD + \lambda_{\text{motion}} \times \text{Rate}(MV - PMV) \quad (1)
\]
\[
\text{RDcost} = SSD + \lambda_{\text{mode}} \times \text{Rate}_{\text{mode}} \quad (2)
\]

In the inter frame prediction procedure, the reference pictures for a frame are previously reconstructed frames, whereas for the field coding, reference data were previously reconstructed fields. If we consider only one reference picture for simplicity, examples of reference pictures are shown in Fig. 5, Fig. 6, and Fig. 7.

Fig. 3. Coding order modified due to using super MBs in MBAFF coding.

Fig. 4. Conventional ME and MD scheme using MBAFF and PAFF in H.264

Fig. 5. Reference data used for top frame MB and bottom frame MB in frame picture coding using MBAFF (a) Reference data for a top frame MB in a super MB, (b) Reference data for a bottom frame MB in a super MB

Fig. 5 shows ME process and reference data used for frame MBs in a super MB in MBAFF coding, where the nearest one of previously reconstructed frames is used as the reference picture for both a top frame MB and a bottom frame MB. Fig. 6 shows ME process applied to field MBs when the super MB is encoded with top/bottom field MB in MBAFF coding, where the last two field pictures of previously reconstructed pictures are used as the reference data. Fig. 7 represents ME process and reference data used in field picture coding. When a MB in a top field picture is encoded for field picture coding procedure, two previous reconstructed fields are used as the reference pictures.
shown in Fig. 7 (a). On the other hand, when a MB in a bottom field is compressed, only one previous field is used as a reference picture as shown in Fig. 7(b).

Fig. 6. Reference data used for top field MB and bottom field MB in frame picture coding using MBAFF (a) Reference data for a top field MB in a super MB, (b) Reference data for a bottom field MB in a super MB

III. PROPOSED ALGORITHMS FOR MBAFF AND PAFF

In this paper, we proposed efficient algorithms to reduce computational complexity of ME and MD modules for MBAFF and PAFF, where three schemes are proposed. Firstly, because both the field MB coding and the field picture coding use same reference data and their processes are very similar to each other, the results (MVs and modes) of field MBs coding in MBAFF can be used for the processes of field picture coding as the predictive information. Secondly, using similarity between search ranges of top MB and bottom MB in a super MB can further reduce the complexity of ME and MD process. Lastly, RDcost calculated in INTRA coding for a super MB can be reused for INTRA coding for top/bottom field coding.

Used MV for proposed algorithm A, B is integer level, and refinement for proposed algorithm A, B is the same method as accuracy of quarter sample for conventional scheme.

A. Efficient ME scheme for a MB in a top field picture (proposed A)

Figure 6(a) and 7(a) shows that optimal MVs of the top field MB in a super MB and a MB in a top field picture are estimated over the same area. Using the fact can simplify the process of ME for a MB in the top field picture. At first, MVs of the top field MB during MBAFF process are estimated by minimizing mcost of (1). Then, the estimated MVs and mode information are used as predictive MVs (PMVs) in ME procedure for a MB in the top field picture. To increase the accuracy of estimation of MVs for a MB in the top field picture, a refinement process is performed over a narrow search region. Since both ME procedures are similar to each other, the complexity of ME’s can be reduced significantly without any additional degradation in the reconstructed images.

To calculate the computational complexity, we denote the number of MBs in a frame picture as $M$. If the frame is split into top and bottom fields, the number of MBs in each field is $M/2$. When MVs for all MBs in a top field picture are estimated by the conventional scheme, the computational complexity is $O(0.5M \times 2 \times C_{ME})$ since the number of MBs in the top field is $M/2$, MVs are searched over both the top and bottom fields of the reference frame. $C_{ME}$ is the complexity of ME performed for a MB. On the other hand, when the proposed scheme A is used, the computational complexity becomes $O(0.5M \times 2 \times C_{p})$ where $O(C_{p})$ denote that for MV refinement.

B. Efficient ME scheme for a bottom field MB in MBAFF coding (proposed B)

Figure 6 shows the procedures and reference data for ME process for a top field MB and a bottom field MB in MBAFF coding. To simplify the procedure, at first, we estimate an optimal MV of a top field MB over a search range. Then, the estimated MV for a top field MB is used as a PMV for a bottom field MB in a super MB. Due to similarity of two MBs, the refinement of ME process can be performed over a very narrow region. Similar to the complexity described in Section 3.1, the complexity of the conventional ME for bottom field MBs in MBAFF is $O(0.5M \times 2 \times C_{ME})$, while the proposed scheme B reduces the burden to $O(0.5M \times 2 \times C_{p})$.

