JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A Fast Inter-prediction Mode Decision Algorithm for HEVC Based on Spatial-Temporal Correlation

Weixin Yao[1] and Dan Yang[2,*]

Abstract

Many new techniques have been adopted in HEVC (High efficiency video coding) standard, such as quadtree-structured coding unit (CU), prediction unit (PU) partition, 35 intra-mode, and so on. To reduce computational complexity, the paper proposes two optimization algorithms which include fast CU depth range decision and fast PU partition mode decision. Firstly, depth range of CU is predicted according to spatial-temporal correlation. Secondly, we utilize the depth difference between the current CU and CU corresponding to the same position of adjacent frame for PU mode range selection. The number of traversal candidate modes is reduced. The experiment result shows the proposed algorithm obtains a lot of time reducing, and the loss of coding efficiency is inappreciable.

Keywords

Inter-prediction, Mode Decision, Spatial-Temporal Correlation

1. Introduction

High efficiency video coding (HEVC) is approved by the Joint Collaborative Team on Video Coding which is composed of ITU Video Coding Experts Group and ISO/IEC Moving Picture Experts Group. HEVC can increase compression efficiency of 1080p video by about 50%. HEVC not only increases quality of video, but also saves a lot of bandwidth. We can enjoy higher quality 4K video, 3D Blu-ray, high-definition television program. Many new coding structures and coding technologies are adopted in HEVC compared with H.264, such as quadtree segmentation structure, 35 intra-prediction, advanced motion vector prediction, and so on. HEVC still adopts block-based hybrid coding framework. Inter-prediction makes use of temporal correlation to reduce temporal redundancy. First, we perform motion estimation to search for a block region which produces optimum correspondence to the current macroblock from reference frame. Then, the different information between matching block and current block is quantized and transformed, which leads to a entropy-coded and temporal-redundancy-reduced result. HEVC adopts coding tree unit (CTU) which is split into quadtree structure. CTU is divided into coding units (CUs). The CU size ranging from 8×8 to 64×64. The CU sizes of 8×8, 16×16, 32×32, 64×64 correspond to depth3, depth2, depth1, and depth0. Each CU is divided into prediction units (PUs). There are eight kinds of PUs for inter-prediction, including "N×N, N×2N, 2N×N, 2N×2N, nR×2N, nL×2N,

2N×nD, and 2N×nU," Fig. 1 shows PU partition mode. To select the best CU size, CU is divided into eight PU types at each depth level [1,2]. All the CUs from 8×8 to 64×64 and prediction modes are implemented to decide the best depth level. Inter-prediction takes up most of the time in the whole encoding process. We can improve encoding speed from both CU and PU to decrease the time spending during the course of inter prediction. Image processing technologies and statistical techniques are adopted to analyze the CU size range.
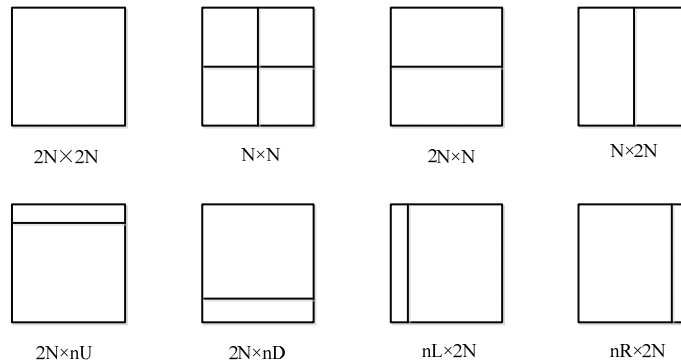
Fig. 1. PU partition mode.

