ELectronics EXpress

LETTER

IEICE Electronics Express, Vol.9, No.21, 1695-1700

Reference frame selection in a hardware-based HEVC encoder

Chae Eun Rhee and Hyuk-Jae Lee[a)]

Inter-university Semiconductor Research Center, Department of Electrical Engineering, Seoul National University, Seoul, 151–742, Korea

a) *hyuk_jae_lee@capp.snu.ac.kr*

Abstract: The multiple-reference-frame motion estimation (ME) is one of the features to improve the compression efficiency. However, the computational complexity for prediction increases in proportion to the number of reference frames. This paper proposes the reference frame selection algorithm for a hardware-based HEVC encoder. The integer-ME explores multiple reference frames to find the best one and the fractional ME is then performed for the best reference frame which is determined by the IME. Simulation results show that a significant time saving of 74% is achieved with a negligible drop in compression efficiency.

Keywords: HEVC, multiple reference frames, reference frame selection

Classification: Electron devices, circuits, and systems

References

[1] Q. Sun, et al., "A Content-adaptive Fast Multiple Reference Frames Motion Estimation in H.264," *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 3651–3654, May 2007.

[2] Y. Su, et al., "Fast multiple reference frame motion estimation for H.264," *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 1, pp. 695–698, June 2004.

[3] N.-J. Kim, et al., "Two-Bit Transform Based Block Motion Estimation Using Second Derivatives," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 902–910, May 2009.

[4] Y.-W. Huang, et al., "Analysis and Complexity Reduction of Multiple Reference Frames Motion Estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 507–552, April 2006.

[5] X. Li, et al., "Fast multi-frame motion estimation algorithm with adaptive search strategies in H. 264," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, pp. 369–372, May 2004.

[6] T.-C. Chen, et al., "Analysis and architecture design of an HDTV720p 30 frames/s H.264/AVC encoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 673–688, June 2006.

[7] P. K. Tsung, et al., "Cache-based integer motion/disparity estimation for quad-HD H.264/AVC and HD multi view video coding," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 2013–2016, April 2009.

[8] Y. K. Lin, et al., "A Hardware-efficient H.264/AVC motion-estimation design for high-definition video," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 6, pp. 1526–1535, 2008.

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012

[9] C. Yang, et al., "High performance VLSI architecture of fractional motion estimation in H.264 for HDTV," *Proc. IEEE Int. Symp. Circuits and Systems*, May 2006.
[10] C.-Y. Kao, et al., "A High-Performance Three-Engine Architecture for H.264/AVC Fractional Motion Estimation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 4, pp. 662–666, 2010.

1 Introduction

Video compression technologies have attracted industry attention due to the increasing popular demand for high-definition (HD) video content and used by video applications such as video conferencing, streaming, video storage and communication. H.264 has been regarded as the state-of-the-art video coding standard. Recently, the next-generation video coding standard known as High Efficiency Video Coding (HEVC) has been developed by ISO/IEC MPEG and ITU-T VCEG. The HEVC standard aims at bitrate saving by a factor of two over H.264 at the expense of an increase in computational complexity. Like H.264, the motion estimation (ME) remains among the most time-consuming computations in HEVC. In an inter-prediction mode decision, more than one frame is used as the reference frames for ME. Although the multiple-reference-frame ME (MRF-ME) improves the compression efficiency, the computational complexity increases in proportion to the number of reference frames. Moreover, the main target resolution of HEVC is full HD (1920×1080) and beyond. Therefore, fast inter-prediction is an important challenge to be solved for HEVC compression to be used in real-time consumer electronic devices. Extensive research effort has been conducted to reduce the number of reference frames for H.264, pursuing an effective trade-off between the rate-distortion (RD) drop and the speed-up. In order to deal with the similar challenge for HEVC, major differences between H.264 and HEVC are investigated. Then a new reference frame selection algorithm is proposed for a hardware-based HEVC encoder. The rest of the paper is organized as follows. Section 2 surveys previous approaches for H.264. The assumed hardware implementation is given in Section 3. The application of the proposed reference frame selection algorithm and simulation results are presented in Sections 4 and 5, respectively. Conclusions are given in Section 6.

