An Simplified Reference Frame Selection Configuration Method

Ping Kuang[†1] and Jianhua Deng[†2*]

School of Information and Software, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 610054
{kuangping[1], jianhua.deng[2]}@uestc.edu.cn

Abstract

Reference Frame Selection (RFS), as reference frame management mechanism in High Efficiency Video Coding (HEVC), plays an important role in constantly developing video coding standard. As RFS has been improved to adapt the introduction of generalized B frame in HEVC. In this paper, a simplified coding configuration method is presented to simplified reference frame management process to cut down on encoding time. This method is applied in real time low-delay encoding configuration with specific simplified rules. Simplified method is conducted on the reference list with biggish Qualitization Parameter (QP) offset and improves coding efficiency with about 10.5% reduction on encoding time and only 0.5% negligible raise on BD-bitrate.

Keywords: HEVC, RFS, depth information, real-time configuration

1. Introduction

HEVC standard is developed by the Joint Collaborative team on video coding and ratified as an international video coding standard to meet the increasing demands for high resolution video.

HEVC is on the block-based hybrid coding framework [1]. Compared with the previous standards, numbers of coding techniques were proposed and accepted to better compress high-difinition video contents. HEVC introduces generalized B frame with more complex GOP structure and reference frame management to reference frame selection (RFS) mechanism [2]. In video coding, a frame encoded by intra prediction is called I frame. And a frame encoded by inter prediction is called P or B frame. For degree of compression, B frame compresses the most, P frame takes the second place and I frame compresses the least. So I frame guarantees the precision of prediction and avoids diffusion of prediction errors, and P frame and B frame is used to meet the demands of real-time coding application. Coding configuration file defines the number of I, P and B frame in a basic coding structure called Group of Pictures (GOP) [3]. According to difference of GOP and other coding parameters, for HEVC reference software HM, encoder defines three kind of configuration: intra, low-delay for real-time and random-access for non-real time application [4]. Hybrid coding structure is applied in low-delay and random-access configuration. In this paper, we mainly focus on simplified method on reference frame management mechanism.

Reference frame management mechanism of video coding goes through lots of improvements. Encoder generally selects the most adjacent restriction frame to be reference frame of current coding frame. However, it is not fit for all the coding structure and application situation, especially application of multi reference frame makes for a huge increase of coding efficiency [5]. The fact that periodically inserting high-quality long-term reference frame contributes to promoting coding performance is proposed in [6], multi- reference frame is proved to be necessary to improve prediction accuracy.

---

[1] †Author Contributions: These authors contributed equally. *Corresponding author

Reference frame selection is modeled as a rate-distortion optimization problem in [7], and then based on reasonable assumption and simplification described in this model, a general situation to this kind of problem is deduced, which is seen as a general method of reference frame setting. "1+x" method for low-delay coding configuration is proposed in [8], "1" stands for a most adjacent reference frame, "x" stands for x frames of high-quality reference frame, this method saves about 3.6% bitrate and accepted by HEVC reference software HM.

In this paper, a simplified reference frame configuration method is proposed to cut down on encoding time in real-time configuration. Configuration file defines encoding structure of the whole coding sequence, QP offset that configuration file assign to each frame reflects its prediction precision when encoder works on this frame to some extent [9]. Our experiment on all the reference frames and cut down on some reference frame selection rules of original HEVC with significant reduction on encoding time and reliable bitrate increase. Then depth information of all the reference frames is used in our judgement to determine whether there is one more frame to be removed in reference frame list. It is proved that reference frame with higher QP offset is more probable to be removed.

The rest of the paper is organized as follows: an overview of RFS mechanism and depth definition in HEVC is presented in Section 2. The proposed simplified reference frame selection method and its depth judgement process are given in Section 3. Experiments are provided in Section 4 to verify the proposed method. Finally, we draw the concluding remarks in Section 5.