C. Efficient intra coding scheme for MBs in field picture coding (proposed C)

Figure 8 shows neighbor pixels used in intra prediction coding for a top/bottom field MB in a super MB and MBs in a top/bottom field picture during field picture coding. Since the pixel values in the top/bottom MBs in a super MB are equal to those in the MBs in the top/bottom field, respectively, the RDcosts calculated in 16×16 Intra and 4×4 Intra coding process for top/bottom MBs in a super MB can be reused in Intra coding for MBs in top/bottom fields, respectively. Based on the calculated RDcost of a super MB, we can simply decide an optimal intra coding mode for MBs during field picture coding procedure.

Fig. 7. Reference data used for field pictures in PAFF coding (a) Reference data for a MB in a top field picture, (b) Reference data for a MB in a bottom field picture

Fig. 8. Neighborhood pixels used in Intra Prediction Coding for (a) a top field MB and a bottom field MB in a super MB using MBAFF coding, and (b) MBs in a top field picture and a bottom field picture during field picture coding
To perform MD process for MBs in field picture coding, the conventional scheme considers 7 modes (Inter 16×16 mode, Inter 16×8 mode, Inter 8×16 mode, Inter P8×8 mode, Intra 16×16 mode, Intra 4×4 mode, SKIP mode), while the proposed method calculates RDcost’s for only 5 modes (Inter 16×16 mode, Inter 16×8 mode, Inter 8×16 mode, Inter P8×8 mode, SKIP mode) by reusing the RDcost’s calculated in frame picture coding. Thus, the computational complexity of MD using the proposed scheme for each field picture is \(O(0.5M \times C_{MD} \times 5/7)\) approximately, where \(O(C_{MD})\) denotes the complexity require to decide an optimal mode for a MB. That for the conventional MD algorithm is \(O(0.5M \times C_{MD})\).

The computational complexities of the paths for ME and MD of H.264 encoder are summarized in Table 1. The entire complexity to complete ME and MD for a frame picture is

Conventional H.264 :
\[
4.50(M \times C_{ME}) + 3.0(M \times C_{MD})
\]

Proposed algorithm :
\[
2.50(M \times C_{ME}) + 2.70(M \times C_{MD}) + 2.0(M \times C_{R})
\]

where note that \(O(C_{ME}) \gg O(C_{D})\).

Table. 1 Complexity Comparison between the conventional and the proposed schemes

<table>
<thead>
<tr>
<th>Coding paths</th>
<th>Coding modules</th>
<th>Complexity of H.264</th>
<th>Complexity of proposed schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding for Inter MBs in MEAPF</td>
<td>ME (Fig. 3)</td>
<td>(O(M \times C_{ME}))</td>
<td>(O(M \times C_{ME}))</td>
</tr>
<tr>
<td>Coding for Inter MBs in MEAPF</td>
<td>MD</td>
<td>(O(M \times C_{MD}))</td>
<td>(O(M \times C_{MD}))</td>
</tr>
<tr>
<td>Coding for top SEAM MBs in MEAPF</td>
<td>ME (Fig. 5)</td>
<td>(O(0.5M \times C_{ME}))</td>
<td>(O(0.5M \times C_{ME}))</td>
</tr>
<tr>
<td>Coding for bottom SEAM MBs in MEAPF</td>
<td>MD</td>
<td>(O(0.5M \times C_{MD}))</td>
<td>(O(0.5M \times C_{MD}))</td>
</tr>
<tr>
<td>Coding for top SEAM MBs in MEAPF</td>
<td>ME (Fig. 3)</td>
<td>(O(0.5M \times C_{ME}))</td>
<td>(O(0.5M \times C_{ME}))</td>
</tr>
<tr>
<td>Coding for bottom SEAM MBs in MEAPF</td>
<td>MD</td>
<td>(O(0.5M \times C_{MD}))</td>
<td>(O(0.5M \times C_{MD}))</td>
</tr>
<tr>
<td>Coding for Inter MBs in MEAPF</td>
<td>ME (Fig. 5)</td>
<td>(O(0.5M \times C_{ME}))</td>
<td>(O(0.5M \times C_{ME}))</td>
</tr>
<tr>
<td>Coding for Inter MBs in MEAPF</td>
<td>MD</td>
<td>(O(0.5M \times C_{MD}))</td>
<td>(O(0.5M \times C_{MD}))</td>
</tr>
<tr>
<td>Coding for top SEAM MBs in MEAPF</td>
<td>ME (Fig. 3)</td>
<td>(O(0.5M \times C_{ME}))</td>
<td>(O(0.5M \times C_{ME}))</td>
</tr>
<tr>
<td>Coding for bottom SEAM MBs in MEAPF</td>
<td>MD</td>
<td>(O(0.5M \times C_{MD}))</td>
<td>(O(0.5M \times C_{MD}))</td>
</tr>
</tbody>
</table>

IV. SIMULATION RESULTS

Computer simulations using video sequences were performed to evaluate the proposed algorithm. Test image sequences are “football” and “bicycle” which are interlaced video sequences. The images are encoded by JM10.0 codec [7]. In this test, GOP structure is “IPPPP…..”. The H.264 encoding parameters are listed in Table 2.