2 Previous works on the reference frame selection in H.264

The MRF-ME technique improves the coding efficiency. However, not all of the video sequences can achieve a substantial gain from the MRF-ME because the MRF-ME gain is closely related to the content and varies from sequence to sequence. A number of previous algorithms are proposed to reduce the number of multiple reference frames. These algorithms can be classified into two categories. For the first category, the complexity of ME is adjusted according to the reference frames. For the reference frames close to the current frame, the accurate ME with a high complexity algorithm is used. On the other hand, the relatively low complexity ME is used for the

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012

farther reference frames. In [1], the search range for every reference frame is adjusted based on motion characteristics of videos or the ME result obtained from the reference frame already processed. To further reduce the ME complexity, MVs obtained from ME in the current reference frame are used to predict MVs for farther reference frames. Thus, the predictions for farther reference frames are performed without full ME operation [2]. In [3], the ME complexity is adjusted by using the ME with low-bit resolution per a pixel for farther reference frame, whereas 8-bit resolution ME is used for the closer reference frames. For the second category algorithm, only a subset of reference frames close to the current frame are used for motion prediction. In [4], the selection criteria to determine if the remaining frames should be estimated or not are obtained from the macroblock (MB) modes, residuals, and MV consistency of the processed frames. In these schemes, the ME operations often stop searching farther reference frames when the texture is homogeneous and the motion activity is low. Similarly, in [5], references 0, 1 and 2 are always searched and the decision for the further search on references 3 and 4 is made according to the motion consistency of references 0, 1 and 2.

3 Hardware organization for an HEVC encoder

To achieve high compression performance for high-resolution videos, HEVC defines the coding unit (CU) as the basic unit instead of the MB. Unlike an MB of which size is fixed as 16×16 pixels, the size of a CU is not fixed, varying from 8×8 to 64×64. Given the CU size, a variable block type of quad tree-structure is adopted. The depth of this tree is as large as four. The largest CU is denoted by LCU. Assuming that the size of a particular CU is $2N \times 2N$, a CU can be split into $2N \times 2N$, $2N \times N$, $N \times 2N$ and $N \times N$ types of prediction units (PUs). Therefore, a significantly large number of block sizes need to be searched in HEVC when compared with H.264.

To examine the impact of the HEVC coding structure on the hardware implementation, it is assumed that the 2-stage-pipeline ME implementation for the HEVC hardware encoder may be similar to the widely-used architectures of H.264 encoders [6] where the integer motion estimation (IME) is performed in the earlier stage, whereas the fractional motion estimation (FME) with the motion compensation (MC) is performed in the later stage. To execute the IMEs for all block sizes of an MB in parallel, the SADs for all 4×4 blocks of an MB are calculated simultaneously. The obtained SAD values are combined in the VBS (variable-block-size) adder tree and 41 SADs for all block sizes are generated in one cycle. Here, each block has an MV predictor (MVP) which is a modified MVP for parallel execution [6]. The solution for parallel IME executions for H.264 is able to be applied for an HEVC encoder. The MVP of $2N \times 2N$ PU in LCU is used for all blocks equally. For FME, it may not be as easy as IME to exploit high parallelism to speed up the execution time. In H.264, FME is conducted one by one for 41 blocks in a 16×16 MB, whereas FMEs for 425 blocks are executed for a 64×64 LCU in HEVC. As a result, the execution time for FME with the MC in HEVC may become significantly larger than that for IME in HEVC.

To estimate the execution times of IME and FME, the IME and FME designs for H.264 are used as a hardware implementation of an HEVC

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012

encoder does not exist yet. Then, the operation times of the IME and FME are compared to find the more critical operation in terms of the encoding speed. To this end, representative and latest proposals for H.264 IME and FME are used for estimation. For IME, the IME execution cycles are 323 and 256 cycles per MB in the designs presented in [7] and [8], respectively, whereas the gate counts are 230 K and 213.7 K, respectively. To process a 64×64 LCU with the same IME module for HEVC, the IME should perform 16 times. After that, VBS adder tree which may be larger than that in an H.264 encoder is used to generate SAD values of all different block sizes. Thus, the expected cycle of the IME operation for an LCU is approximately 4,096 with a marginal PSNR drop of 0.005 dB when the IME of [8] is used. Meanwhile, the FME in [9] uses 16-pixel-width interpolators and can process an MB in 790 cycles, whereas the FME in [10] adopts multiple interpolators and the FME operation time is 631 cycles. The gate counts are 311 K and 321 K for FMEs in [9] and [10], respectively. In both FMEs, the mode reduction schemes are not applied to avoid PSNR degradation. Considering the increased number of block sizes, the processing cycles per LCU in HEVC are estimated as 16,000 cycles based on the processing cycles per MB. Thus, the encoding time ratio between the IME and FME is approximately 1 versus 3.9. This estimation of execution time leads to a conclusion that the encoding time is most likely determined by the time for the FME.