In [3], a fast inter-mode decision algorithm (FIMDA) analyzes the prediction mode distribution at each depth level, the motion vector and rate distortion cost of neighboring CUs select adaptively inter modes. The proposed efficient inter-prediction mode (EIPM) algorithms compute the priority of all inter-prediction modes. Motion estimation is executed in the high priority inter-prediction modes [4]. Lee et al. [5] propose fast CU size decision algorithms include early CU termination (ECUT), CU skip estimation (CUSE) and skip mode decision (SMD). In [6], a fast CU encoding algorithm is proposed. Spatio-temporal encoding parameters are used to decide CU partition. An algorithm by Huang et al. [7] adopts edge information of the partition to reduce computational complexity, and is especially suitable for simple texture and smooth movement. In [8], a fast reference frame selection algorithm makes use of content correlation between the sub-PU and parent PU to decrease computational burden. In [9], the algorithm utilizes statistical analysis to predict the rate distortion cost and bit cost, which skip unnecessary computation. Pareto-based method adopts fine- and medium-granularity encoding time to precisely set a limit to encoding time [10]. Efficient mode decision algorithm consists of asymmetric motion partition (AMP), symmetric motion partition (SMP), and square motion partition [11]. The proposed algorithm adopts CU classification technique to fast select CU size in [12]. The above methods take advantage of temporal-spatial correlation in the process of inter-mode decision. However, these methods also maintain computational complexity. Inter-prediction mode decision is the driving force of our research.

In the paper, the proposed algorithm includes fast CU depth range decision and fast PU partition mode decision. We can pre-estimate depth range to skip unnecessary CU depth level decision according to spatial-temporal correlation. We make use of depth difference between the co-located CU and the current CU to select PU partition mode range. The proposed algorithm can save about 50% of encoding time, while the video quality remains approximately the same.

In the following sections, the fast decision algorithm is introduced to analyze the inter-prediction mode.

A fast CU depth range decision (FCDRD) is proposed in Section 2. In Section 3, temporal correlation of PU partition mode is verified. A fast PU partition mode decision (FPPMD) is proposed. In Section 4, the overall algorithm combines FCDRD with FPPMD. Section 5 describes simulation results.

2. Fast CU Depth Range Decision

There are four cases of CU depth level, which are labelled by zero, one, two, and three. The probability that CU depth level of 0 or 1 is high for the smooth region, while the probability selection of 2 or 3 is high for complicated motion region [13]. We can pre-estimate depth range to skip unnecessary CU depth level decision. In [14], adaptive CU depth range makes use of depth level of the left and upper macroblocks to pre-estimate depth level range of the current macroblock. However, it does not save much time. The relationship of the current CU and its neighboring and co-located CU is heavy, as shown in Figs. 2 and 3. The paper proposes the algorithm which makes use of spatial and temporal correlation to predict depth range. D_{pre} is defined as follows:

$$D_{pre}=\omega_L D_L+\omega_U D_U+\omega_{U\text{-}L} D_{U\text{-}L}++\omega_{U\text{-}R} D_{U\text{-}R}+\omega_P D_P \tag{1}$$

where D represents the CU depth level, L, U, $U\text{-}L$, $U\text{-}R$, and P represent left, up, up-left, up-right, and the co-located position of the neighboring frame, respectively. To obtain more accurate prediction effect, rate distortion (RD) cost from spatial neighboring CUs and co-located CU is considered. An adaptive weight ω is proposed, as shown in formula (2). If the neighboring coding units of the current coding units is unavailable, D_{pre} is set to 0 as starting value. We make use of D_{pre} to select candidate depth level, as shown in Table 1.

$$\omega_L=\frac{RD_L}{RD_L+RD_U+RD_{U\text{-}L}+RD_{U\text{-}R}+RD_P}$$

$$\omega_U=\frac{RD_U}{RD_L+RD_U+RD_{U\text{-}L}+RD_{U\text{-}R}+RD_P}$$

$$\omega_{U\text{-}L}=\frac{RD_{U\text{-}L}}{RD_L+RD_U+RD_{U\text{-}L}+RD_{U\text{-}R}+RD_P} \tag{2}$$

$$\omega_{U\text{-}R}=\frac{RD_{U\text{-}R}}{RD_L+RD_U+RD_{U\text{-}L}+RD_{U\text{-}R}+RD_P}$$

$$\omega_P=\frac{RD_P}{RD_L+RD_U+RD_{U-L}+RD_{U-R}+RD_P}$$
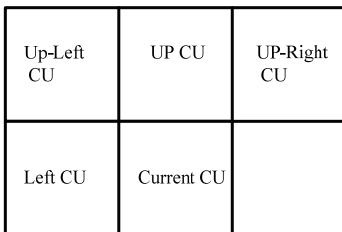
Fig. 2. Neighboring CUs of the current CU.
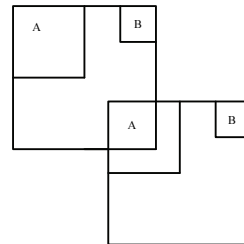
Fig. 3. Corresponding CU of neighboring frame.