2. The Related Work of RFS

Modern mainstream video coding mainly uses multiple reference frame technique in reference frame management which definitely influences coding efficiency [10]. Encoder traverses the entire reference frame in motion estimation. It improves the precision of prediction. On the other hand, calculations of encoding significantly increase.

This paper aims to find out a simplified method for real-time encoding configuration which is named "low-delay" among HEVC configuration files. In this configuration, only the first frame in a sequence of frames is encoded by intra configuration as an "I" frame, and its subsequent frames are all encoded as P frame and generalized B frame. To take full advantage of the high efficiency of "bi-directional" prediction, HEVC introduces generalized B frame, its defines two reference frame lists which store the same reference frames in low-delay configuration and differs a lot from each other in random-access configuration.

HEVC uses Reference Picture Set (RPS) technique to manage the frames in Decoded Picture Buffer (DPB). The reference frames of RPS are classified into two groups: short-term reference frame set, and long-term reference frame set [11]. Short-term reference frame set can be classified into five subsets. Reference of these subsets is classified into three parts: RefPicSetStCurrBefore, RefPicSetStCurrAfter and RefPicSetStFoll. Picture Order Count (POC) is defined to denote display order of frames in video coding. RefPicSetStCurrBefore contains the reference frames whose POC is before current coding frame, and RefPicSetStCurrafter contains the reference frames whose POC is after current coding frame. RefPicSetStFoll contains the frames that are not used by current frame. And long-term reference frame set is configured in slice header and not divided into subsets.

Two reference list "list0" and "list1" are created to be used in RFS through two processes: initialize reference lists and modification [12]. The first reference list list0 selects reference frames from subset of short-term reference frame set RefPicSetStCurrBefore, RefPicSetStCurrAfter and RefPicSetLtCurr in order, until the number of reference frames selected into list0 reaches configured limitation. And the

second reference list list1 selects reference frames from RefPicSetStCurrAfter, RefPicSetStCurrBefore and RefPicSetLtCurr in order. If the number of frames in subset RefPicSetStCurrAfter is set to 0, list0 and list1 are the same, as generalized B frame defines. Modification for reference list is not essential and flags exit in bit stream to indicate whether modification should be performed or not. Modification process can adjust the position of reference frames as expected.

3. Proposed Simplified Reference Frame Selection Method

To reduce computational complexity of encoder, statistics of reference frames distribution of pictures in GOP are made in the "1+3" coding structure of low-delay configuration. The flowchart of proposed simplified RFS method is given in Figure1.
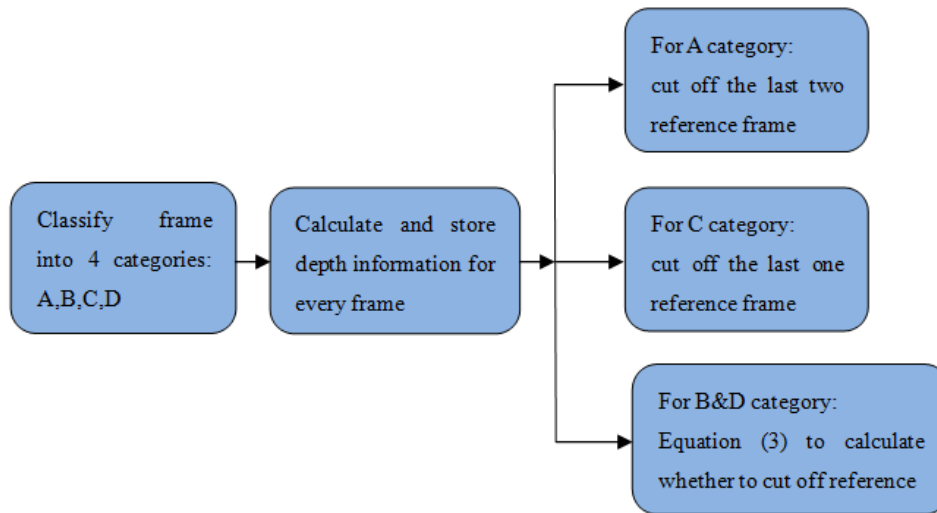

