LETTER

Skipping Prediction Directions Based on the Cost Relationship between Multi-Directional Predictions for an HEVC Encoder

Chae Eun Rhee*†, Member

SUMMARY The emerging high-efficiency video coding (HEVC) standard attempts to improve the coding efficiency by a factor of two over H.264/AVC through the use of new compression tools such as various block sizes with multiple directions. Although multiple-directional predictions are among the features contributing to the improved compression efficiency, its high computational complexity keeps it from being used widely. This paper presents an algorithm to skip backward and bi-directional predictions when merge or forward prediction modes are likely to be determined as the best mode. The proposed algorithm takes advantage of the fact that there is a cost relationship among multi-directional predictions and that the results of backward and bi-directional predictions are therefore predictable before the actual operations. After merge and forward predictions, if the expected results of backward and bi-directional predictions are worse than the results up to that point, then additional backward and bi-directional predictions to search for more accurate motion vectors are not performed. A simulation shows that the encoding time is reduced by about 15.18% with a marginal degradation in compression efficiency.

key words: HEVC, H.264/AVC, prediction direction, mode decision

1. Introduction

Recently, the next-generation video coding standard [1], [2] known as high-efficiency video coding (HEVC) was adopted by ISO/IEC MPEG and ITU-T VCEG. In the emerging HEVC standard, several new features are introduced. In particular, various block sizes from 8×8 to 64×64 and flexible block size selection schemes for predictions and transforms have been shown to support high resolutions effectively. In the generalized P and B picture (GPB) as well as the B picture, various inter-prediction modes are feasible for use, such as forward, backward and bi-directional predictions. The final prediction direction is determined considering the compression efficiency of all of the prediction modes. The multi-directional prediction improves the compression efficiency as the temporal correlations among the current frame and various reference frames are exploited. However, the computational complexity increases significantly and the encoding speed decreases as a result. Therefore, it is increasingly important to reduce the complexity of multi-directional predictions without a noticeable degradation of the rate-distortion (R-D) performance.

In H.264/AVC, an early termination algorithm for multi-directional predictions has been proposed for a hardware-based encoder system [3]. The condition for unidirectional motion estimation (ME) is defined based on information from neighboring macroblocks (MBs). If this condition is satisfied, integer ME (IME) and fractional ME (FME) are performed only in the selected direction, i.e., forward or backward. The bi-directional prediction is also omitted. In this way, a speed-up is achieved by selecting unidirectional prediction with lower complexity and discarding multi-directional prediction with higher complexity. However, the direct application of an existing algorithm for a fast HEVC encoding is not proper. The size of the largest coding unit (CU) in HEVC is 64×64, which is much larger than the 16×16 MB in H.264. Thus, the spatial correlation decreases between neighbors with small block sizes. This can severely degrade the quality when a fast decision of the prediction direction for H.264 is directly applied to HEVC. Recently, algorithms for efficient bi-directional predictions have been proposed for HEVC [4]–[6]. It was empirically observed that the R-D costs of forward and backward predictions increase when the bi-directional mode is the best mode. Thus, a bi-directional prediction is made only when the R-D costs of forward and backward predictions are larger than the average R-D cost of previous blocks coded in the forward or backward mode [4]. In Kim’s approach [5], if the difference in the forward and backward MV is less than four pixels, a bi-predictive merge prediction is performed. Otherwise, bi-directional ME is removed. This method reduces the encoding time significantly, but the quality loss is considerable. Another study [6] took advantage of the fact that there is a deep CU hierarchy in HEVC and that there is a strong correlation among the prediction directions of different CUs which share the same area. The prediction of the current CU is determined early as the forward direction when the direction of the upper CU is not backward or bi-directional. This approach is effective in most cases. However, the amount of speed-up is reduced when the depth of the CU hierarchy is decreased.

In this paper, backward/bi-directional predictions are skipped when the expected cost of these predictions are higher than the lowest cost up to that point. The cost of backward/bi-directional predictions is estimated from recent statistics pertaining to cost relationships during encoding. In the HM reference software [7], for one block size, predictions are performed in various directions sequentially. After merge mode and forward predictions, if the expected R-D cost of a backward/bi-directional prediction is greater than the lowest R-D cost up to that point, additional backward

*The author is with the Department of Information and Communication Engineering, Inha University, Incheon, 402–751 Korea.
†E-mail: chae.rhee@inha.ac.kr
DOI: 10.1587/transinf.2014EDL8006
and bi-directional predictions to find more accurate MVs are not performed. Simulation results show that the proposed algorithm achieves encoding time savings of about 15.18% with negligible degradation of the compression efficiency.

