Depth-of-interest–based Bypass Coding–unit Algorithm for Inter-prediction in High-efficiency Video Coding

Chae Eun Rhee

Department of Information and Communication, Inha University / Incheon, South Korea chae.rhee@inha.ac.kr

Received June 27, 2016; Accepted July 21, 2016; Published August 30, 2016

** Short Paper*

***Abstract*:** The next-generation video coding standard known as High-Efficiency Video Coding (HEVC) was developed with the aim of doubling the bitrate reduction offered by H.264/Advanced Video Coding (AVC) at the expense of an increase in computational complexity. Mode decision with motion estimation is still one of the most time-consuming computations in HEVC, as it is with H.264/AVC. Several schemes for a fast mode decision have been presented in reference software and in other studies. However, a possible speed-up in conventional schemes is sometimes insignificant for videos that have inhomogeneous spatial and temporal characteristics. This paper proposes a bypass algorithm to skip large-block-size predictions for videos where small block sizes are preferred over large ones. The proposed algorithm does not overlap with those in previous works, and thus, is easily used with other fast algorithms. Consequently, an independent speed-up is possible.

***Keywords*:** HEVC, Inter-prediction, Fast mode decision

1. Introduction

Recently, the next-generation video coding standard [1-3] known as High-Efficiency Video Coding (HEVC) was developed by the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). In the emerging HEVC standard, several new features were 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 be very suitable for large resolutions. Thus, mode decisions with motion estimation (ME) are among the most time-consuming computations in HEVC.

HEVC reference software and other recent studies suggest several schemes for rapid mode decisions. These algorithms are useful for speed-up in most cases. However, there can be very little benefit for videos that have inhomogeneous spatial and temporal characteristics. To make up for this weakness, there are numerous ongoing studies for HEVC fast mode decisions that pursue an effective trade-off between degradation of compression efficiency and speed-up degree [4-7].

This paper proposes fast algorithms that attempt to determine the block size early, and consequently, speed up the encoding time by avoiding computations with unselected block sizes. Depending on the video content, large and small blocks have a different probability of being selected as the optimal block size. To skip large blocks for videos that have complex motion and texture characteristics, an intra-prediction routine is executed first. If the rate-distortion (R-D) cost of intra-prediction in the current block size is small enough, compared to that of inter-prediction, smaller block sizes are expected to be more suitable than the current selection due to the high spatial complexity. Thus, the current block size is bypassed and searches continue on smaller block sizes. The fast decision points of the proposed algorithms do not overlap those proposed in the HEVC reference software or in previous studies. Thus, the proposed algorithms are easily used with other fast-decision algorithms, and the resulting speed-up effect is quite independent. Simulation results show that the encoding speed is improved by over 10% for videos that have complex texture and motion, whereas the R-D degradation is marginal. The rest of this paper is organized as follows. Section 2 gives previous fast-decision algorithms. The proposed algorithm and simulation results are presented in sections 3 and 4, respectively. Conclusions are given in Section 5.

2. Previous Works

In the existing video standards, the macroblock (MB) has served as the basic processing unit for a long time. The size of an MB is 16×16 pixels in terms of the luma component. To achieve high compression performance for high-resolution videos, instead of the MB, HEVC defines the coding unit (CU) as the basic unit. Unlike an MB, the size of a CU is not fixed, varying from 8×8 to 64×64. Given the CU size, a variable block type in a quadtree structure is adopted. The largest CU at depth 0 is denoted as the LCU. When the tree depth is 4, the size of the LCU is 64×64. The LCU can be split into as many as four 32×32 CUs. Each 32×32 CU can further be split into four 16×16 CUs. Assuming that the size of a particular CU is 2N×2N, a CU can be split into 2N×2N, 2N×N, and N×2N types of prediction units (PUs). Inter- and intra-predictions are performed for each PU.