Figure 1. The Flowchart of Proposed Simplified RFS Method

3.1. Simplified Low-Delay Configuration Method

Low-delay configuration includes two kind of encoding structure: IBBB and IPPP, for generalized B frame and P frame, respectively. IBBB encoding structure can be described in Table. 1. POC of reference picture in Table. 1 indicates the POC of reference frame of current encoding frame. The position of reference frames and GOP structure correlations of encoding frame and reference frame is shown in Figure2. The first frame whose POC is 0 is processed by intra prediction as an "I" frame. Every bar's height corresponding to every frame indicates the value of QP offset when encoding this frame [13]. QP offset can be 1,2,3,4 corresponding to different height.
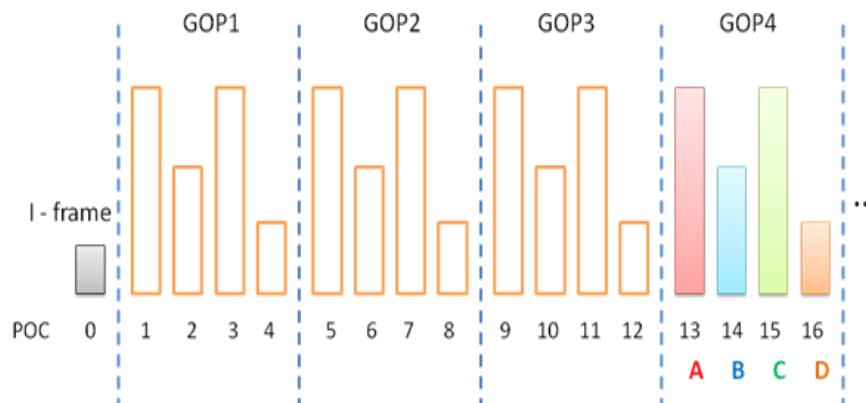
Figure 2. GOP Structure of Low-Delay Configuration in HEVC

Table 1. Low-Delay Configuration IBBB Structure Reference Frame Setting for GOP4

POC of current encoding frame	POC of reference frame
13	12、8、4、0
14	13、12、8、4
15	14、12、8、4
16	15、12、8、4

The four frames in GOP4 select their reference frames utilizing a fixed "1+3" structure, meaning one frame which is most adjacent to current frame and three high-quality frames whose QP offset is the least in their respective GOP, usually located in the end of forward GOP as general low-delay configuration defines [14].

Four pictures of a GOP can be classified into four parts denoted by A, B, C, and D. Four reference frames of every kind of picture are denoted by R0, R1, R2, R3, respectively. The percentage of current frame encoded by intra prediction is denoted by Pintra. The percentage of current frame selecting reference frame labelled Ri is denoted by Pi. It is obviously that,

$$Pintra+P0+P1+P2+P3=1 \qquad (1)$$

Statistics experiments are performed in HEVC reference software HM, version 13, low-delay configuration with QP equal to 27. First 113 frames of 10 Test sequences are selected as experimental samples. Reference frames in GOP start to possess universality from the fourth GOP, with POC starting from 12.

Statistics of reference frames distribution are shown in Table.2. Average percentage distribution can be described in Figure3.

Table 2. Percentage Distribution of Reference Frame Combinations for Pictures of Class A

combination test sequence	Pintra+ P0 (%)	Pintra+ P0+ P1 (%)	Pintra+ P0+ P2 (%)	Pintra+ P0+ P3 (%)	Pintra+ P0 + P1+ P2 (%)
BasketballDrive	97.26	99.02	98.34	97.47	99.35
RaceHorses	93.41	97.15	94.28	94.51	97.20
BasketballDrill	90.29	95.26	92.08	93.49	97.34
BasketballPass	97.33	98.55	97.00	96.07	98.79
BlowingBubbles	96.86	98.19	97.51	96.41	98.90