Table 1. Candidate depth of the CU

	$D_{pre} = 0$	$D_{pre} \in (0, 1)$	$D_{pre} \in [1, 2)$	$D_{pre} \in [2, 3]$
Candidate depth level	0	0,1	1,2	2,3

3. Fast PU Partition Mode Decision

The current CU and the co-located CU are alike. The current CU depth is in accordance with depth of the co-located CU, and then PU partition mode (PPM) range of the current CU may be determined by PPM of the co-located CU. Some unnecessary PPMs are skipped and computational complexity is reduced. For purpose of verifying temporal correlation, Traffic, BQTerrace, BQMall, and BQSquare are tested with QP=32. The first 2 seconds of each test sequences are adopted at the experiment.

The medial probability of PU with "2N×2N, 2N×N, N×2N, N×N, 2N×nU, 2N×nD, nL×2N, nR×2N" is 97.90%, 0.66%, 0.84%, 0.06%, 0.14%, 0.11%, 0.16%, 0.13% from Table 2, when 2N×2N is decided as the optimal PPM at the co-located CU. It is that the total probability selection of "2N×2N, 2N×N, N×2N" is up to 99.4%. So when "2N×2N" is decided as the optimal PPM at the co-located CU, we only perform RD cost calculations on the three modes of "2N×2N, 2N×N," and "N×2N." RD cost of the current CU are calculated by the PPM of "2N×2N, 2N×N, N×2N, 2N×nU, 2N×nD" and the rest PPMs are skipped when the PPM of co-located CU is "2N×N." When the PPM of co-located CU is "N×2N," the candidate PPMs is "2N×2N, 2N×N, N×2N, nL×2N, nR×2N" at the current CU. The candidate PPMs of the current CU are "2N×2N, 2N×N, N×2N, N×N," when the PPM of co-located CU is "N×N." When the PPM of co-located CU is "2N×nU," the candidate PPMs are "2N×2N, 2N×N, N×2N, 2N×nU" at the current CU. The candidate PPMs of the current CU are "2N×2N, 2N×N, N×2N, 2N×nD," when the PPM of co-located CU is "2N×nD." When the PPM of co-located CU is "nL×2N," the candidate PPMs are "2N×2N, 2N×N, N×2N, nL×2N" at the current CU. The candidate PPMs of the current CU are "2N×2N, 2N×N, N×2N, nR×2N," when the PPM of co-located CU is "nR×2N." We make use of temporal correlation to decrease the number of the PPMs and complexity of PPM decision.

Table 2. Average probabilities of PU partition mode of the current CU

Best PU mode of the co-located CU	PPM of the current CU (%)							
	2N×2N	2N×N	N×2N	N×N	2N×nU	2N×nD	nL×2N	nR×2N
2N×2N	97.90	0.66	0.84	0.06	0.14	0.11	0.16	0.13
2N×N	93.68	2.40	2.22	0.25	0.35	0.50	0.34	0.26
N×2N	94.39	1.83	2.94	0.34	0.31	0.19	0.42	0.48
N×N	95.78	0.97	2.13	0.26	0.24	0.20	0.18	0.24
2N×nU	93.39	2.25	2.14	0.31	0.85	0.37	0.35	0.34
2N×nD	94.23	2.26	1.90	0.16	0.26	0.62	0.34	0.23
nL×2N	92.67	1.63	3.42	0.28	0.23	0.24	0.96	0.57
nR×2N	92.97	1.52	3.40	0.22	0.34	0.28	0.48	0.79