HEVC supports variable block sizes of a quadtree structure, and the depth of this tree can be as large as 4. Two hundred fifty-five block partitions in total need to be searched. Thus, accurate selection of candidate block sizes before ME is very difficult. In contrast, a hierarchical decision algorithm is very effective for HEVC because there are many opportunities to terminate further predictions while searching a tree of CUs. For a hierarchical block size decision, the reference software includes several fast-decision algorithms. One of these algorithms is the early CU (ECU) determination. If the SKIP mode is the best mode at the current CU, the predictions for the smaller CUs of the next depth are not performed. For example, if the SKIP mode cost of a 64×64 PU in a 64×64 CU is the smallest among 64×64, 64×32, and 32×64 PUs, the block size search is terminated at this depth, and it becomes unnecessary to run the predictions for the smaller CUs, i.e., the 32×32, 16×16, and 8×8 CUs. The ECU saves a lot of search time through the early mode decision as to the best CU. The goal of the ECU is not to search small CUs. However, the ECU is not useful for videos that have an inhomogeneous texture or complex motion. After encoding these types of videos, it is observed that many small CUs are included and that their coefficients are quite large. This ECU scheme is effective only for videos that have spatially and temporally homogeneous characteristics or that are encoded with a high quantization parameter (QP) value. Many studies have attempted to improve the ECU. Cassa et al. proposed schemes called Top Skip and Early Termination to avoid inter-predictions for large and small CUs when they have little chance to be selected [8]. In the Top Skip algorithm, the starting depth of the current CU is determined according to the depth of a co-located CU, whereas the Early Termination scheme stops the splitting of the CU if the best R-D cost of the current CU is lower than a given threshold. Leng et al. also chose the candidate depth range according to the video contents [9]. The utilization rate of CUs at all depths is analyzed at the frame level; rarely used CUs are skipped. In addition, the neighboring CU information is checked for further CU skipping. In Wang et al.'s study [10], if the average and variance of the best residual of the current CU are less than predefined thresholds, further CU splitting is not performed. Tan et al. made a decision on whether to split the CU by changing the CU processing order [11], as defined in the reference software. Before making a recursive call for each sub-CU, the best R-D cost of the current CU and the sum of the best R-D costs from four sub-CUs are compared, assuming that the sub-CUs are not split further. These previous works try to make up for the weakness of the ECU. However, these CU-level approaches usually incur a large degradation in compression efficiency.

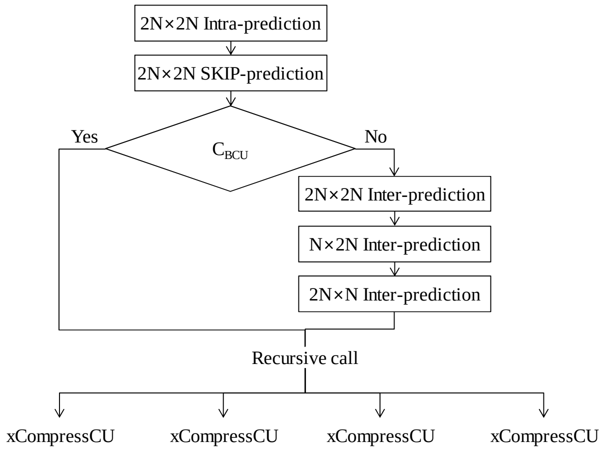

Fig. 1. Mode decision flow with the BCU condition.

3. Depth-of-interest–based Bypass CU Algorithm

To offset the weakness of the ECU, the Bypass CU (BCU) algorithm is proposed here to speed up mode decisions for videos that have complex motion and texture characteristics. In these videos, a small CU is determined as the best CU with a high probability. In the reference software, a search for the best CU is carried out from large to small CUs. Thus, a search for large CUs should be bypassed to reach a small CU as rapidly as possible. In Fig. 1, a flow chart for the proposed BCU algorithm is shown. Initially, the intra- and SKIP-predictions of a 2N×2N PU for the current CU are performed sequentially. Then, if the BCU condition (C_{BCU}) is satisfied, the mode decision flow goes directly to the smaller CUs at the next depth. The predictions for the remaining PUs of the current CU are skipped. C_{BCU} is described in Eq. (1). If the R-D cost of the 2N×2N intra-prediction ($COST_{INTRA}$), is less than that of the 2N×2N SKIP prediction ($COST_{SKIP}$) and multiplied by W_{BCU}, the C_{BCU} is satisfied. The weight value (W_{BCU}) is set to 2. The C_{BCU} condition indicates that the current CU has a low temporal correlation; thus, a more precise inter-prediction is required with smaller CUs.