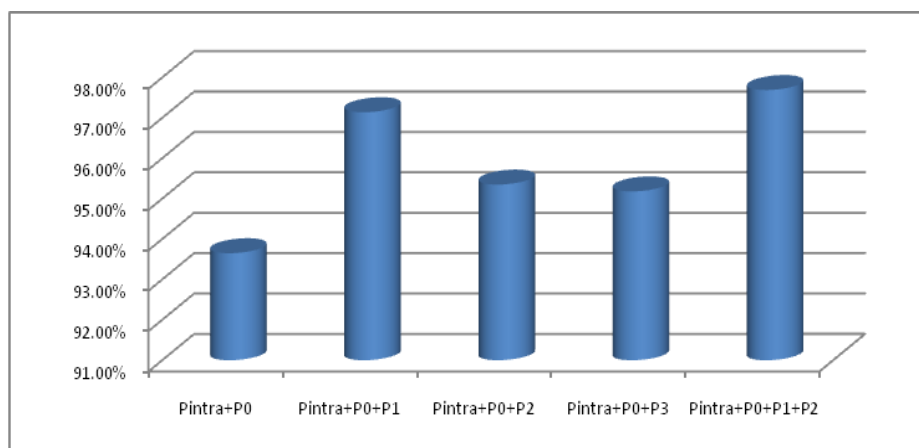|				
BQSquare	91.30	97.67	94.26	92.62	97.71
FourPeople	91.43	95.90	98.04	96.35	96.12
Johnny	91.67	94.45	93.22	93.57	94.63
SlideShow	95.93	97.21	96.04	97.05	98.33
Traffic	91.02	97.98	92.75	94.29	98.49
Average	93.65	97.14	95.35	95.18	97.69

**Figure 3. Average Percentage Distribution of Reference Frame
Combinations for Pictures of Class A**

As is shown in Table 3 and Figure 3, when reference frames are set as R0+R1, seven sequences in ten test sequences indicate more than 97% percentage hit rate of reference frame selection, and R0+R1+R2 combination does not significantly increase hit rate of selection with the percentage raise about 0.5%. So it is reasonable to select R0+R1 or R0+R1+R2 as simplified reference frame selection settings for pictures of class A. In our proposed method, R0+R1 are selected considering real-time encoding demand for HEVC.

Furthermore, statistic experiments are also performed on pictures of class B, C, D, and the results indicate that class C presents similar regulations with class A, as they both correspond to the biggest QP offset in GOPs, which means a relatively cursory prediction compared with frames with smaller QP offset. Experimental statistics for pictures of class C are given in Table 3 and Figure 4.

**Table 3. Percentage Distribution of Reference Frame Combinations for
Pictures of Class C**

| combination ⟋ test sequence	Pintra+ P0+ P1 (%)	Pintra+ P0+ P2 (%)	Pintra+ P0+ P3 (%)	Pintra+ P0+ P1+ P2 (%)	Pintra+ P0 + P1+ P3 (%)
BasketballDrive	93.77	83.90	82.56	97.45	96.46
RaceHorses	94.61	91.83	88.30	98.39	94.19
BasketballDrill	90.29	91.06	88.95	91.29	95.38
BasketballPass	95.33	86.62	85.81	98.05	97.37
BlowingBubbles	94.36	87.24	87.35	99.48	95.74
BQSquare	82.68	77.36	73.94	94.19	88.09
FourPeople	95.83	88.40	87.64	99.03	98.54
Johnny	92.19	84.45	85.64	98.80	95.02
SlideShow	93.37	82.57	82.76	97.26	95.99
Traffic	94.27	90.57	83.76	97.11	92.48
Average	92.67	86.40	84.47	97.10	94.93
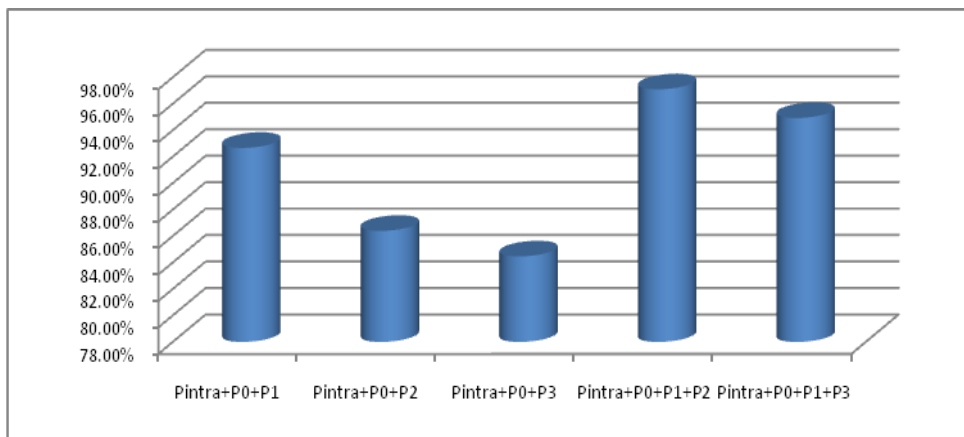
**Figure 4. Average Percentage Distribution of Reference Frame
Combinations for Pictures of Class C**