4. The Overall Algorithm

Firstly, depth range of the current CU is calculated by formula (1). Secondly, we calculate depth

difference between the co-located CU and the current CU. If the depth difference is 0, PPM range is selected at the current CU by the PPM of the co-located CU. If the depth difference is 1 or 2, PPM decision is performed by HM original algorithm. If the depth level difference is 3, PPM traversed by the current CU are skipped. Fig. 4 shows flow diagram of the whole algorithm.
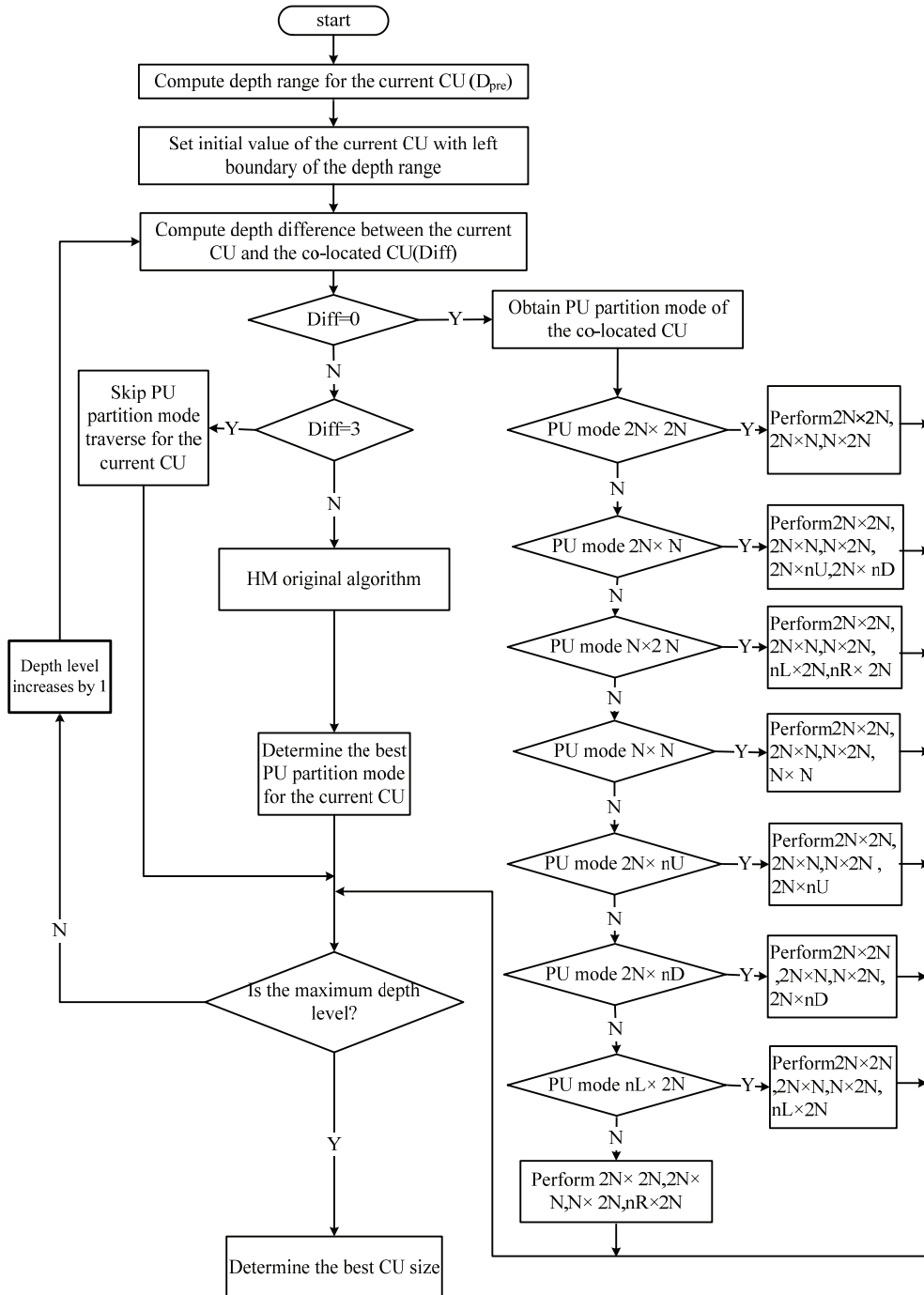

Fig. 4. Inter-prediction mode decision algorithm.