Table. 2 Conditions of coding parameters

<table>
<thead>
<tr>
<th>Condition</th>
<th>Parameter values</th>
<th>Condition</th>
<th>Parameter values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial resolution</td>
<td>720×480</td>
<td>Ref frame</td>
<td>1</td>
</tr>
<tr>
<td>Frame number</td>
<td>100(football), 300(bicycle)</td>
<td>ED Opt</td>
<td>Yes</td>
</tr>
<tr>
<td>MC</td>
<td>164 sample accurate</td>
<td>Search range</td>
<td>16</td>
</tr>
</tbody>
</table>

The PSNR (Peak Signal to Noise Ratio)’s of the images encoded at various bit rates are shown in Fig. 9. The PSNR’s of the encoded images are evaluated with respect to the original images. The test images are encoded by “Full search”, a conventional scheme [4], and the proposed scheme. As shown in Fig. 9, the PSNR’s of the images encoded by the proposed schemes are almost equal to those by the full search scheme. It implies that using the proposed scheme does not result in the degradation of the image quality when compared to the conventional scheme.

![Fig. 8. Comparison of Rate distortion curves between the conventional H.264 encoder, Y.Qiu[4]'s scheme, and the proposed scheme for (a) football, (b) bicycle sequence](image)

To compare the computational complexity of the proposed method with those of the conventional schemes, the CPU times consumed by the ME and total encoding module for various sequences are checked in Fig. 10 and Fig. 11, respectively, where the consumed time is displayed in the resolution of msec/frame. As shown in these figures, the proposed scheme requires much smaller computing time than the conventional schemes. From results in Fig. 10 and Fig. 11, we can see that proposed method has good performance for interlaced video sequences.

![Fig. 10. Comparison of consumed CPU times between the conventional H.264 encoder, Y.Qiu[4]'s scheme, and the proposed scheme for football and bicycle sequences](image)
In Table 3, the reduction ratio of the computational complexities of the algorithms is evaluated by

\[
\text{Reduction ratio} = \frac{(\text{CPU time consumed by H.264} - \text{CPU time consumed by a scheme})}{\text{CPU time consumed by H.264}} \times 100\%.
\]

Table 3 Performance comparisons between the conventional and the proposed schemes for football sequence.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>Bit-rate (Kbit/sec)</th>
<th>Computational complexity reduction ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>in the ME-modules</td>
</tr>
<tr>
<td>H.264 [7]</td>
<td>32.02</td>
<td>87.24</td>
<td>0</td>
</tr>
<tr>
<td>Proposed A</td>
<td>32.11</td>
<td>89.00</td>
<td>16.1</td>
</tr>
<tr>
<td>Proposed B</td>
<td>31.77</td>
<td>84.67</td>
<td>19.15</td>
</tr>
<tr>
<td>Proposed C</td>
<td>32.07</td>
<td>83.94</td>
<td>0.44</td>
</tr>
<tr>
<td>Proposed A+B</td>
<td>32.04</td>
<td>91.13</td>
<td>32.39</td>
</tr>
<tr>
<td>Proposed A+B+C</td>
<td>31.11</td>
<td>91.54</td>
<td>39.46</td>
</tr>
</tbody>
</table>

From this table, we can see that the computational complexity of the encoder using the proposed scheme is much less than those of the conventional H.264 encoder while PSNR’s of the encoded images and the generated bit rates are maintained. The proposed scheme can skip some paths in ME and MD process. That is why the proposed scheme can reduce the computational complexity.

V. CONCLUSION

We have proposed an efficient scheme to estimate the motion vector and to decide the mode in H.264 encoder. Since the proposed ME and MD utilize the correlation between field MB coding and field coding for the interlaced fields, the proposed scheme can reduce the computational complexity while the bit rate and the image qualities are unchanged.

Various computer simulations show that the proposed algorithm significantly reduces computational complexity compared with performance of conventional schemes.

REFERENCES