It proves that the combination of R0+R1+R2 indicate over 97% hit rate of reference selection, far more than other groups, and it is obvious that R0+R1 is not enough to guarantee the precision of prediction with the hit rate less than 93%. Above all, in proposed simplified method, R0+R1 is retained for pictures of class A and R0+R1+R2 is retained for class C. New configuration of RFS in proposed method can be described in Table. 4.

**Table 4. New Configuration of Reference Frame Setting for GOP4 in
Proposed Method**

category	POC of current encoding frame	POC of reference frame(original)	POC of reference frame(proposed)
A	13	12、8、4、0	12、8、(4)
B	14	13、12、8、4	13、12、8、4
C	15	14、12、8、4	14、12、8
D	16	15、12、8、4	15、12、8、4

3.2. Depth Judgment Simplification for High Quality Reference Frame

Each video frame is divided into a sequence of Coding Tree Unit (CTU) to be processed in splitting process of intra and inter prediction. Each CTU can be recursively divided into four CUs with equal size, thus forming a coding quad-tree splitting structure [15]. The CUs size may range from 8x8 to 64x64 luma samples as configuration file defines.
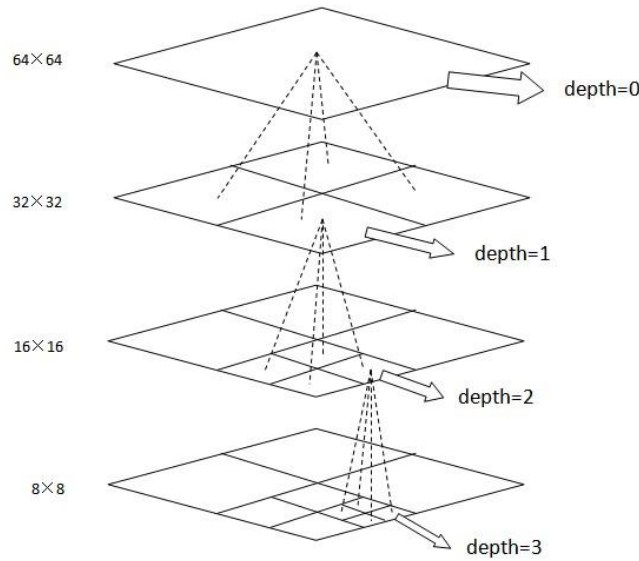
Figure 5. Depth Definition Corresponding to CU Size

CU is basic region of coding process, HEVC defines a variable depth corresponding to each kind size of CU. Depth is 0 when CU size equals CTU, usually is 64x64. As is shown in Figure5, depth plus one as long as CU is divided into four equally sized sub-CUs' until sub-CUs' size equals the minimum size 8x8 corresponding to depth that is equal to 3. Depth of a frame reflects its prediction precision in coding process, providing basis to determine whether a frame is high quality or not together with QP offset assignment regulations [16].