The rest of the paper is organized as follows. Section 2 presents the proposed algorithm, and the simulation results are given in Sect. 3. Conclusions are made in Sect. 4.

2. A Skip of Backward and Bi-Directional Predictions Based on a Cost Relation

Figure 1 shows the rate of backward and bi-directional predictions. KristenAndSara with a resolution of 1280 × 720 (denoted as K) and RaceHorses with a resolution of 832 × 480 (denoted as R) are encoded with QPs of 24, 28, 32, and 36. Approximately 9% and 19% of blocks are determined as a backward or a bi-directional prediction in the KristenAndSara and RaceHorses sequences, respectively. In the RaceHorses sequence, which has fast and irregular motion, the rate of a backward or a bi-directional prediction increases compared to the static sequence of KristenAndSara. Here, the dominant direction is a forward prediction in both sequences shown in Fig. 1. If backward and bi-directional predictions (BW and BI, hereafter) are skipped for the block of which the best direction is the forward or merge modes (FW and MRG, hereafter), the complexity of the prediction operation will decrease significantly.

In the HM reference software, FW, BW and BI are performed in sequence. The smallest cost of a multi-directional prediction is compared to the cost of MRG. After that, the best prediction mode for the current block size is determined. If the R-D costs of BW and BI can be estimated early, it becomes possible to decide on the best direction without performing BW or BI. To test the feasibility of estimating the R-D cost, a simple experiment is conducted. Eight video sequences of BasketballDrive and Kimono1 with a resolution of 1920 × 1080; BasketballDrill and RaceHorses with a resolution of 832 × 480; BQSquare and Mobisode2 with a resolution of 416 × 240; and KristenAndSara and Vidy01 with a resolution of 1280 × 720 are encoded with QPs of 24, 28, 32, and 36 and the R-D costs of FW, BW, BI and MRG are then logged to analyze the relationship associated with each cost. For the five recent cases when the best mode is determined as BW or BI, the average decrease in the R-D cost of BW and BI denoted, as R_{BW/BI}, is calculated, as shown in (1), where BestCost_{BW/BI} represents the R-D cost of the best mode, while NonBestCosts represents the costs of the other modes. Using R_{BW/BI} from the cost relationship of the block sizes processed previously, EstimatedCost_{BW/BI} is calculated, as in (2), after MRG and FW assuming that the best mode is BW or BI. The R-D costs of FW and MRG, denoted by Cost_FW and Cost_MR, respectively, are obtained beforehand.

\[
R_{BW/BI} = \frac{\text{BestCost}_{BW/BI}}{\max(\text{NonBestCosts})} \quad (1)
\]

\[
\text{EstimatedCost}_{BW/BI} = \frac{R_{BW/BI} \times \max(Cost_FW, Cost_MR)}{\max(Cost_{FW}, Cost_{MRG})} \quad (2)
\]

In Table 1, \(A_{best}\) denotes the actual best mode, whereas \(E_{best}\) denotes the best mode as determined by a comparison of Cost_FW, Cost_MR and EstimatedCost_{BW/BI}. Here, Cost_FW and Cost_MR are actual values, whereas EstimatedCost_{BW/BI} is calculated using (2). As shown in the third column of Table 1, \(A_{best}\) and \(E_{best}\) show a close match. The fourth column shows the rate of the mismatch between \(A_{best}\) and \(E_{best}\), where \(A_{best}\) is either BW or BI, while \(E_{best}\) is either FW or MRG. The fifth column also shows mismatch cases for which \(A_{best}\) is either BW or MRG while \(E_{best}\) is either BW or BI. If BW and BI are skipped based on the decision with \(E_{best}\), the rate of the fourth column degrades the compression efficiency, whereas the rate of the fifth column represents time which is unfortunately not saved. From the results in Table 1, Cost_{BW/BI} is shown to be fairly predictable from the prior predictions.