$$C_{BCU}: COST_{INTRA} < W_{BCU} \times COST_{SKIP} \quad (1)$$

A CU pre-decision should be performed very carefully, because PU decisions are influenced by a CU decision. In this paper, depth-of-interest (DOI) is adopted. The BCU algorithm is executed only when the depth of the current

```
xCompressCU(){
    Get the DOI for the current CU;

    If (the current depth==DOI) UseBCU = False;
    Else                        UseBCU= True;

    If (UseBCU ==True)
        Prediction in an 'intra-first' order of Fig. 1
        The C_BCU is checked to bypass the remaining predictions
    Else
        Prediction in the same order with reference software
        The C_BCU is not checked
```

Fig. 2. A DOI-based early CU decision.

Table 1. R-D performance and time savings of the proposed DOI-based BCU algorithm.

Size	Videos	BDBR (%)	BDPSNR (dB)	TimeSaving (%)		
				low QP	high QP	Average
ClassA	NebutaFestival	0.14	-0.009	8.01	9.75	8.88
	PeopleOnStreet	0.11	-0.003	17.21	13.66	15.43
	Traffic	-0.15	0.004	7.60	2.28	4.94
ClassB	BQTerrace	0.03	-0.003	10.20	2.18	6.19
	Kimono1	0.31	-0.007	7.56	4.28	5.92
	ParkScene	-0.01	0.000	8.16	2.88	5.52
ClassC	BQMall	0.08	-0.002	5.72	2.36	4.04
	PartyScene	0.11	-0.005	4.14	3.50	3.82
	RaceHorses	0.14	-0.006	15.71	8.96	12.34
ClassD	BasketballPass	0.20	-0.007	1.67	1.39	1.53
	BlowingBubbles	0.15	-0.006	6.05	2.74	4.40
	BQSquare	-0.06	0.002	4.78	2.33	3.56
	Average	0.09	0.00	8.07	4.69	6.38

CU is outside the DOI. The minimum and maximum depths (denoted as DMIN and DMAX, respectively) of the DOI are obtained from the left, upper-left, upper, and upper-right LCUs and the co-located LCU. DMIN and DMAX range from 0 to 3. For example, if DMIN and DMAX are 1 and 3, respectively, the DOI of the current LCU ranges from 1 to 3. If the depths of all neighboring LCUs are equally 3, the DOI is only 3. As shown in Fig. 2, the DOI is obtained before the current CU search starts. If the depth of the current CU belongs to the DOI, UseBCU is deactivated. If UseBCU is true, several PU predictions for the current CU are executed in an intra-first order, as shown in Fig. 1, and the C_{BCU} in Eq. (1) is checked to determine whether or not the remaining predictions are bypassed. Otherwise, predictions are executed as described in the reference software, and the BCU algorithm is not applied.

4. Simulation Results

The proposed algorithm was implemented in the HM8.2 reference software and simulated on a server with an Intel Core2 processor at 3GHz with 8GB of DDR2 RAM. For the simulation, configurations for the encoding are low-complexity, low-delay, and generalized P and B picture (GPB). There is one reference frame. The fast encoding (FEN) flag is turned on. Twelve video sequences, NebutaFestival, PeopleOnStreet, and Traffic in Class A at a resolution of 2560×1600; BQTerrace, Kimono1, and ParkScene in Class B at a resolution of 1920×1080; BQMall, PartyScene, and RaceHorses in Class C at a resolution of 832×480; and BasketballPass, BlowingBubbles, and BQSquare in Class D at a resolution of 416×240, were used in the simulation. Each test sequence consisted of 50 frames and was encoded with four QPs (20, 24, 28, and 32).

The DOI-based BCU algorithm was applied to the

HM8.2 reference software; the results were tabulated and are presented in Table 1. The first and second columns represent the video sizes and test sequences. In the third and fourth columns, the increase in the Bjøntegaard-Delta bitrate (BDBR) and the drop in the Bjøntegaard-Delta peak signal-to-noise ratio (BDPSNR) are shown, respectively. The definitions of BDBR and BDPSNR were presented in earlier works [12]. From the fifth to seventh columns, the encoding time saved is shown. The fifth column shows results when QP values are 20 and 24, whereas the sixth column shows results when QP values are 28 and 32. The average time-saving values are presented in the seventh column. The encoding time reduction achieved is 8.07% with low QP values and 4.69% with high QP values. The encoding time was reduced by 6.38%, on average, whereas the R-D drop is marginal. Specifically, the proposed DOI-based BCU algorithm shows significant time savings in the PeopleOnStreet and RaceHorses samples, which have complex texture and motion. In these cases, small CUs were determined as the best type of CU with high probability. Thus, bypassing the large CU in the proposed algorithm reduces the encoding time more effectively.