CU splitting depth reflects the prediction quality along with its correlation with QP. Depth of every frame derives from the average depth of all the 4×4 pixels blocks in this frame, which can be calculated as

$$depth_{frame} = \sum_{0}^{M-1} (\frac{\sum_{0}^{N-1} depth_{4\times4}(i)/N}{N})/M$$

(2)

where depth*frame* denotes the depth information of current encoding frame, depth4×4(i) denotes a 4×4 pixels block inside a CU, N denotes the number of 4×4 block in corresponding CTU, M denotes the number of CTUs in corresponding frame.

Simplified setting method for pictures of class B and class D derives from depth relationship of reference frames. Depth of reference frames that are labelled R1, R2, and R3 are selected to be compared with each other. These frames are assigned with the same QP offset as configured.

For frames of class B and class D, three long term reference frames are included in reference list of low-delay configuration. Depth information is utilized to distinguish the quality of frames in reference list. If one of them is calculated by its depth to be not high-quality enough compared to the other two reference frame, this frame is removed from reference list.

Experiments are conducted to determine the threshold used in determining how far current frame which is labelled Ri is away from the average depth of three frames' depth which is denoted by Taway ranging from 0.1 to 0.4 and how close the other two reference frames are close to each other denoted by Tclose ranging from 0.01 to 0.04. Tclose is scaled by variance of average depth of the other two reference frames. Statistics of the percentage of best reference frame still included in simplified reference list are shown in Table. 5.

Table 5. Statistic of the Percentage of Best Reference Frame Included in Simplified Reference List

T_{away}	T_{close}	Hit rate of simplified reference list (%)
0.1	0.01	97.1
0.2		96.3
0.3		74.5
0.4		61.3
0.1	0.02	95.1
0.2		97.5
0.3		84.3
0.4		77.2
0.1	0.03	92.1
0.2		89.3
0.3		75.4
0.4		53.9
0.1	0.04	79.2
0.2		66.4
0.3		60.5
0.4		56.1

When Taway equals 0.2 and Tclose equals 0.02, simplified reference list shows the best hit rate of including the original best reference frame. So specific criterion is made for simplification for pictures of class B and class D, when their depth correlations meet the rule, specific reference frame labelled Ri can be removed from original method. The criterion C is given by

$$C = c1 \,\&\, c2 \tag{3}$$

where $c1$ is given by

$$depth_{Ri} < \frac{depth_{Rj} + depth_{Rk} + depth_{Ri}}{3} - 0.2 \tag{4}$$

And $c2$ is calculated as

$$(\frac{depth_{Rj}^2 + depth_{Rk}^2}{2}) - [\frac{depth_{Rj} + depth_{Rk}}{2}]^2 < 0.02 \tag{5}$$

where depthRi denotes the depth of reference frame which is being calculated, i values from 1 to 3, denoting corresponding frame of current reference list when calculating. Rj and Rk denotes the other two reference frame in the same reference list besides Ri, depthRj and depthRk denote depth of the two reference frame, respectively.

If depth information meets specific criterion given in equation(3), Ri is considered as a relatively cursory reference compared with the other two reference frame and immediately removed from original reference frame set, as the number of reference frames of class B and class D pictures is cut down from 4 to 3, impressively saving encoding time.