4 Reference frame selection algorithm for HEVC

For the early termination of searching multiple reference frames, a number of statistics and proper conditions have been presented to determine whether or not farther reference frames should be searched for H.264 encoding [4]. Among these conditions, three are chosen to be applied to an HEVC encoder. After searching the current reference frames, the following conditions are tested. First, the transformed and quantized coefficients of each CU are close to zero. If the current residual is zero, the more computation for the remaining frames will not further reduce prediction errors. Second, the location of the MV of each CU is the integer pixel. In the real world, the motion and the texture are continuous. However, input videos from a camera are the sampled images, spatially and temporally. One of the usefulness of the multiple reference frames is to relieve the discrete sampling defects. Therefore, if the best MV of the current CU is the integer point, multiple reference frames ME may not be very helpful. Third, the MVs of large CUs and small CUs are consistent. This condition detects whether the motion is homogeneous or not. If these three conditions are satisfied for the CU in the current reference frame, searches for farther reference frames are terminated. Otherwise, the next reference frame is searched and the conditions are tested again.

The performance of the early termination algorithm varies depending on video contents. In other words, in most cases, it is difficult to terminate searching multiple reference frames in videos which have fast and complex motion characteristics. To compensate this weakness of the early termination algorithm, the IME-based reference frame selection algorithm is proposed. As explained in Section 3, the reduction of FME operation time is important in HEVC, whereas the IME explores multiple reference frames

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012

to find the best one without difficulty because the IME which executes with a high parallelism affords time with a small loss in compression performance compared to the FME. For the HEVC hardware implementation, the proposed algorithm considers the dominance of the FME time in the total ME time in a hardware-based HEVC encoder. After IME operation, each block in an LCU has the best reference frame which is determined in the IME phase. The FME of each block is then performed not for all reference frames but for the best reference frame. The proposed algorithm achieves a speed-up in encoding time by reducing the dominant FME operation time.

5 Simulation results

For the simulation of a hardware-based HEVC encoder, 2-stage pipeline schedule is assumed where the time ratio between the IME and the FME is set as 1 versus 3.9 as explained in Section 3. The configurations for the encoding are low-complexity, low-delay, and P picture-only and the number of reference frames is four. Twelve video sequences are used in the evaluation. Each test sequence consists of 50 frames and is encoded with four QPs (20, 24, 28 and 32).

In Table I, the first and second columns represent the resolutions and test sequences used in the simulation, respectively. From the third to fifth columns, the increase in the bitrate, the PSNR drop and the saved time for the ME operation, denoted by ΔB, ΔP and ΔT, respectively, are shown when the number of reference frames is set to two which are the nearest and the second nearest reference frames compared to when four reference frames are used. 50% of time saving is achieved but the RD drop is significant. From the sixth to eleventh columns in Table I, the simulation results of the early termination algorithm are shown. From the sixth to eighth columns, the change of the RD performance and time saving are shown when the decision to search farther reference frame is made after Ref0 is searched. The time saving is 43.37% on average, whereas the increase in the bitrate and PSNR drop are 0.16% and 0.03 dB, respectively. However, for videos which have fast or complex motion characteristics such as *Keiba*,

Table I. RD performance degradation and the ME time saved when the previous early reference frame selection scheme is used