5. Conclusion

The main contribution of the proposed algorithms is that the fast decision points do not overlap those proposed in previous studies. Thus, the proposed algorithms are easily used with other fast-decision algorithms, and the consequent speed-up effect is quite independent.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1C1A1A02037625)

References

[1] ISO/IEC JTC 1 SC29 WG11, "Joint Call for Proposals on Video Compression Technology," Doc. N11113, Jan. 2010. Article (CrossRef Link)

[2] ISO/IEC JTC 1 SC29 WG11, "Vision, Applications and Requirements of High-Performance Video Coding," Doc. N11096, Jan. 2010. Article (CrossRef Link)

[3] Wiegand, T., Han, W.J., Bross, B., Ohm, J. R., and Sullivan, G. J.: 'WD4: Working Draft 4 of High-Efficiency Video Coding', JCTVCF803, Torino, Itary, July 2011. Article (CrossRef Link)

[4] G. Correa, P. Assuncao, L. Agostini and L.A. da Silva Cruz, "Complexity control of high efficiency video encoders for power-constrained devices", *IEEE Trans. Consumer Electronics*, Vol. 57, No. 4, pp.1866-1874, Nov. 2014. Article (CrossRef Link)

[5] L. Shen, Z. Zhang and P. An, "Fast CU size Decision and moDe Decision algorithm for HEVC intra coding", *IEEE Trans. Consumer Electronics*, Vol. 59, No. 1, pp. 207-213, Feb. 2013. Article (CrossRef Link)

[6] J. Kim, S. Jeong, S. Cho and J. S. Choi, "Adaptive coding unit early termination algorithm for HEVC", *in Proc. Of IEEE International Conference on Consumer Electronics*, pp. 261-262, 2012. Article (CrossRef Link)

[7] J. Kim, Y. Choe and Y.-G. Kim, "Fast coding unit size decision algorithm for intra coding in HEVC", *in Proc. Of IEEE International Conference on Consumer Electronics*, pp. 637-638, 2013. Article (CrossRef Link)

[8] M. B. Cassa, M. Naccari, F. Pereira, "Fast rate distortion optimization for the emerging HEVC standard," *in Proc. of Picture Coding Sym.*, pp.493–496, May, 2012. Article (CrossRef Link)

[9] J. Leng, S. Lei, T. Ikenaga, S. Sakaida, "Content Based Hierarchical Fast Coding Unit Decision Algorithm for HEVC," *in Proc. of Int. Conf. on Multimedia and Signal Processing*, pp.56–59, May, 2011. Article (CrossRef Link)

[10] H. Wang, Y. Wee, Y. Kim, D.-S. Jun, S.-h. Jung, J. S. Choi, "An early termination method using the residual in High Efficiency Video Coding," *in Proc. of IEEE Int. Sym. on Broadband Multimedia Systems and Broadcasting*, pp.1–2, June, 2012. Article (CrossRef Link)

[11] H. L. Tan, F. Liu, Y. H. Tan, C. Yeo, "On fast coding tree block and mode decision for high-Efficiency Video Coding (HEVC)," *in Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp.825–828, March, 2012. Article (CrossRef Link)

[12] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," presented at the 13th VCEG-M33 Meeting, Austin, TX, Apr. 2–4, 2001. Article (CrossRef Link)

Chae Eun Rhee received the B.S., M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2000, 2002 and 2011, respectively. From 2002 to 2005, she was with the Digital TV Development Group, Samsung Electronics Company Ltd., Suwon City, Korea, as an Engineer, where she was involved in bus architecture and MPEG decoder development. In 2013, she joined the Department of Information and Communication Engineering at Inha University, Korea, where she is currently working as an assistant professor. Her research interests include algorithm and architecture design of video coding for HEVC and H.264/AVC and configurable video coding for real time systems.