4. Experimental Results

In this section, the performance of the proposed simplified reference frame selection method is evaluated in terms of the difference in computational time Average Saving Time Percentage (ASTP), Bjontegaard Delta bit rate (BDBR) and BD-PSNR (dB) [17]. The performance gain or loss is measured with respect to the HEVC reference software

platform (HM 12.0) [18]. The experiments are implemented based on the "encoder_lowdelay_main" (LB-Main) settings [19] as stipulated by the common condition which was proposed in [20]. The quantization parameters values are set to 22, 27, 32 and 37 for a wide range of qualities and bit-rates. The simulation is performed on a computer with an Intel Core i7-2600 3.4GHz processor. The time savings are by a customized variable ASTP calculated as

$$ASTP \; = \; \frac{Enc.Time_{HM\,12.0} - \; Enc.Time_{Proposed}}{Enc.Time_{HM\,12.0}} \times 100\%$$

(6)

where Enc.TimeHM12.0 denotes the average encoding time through all the QP value of original HEVC Reference Software HM12.0, and Enc.TimeProposed denotes encoding time of the proposed simplified reference frame selection method. Table. 6 shows the results of single test sequence base on the "LB-Main" setting with proposed method applied.

Table 6. Proposed Method Experimental Results of Single Sequence in LB-Main Configuration

variable / test sequences	Increase of BDBR of every component (%)			ASTP (%)
	Y	U	V	
BasketballDrive	0.4	0.9	0.7	9.3
RaceHorses	0.1	0.6	0.3	12.9
BasketballDrill	0.9	1.3	0.8	10.7
BasketballPass	0.2	0.4	0.5	9.2
BlowingBubbles	0.3	0.2	0.2	9.9
BQSquare	0.4	−0.9	−0.2	8.2
FourPeople	0.3	0.8	0.0	11.0
Johnny	0.3	−0.3	−0.4	12.1
SlideShow	0.2	0.1	0.1	11.8
Traffic	1.1	1.0	0.7	10.3
Average	0.42	0.41	0.27	10.54

Experiments are also performed on different kinds of sequences with different specific definition. The range of definition from 416×240 to 2560×1600, and its average increase of BDBR, average ASTP and total average are shown in Table. 7.

Experimental results indicate that proposed method brings more than 10% reduction on encoding time. And the BDBR increase of all the color components are limited no more than 0.5%. Proposed method shows great adaptability and stability referring to the effect that experiment performed on all kinds of video sequences with different definition, with impressive reduction on coding time and negligible BDBR increase in average.

Table 7. Proposed Method Experimental Results on Different Definition Video Sequences in LB-Main Configuration

test variable / sequence definition	Average increase of BDBR of every component (%)			ASTP (%)
	Y	U	V	
416×240	0.64	0.59	0.73	9.5
832×480	0.60	0.82	0.34	13.2
1280×720	0.81	0.31	0.55	10.1
1920×1080	0.23	0.17	0.09	11.3

| 2560×1600 | 0.77 | 0.82 | 0.72 | 8.5 |
| Total Average | 0.61 | 0.54 | 0.49 | 10.52 |

5. Conclusion

In this paper, a simplified reference frame selection method is proposed. The implicit correlation between the prediction precision of reference frame settings and splitting depth along with QP is investigated. Reliable reference frame is cut off from original RFS method for these frames with the biggest QP offset which indicate relatively cursory prediction. And depth information is calculated and stored for every frame, and is utilized to make judgement for reference frames with smaller QP offset. The experimental results show that the proposed method reduces the encoding time significantly and has a negligible loss of prediction precision at the same time.

Acknowledgement

References

[1] Ahmad, X. Wei, Y. Sun and Y. Zhang, "Video transcoding: an overview of various techniques and research issues", IEEE Trans. Multimedia, vol. 7, **(2005)**, pp. 793-804.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Transaction on Circuits and Systems for Video Technology, vol. 13, no. 7, **(2003)**, pp. 560-576.

[3] G. J. Sullivan, J. R Ohm, W. J Han and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard", IEEE Transaction on Circuits and Systems for Video Technology, vol. 22, no. 12, **(2012)**.

[4] "High efficiency video coding (HEVC/H.265)", ITU-T Telecommunication Standardization Sector of ITU, **(2013)**.

[5] F. Bossen, B. Bross, K. Shring and D. Flynn, "HEVC complexity and implementation analysis", IEEE Transaction on Circuits and Systems for Video Technology, vol. 22, no. 12, **(2012)**, pp. 1684-1695.