Size	Videos	2 reference frames (Ref0 & Ref1)			Early termination after 1 reference frame			Early termination after 2 reference frames		
		ΔB (%)	ΔP (dB)	ΔT (%)	ΔB (%)	ΔP (dB)	ΔT (%)	ΔB (%)	ΔP (dB)	ΔT (%)
832 ×480	BQMall	0.67	-0.03	50	0.34	-0.03	56.77	0.28	-0.03	37.41
	FlowerVase	3.68	-0.10	50	1.17	-0.08	50.68	0.93	-0.05	32.83
	Keiba	2.92	-0.04	50	0.10	-0.01	11.54	0.31	0.00	6.26
	RaceHorses	1.56	-0.03	50	0.00	-0.01	9.20	0.00	0.00	5.73
1280 ×720	FourPeople	-0.04	-0.01	50	-0.07	-0.02	68.92	0.03	-0.01	45.83
	Johnny	-0.64	-0.04	50	-0.14	-0.05	60.98	-0.52	-0.03	39.95
	KristenAndSara	-0.26	-0.02	50	-0.01	-0.04	60.02	0.07	-0.01	38.94
	Vidyo1	0.31	-0.02	50	0.05	-0.03	65.30	0.02	-0.01	43.18
1920 ×1080	Aspen	6.42	-0.08	50	0.15	-0.02	31.63	0.43	-0.02	20.99
	BasketBallDrive	0.47	-0.03	50	0.08	-0.01	18.75	-0.08	0.00	11.85
	SnowMoutain	0.10	-0.02	50	0.17	-0.02	62.74	0.01	-0.02	41.67
	Kimono1	0.14	-0.01	50	0.06	0.00	23.87	-0.01	0.00	14.93
	Average	1.28	-0.04	50	0.16	-0.03	43.37	0.12	-0.01	28.30

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012

RaceHorses and *BasketBallDrive*, the saving in the ME operation time is very small. The ninth to eleventh columns show the simulation results when the decision to search farther reference frames is made after Ref1. The RD performance is enhanced compared to the early termination after Ref0, whereas the time saving is 28.3%.

Table II shows the simulation results of the proposed algorithm. From the third to fifth columns, only a single reference frame which is the best frame from the IME phase is used for FME operation. 74.3% of time saving is achieved, whereas the increase in the bitrate and PSNR drop are 0.57% and 0.03 dB, respectively. In the sixth and eighth columns, the simulation results with two best reference frames are presented. The increase in bitrate is just 0.1% and the PSNR is degraded by 0.01 dB, whereas the saving in the ME operation time is 50%. Compared to the previous approaches in Table I, the significant time saving is achieved for all videos, whereas and the consequent RD drop is negligible.

Table II. RD performance degradation and the FME time saved when the proposed reference frame selection algorithm is used

Size	Videos	1 reference frame			2 candidate reference frames		
		ΔB (%)	ΔP (dB)	ΔT (%)	ΔB (%)	ΔP (dB)	ΔT (%)
832	BQMall	0.86	-0.03	74.3	0.35	-0.02	50
×480	FlowerVase	2.20	-0.07	74.3	0.48	-0.04	50
	Keiba	0.69	-0.03	74.3	0.30	-0.01	50
	RaceHorses	0.81	-0.03	74.3	0.08	-0.01	50
1280	FourPeople	0.11	-0.01	74.3	0.00	0.00	50
×720	Johnny	0.13	-0.03	74.3	-0.30	-0.01	50
	KristenAndSara	0.01	-0.02	74.3	-0.16	-0.01	50
	Vidyo1	0.50	-0.02	74.3	0.23	0.00	50
1920	Aspen	0.62	-0.02	74.3	0.08	-0.01	50
×1080	BasketBallDrive	0.33	-0.02	74.3	0.08	0.00	50
	SnowMoutain	0.44	-0.02	74.3	0.13	-0.01	50
	Kimono1	0.10	-0.01	74.3	-0.02	0.00	50
Average		0.57	-0.03	74.3	0.10	-0.01	50

6 Conclusion

In the hardware implementation for the HEVC standard, the execution time for FME becomes much larger than that for IME because IME execution can be speed up by exploiting parallelism. In this paper, the reference frame selection algorithm is proposed focusing on the FME operation time. In the future, various fast encoding schemes for the FME need to be further researched.

Acknowledgments

This work was supported by Industrial Strategic Technology Development Program funded by the Ministry of Knowledge Economy (MKE, Korea) (10039188, Development of multimedia convergence programmable platform for smart vehicles).

DOI: 10.1587/elex.9.1695
Received September 24, 2012
Accepted October 25, 2012
Published November 13, 2012