This paper proposes a scheme to skip BW and BI predictions when it is expected that the R-D costs of BW or BI are not lowest among the costs of all other predictions. In Fig. 2, the merge mode is performed first. The merge mode plays an important role in compression efficiency and is thus always performed ahead of ME in this paper. The forward prediction is then performed. The smaller value between Cost_MR and Cost_FW is compared to EstimatedCost_{BW/BI}, which is obtained using (1) and (2). If this min (Cost_FW, Cost_MR) value is smaller than EstimatedCost_{BW/BI}, the min (Cost_FW, Cost_MR) value of the current block size is small enough to be determined as the best mode considering recent cases. Thus, backward and bi-directional predictions are not performed. Otherwise, backward and bi-directional predictions are performed due to the possibility that the backward or bi-directional predic-

<table>
<thead>
<tr>
<th>Size</th>
<th>Videos</th>
<th>Match (%)</th>
<th>Mismatch (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920</td>
<td>BasketballDrive</td>
<td>76.78</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>KristenAndSara</td>
<td>77.65</td>
<td>2.17</td>
</tr>
<tr>
<td></td>
<td>BQSquare</td>
<td>72.06</td>
<td>2.11</td>
</tr>
<tr>
<td>832</td>
<td>BasketballDrill</td>
<td>69.98</td>
<td>4.78</td>
</tr>
<tr>
<td></td>
<td>RaceHorses</td>
<td>68.31</td>
<td>1.94</td>
</tr>
<tr>
<td>416</td>
<td>Mobisode2</td>
<td>89.64</td>
<td>0.40</td>
</tr>
<tr>
<td>1280</td>
<td>KristenAndSara</td>
<td>82.96</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Vidy01</td>
<td>83.86</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>77.65</td>
<td>1.78</td>
</tr>
</tbody>
</table>
3. Simulation Results

The proposed algorithm is implemented on the HM10.0 reference software and simulated on a server with two processors at a speed of 3GHz with 8GB of DDR2 RAM. For the simulation, a random-access (RA) scheme is used. Additionally, a low-delay configuration with a picture (LD) is used as an anchor. The fast encoding (FEN) flag is turned on. Sixteen video sequences, BasketballDrive, BQTerrace, Kimono1 and ParkScene with a resolution of 1920 × 1080; BasketballDrill, BQMall, PartyScene and RaceHorses with a resolution of 832 × 480; BasketballPass, BlowingBubbles, BQSquare and Mobisode2 with a resolution of 416 × 240; and FourPeople, Johnny, KristenAndSara and Vidyo1 with a resolution of 1280 × 720, are used in the simulation. Each test sequence consists of 100 frames and is encoded with four QPs (20, 24, 28 and 32).

Table 2 shows the simulation results of the proposed algorithm described in Fig. 2. The first and the second columns represent the resolutions and test sequences used in the simulation. From the third to fifth columns, the increase in the BDBR, the drop in the BDPSNR and the encoding time saved are shown when an LD with four reference frames is applied. Compared to an RA with two reference frames, a time reduction of 15.01% is possible. The consequent R-D drop is an 11.82% increase in the BDBR and a 0.346dB drop in the BDPSNR. An LD shows poor compression efficiency compared to an RA even though more reference frames are used. From the sixth to eighth columns, the proposed algorithm is applied where an RA configuration is used with two reference frames. Compared to an RA with two reference frames, a time reduction of 15.18% is achieved, whereas the R-D performance is a 0.54% increase in the BDBR and a 0.015dB drop in the BDPSNR. The time saved is comparable to that with an LD, whereas the degradation of the R-D performance is marginal.

The performance of the proposed algorithm is compared with that of previous works in terms of time reduction and an increase in BDBR. In Fig. 3, ref4 [4] shows the smallest time reduction, whereas the time reduction of ref5 [5] is significant. However, an increase in BDBR of ref4 is over 4%. Ref6 [6] shows large time reduction with a small BDBR increase. Unfortunately, this good performance of the scheme used in ref6 is only possible when the depth of the CU hierarchy is as many as 4. When the depth is decreased to 3 denoted by ref6_low_depth, the time reduction is reduced almost by half. If the depth is smaller than 3, the scheme in ref6 will not be applicable. Finally, the proposed one shows quite large time reduction with a reasonable BDBR increase.

4. Conclusion

The HEVC standard supports multi-directional predictions as well as variable block sizes. Although a random-access configuration with multi-directional predictions significantly improves the compression efficiency, its high complexity keeps the random-access configuration from be-
ing used widely. Also, an approach for the reduction of prediction directions has not been tested extensively compared to the effort to reduce the number of block size. The proposed algorithm takes advantage of the fact that there is a cost relationship among BW/BI and other directions; thus, the R-D cost of BW/BI is predictable without BU/BI predictions. In the future, the proposed algorithm needs to be elaborated upon further and enhanced considering various factors, such as QP and block size.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2014-H0301-14-1018) and by INHA UNIVERSITY Research Grant (INHA-49292).

References