[6] V. Chellappa, P. C. Cosman and G. M. Voelker, "Dual frame motion compensation with uneven quality assignment, Circuits and Systems for Video Technology", IEEE Transactions on, vol. 18, no. 2, **(2008)**, pp. 249-256.

[7] H. L, B. Li and J. Xu, "Rate-distortion optimized reference picture management for high efficiency video coding", IEEE Transaction on Circuits and Systems for Video Technology, vol. 22, no. 12, **(2012)**, pp. 1844-1857.

[8] B. Li, J. Xu, F. Wu and Q. Li, "Encoding optimization to improve coding efficiency for low delay cases", JCTVC-F701, Torino, IT, **(2011)**.

[9] Z. X Chen and S. S Goto, "A QP and partition-size statistic based fuzzy algorithm for fast inter&intra mode decision in video coding", The 2th International Congress on Image and Signal Processing, **(2009)**.

[10] L. Bin, "Optimized reference frame selection for video coding by cloud", IEEE the 13th International Workshop on Multimedia Signal Processing (MMSP), **(2011)**.

[11] W. J. Han, J. Min, I. K. Kim, E. Alshina, A. Ashin, T. Lee, J. Chen, V. Seregin, S. Lee, Y. M. Hong, M. S. Cheon, N. shlyakhow, K. McCann, T. Davies and J. H. Park, "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools", IEEE Trans. Circuits Syst. Video Technology, vol. 20, no. 12, **(2010)**, pp. 1709-1720.

[12] K. Ugur, K. Andersson, A. Fuldseth, G. Bjontegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Rusert, J. Samuelsson, R. Sjoberg and Z. Wu, "High performance, low complexity video coding and the emerging HEVC standard", IEEE Trans. Circuits Syst. Video Technology, vol. 20, no. 12, **(2010)**, pp. 1688-1697.

[13] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan and T. Wiegand, "Comparison of the coding effciency of video coding standards including high efficiency video coding (HEVC)", IEEE Transaction on Circuits and Systems for Video Technology, vol. 22, no. 12, **(2012)**.

[14] L. C. Soon, "Thet NSM, Reference lists for low delay settings", JCT-VC Document, JCTVC-F433, Torino, Italy, **(2011)**.

[15] P. Helle, S. Oudin and B. Bross, "Block merging for quadtree-based partitioning in HEVC", IEEE Transactions on Circuits and Systems for Video Technology, **(2012)**.

[16] J. Xiong, "Fast coding unit selection algorithm for high efficiency video coding intra prediction", Optical Engineering, vol. 52, no. 7, **(2013)**, pp. 071 504-071 504.

[17] G. Bjontegaard, "Calculation of average PSNR differences between RD curves", in No. ITU-T SC16/Q6, VCEG-M33, Austin, TX, USA, **(2001)**.

[18] "JCT-VC HEVC reference software version HM 12.0", available online at https://hevc.hhi.fraunhofer.de/svn/Software/tags/HM-12.0.

[19] "JCT-VC, HEVC reference software manual", JCTVC-J0470, JCT-VC Meeting, Stockholm, Sweden, **(2012)**.

[20] F. Bossen, "Common HM test conditions and software reference configurations", document JCTVC-L1100, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), **(2013)**.

Authors

Ping Kuang, Associate professor in University of Electronic Science and Technology of China(UESTC). In 2000, 2003 and 2006, he obtained his Bachelor, Master and Ph.D. in UESTC, respectively. His interests include image processing and cloud computing.

Jianhua Deng, he graduated in information security from the University of Electronic Science and Technology of China (UESTC), China, in 2006. After graduated, he joined the School of Computer Science and Engineering at UESTC as a staff. From 2009 to 2014, he was a Ph.D. student in Dublin Institute of Technology (DIT) in Ireland and received Ph.D. degree in electrical engineering from DIT in 2014. Now, he is a lecturer in the School of Information and Software Engineering at UESTC. His research activities are focused on wireless communication and video streaming.