5. Experimental Results

In order to verify the property of the whole algorithm, FCDRD and FPPMD are performed on HM14.0 reference software. HEVC standardization has been performed for two different scenarios, random access (RA) and low delay (LD). Entertainment applications are supported by random access. Interactive applications are restrained by low delay. Table 3 shows the test sequences are used under both RA and LD conditions. Each test sequence is performed with QP=37, 32, 27, and 22. The experimental results are measured in terms of Bjontegaard delta peak signal-to-noise ratio (BDPSNR), Bjontegaard delta bitrate (BDBR), and time saving (TS). TS represents the saved coding time compared to HM14.0. A positive or negative value of TS indicates the increase or decrease of time.

$$TS(\%) = \frac{T_{pro} - T_{HM}}{T_{HM}} \times 100\% \tag{3}$$

Table 3. Parameters of test sequences

Sequence	Resolution	Frame rate (fps)
Traffic	2560×1600 (A)	30
PeopleOnStreet	2560×1600 (A)	30
Kimonol	1920×1080 (B)	24
ParkScene	1920×1080 (B)	24
Cactus	1920×1080 (B)	50
BQTerrace	1920×1080 (B)	60
BasketballDrill	832×480 (C)	50
BQMall	832×480 (C)	60
PartyScene	832×480 (C)	50
BasketballPass	416×240 (D)	50
BlowingBubbles	416×240 (D)	50
BQSquare	416×240 (D)	60
SlideShow	1280×720 (F)	30
SlideEditing	1280×720 (F)	20
ChinaSpeed	1024×768 (F)	30
BasketballDrillText	832×480 (F)	50

The experiment result is shown in Tables 4 and 5. The FCDRD algorithm can save coding time by 31.3%–34.6% on average, while BDPSNR slightly lose 0.07–0.11 dB and BDBR slightly increase 1.00%–1.34% under both RA and LD conditions. The procedures of CU depth decision skip some unnecessary depth levels for test sequences. The SlideEditing sequence saves time by a maximum of 42.7%. The BlowingBubbles sequence saves time by a minimum of 19.8%. The FPPMD algorithm reduces the number of PU candidate modes traversed. Since the low-activity sequences have simple motion characteristics, the number of candidate modes traversed by the low activity sequences is less than that of the high activity sequences. The FPPMD can decrease the coding time of SlideEditing sequence by up to 36.4% (RA) and PeopleOnStreet (RA) sequence by a minimum of 10.2%. The average BDPSNR is decreased by 0.02–0.05 dB and the average BDBR is increased by 0.66%–0.74% under both RA and LD conditions, which are negligible. The proposed overall algorithm which combines the

FCDRD algorithm with the FPPMD algorithm is applied to all test sequences. The proposed overall algorithm can save 41.9% and 44.3% coding time compared with HM14.0, while 0.10–0.16 dB PSNR loss and 1.49%–1.88% bitrate increment can be ignored. Tables 4 and 5 show results of FCDRD, FPPMD, and overall algorithm compared with HM14.0.

Table 4. Comparison results of the proposed algorithm and HM reference software under RA condition

Test sequence	FCDRD			FPPMD			The overall algorithm		
	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)
Traffic	-0.07	1.10	-38.1	-0.06	0.84	-31.1	-0.13	1.78	-54.2
PeopleOnStreet	-0.05	0.70	-25.4	-0.04	0.95	-10.2	-0.09	1.64	-27.7
Kimonol	-0.09	0.56	-32.7	-0.02	0.78	-23.8	-0.08	1.32	-40.7
ParkScene	-0.05	1.21	-38.3	-0.04	0.67	-29.2	-0.09	1.65	-50.3
Cactus	-0.07	0.80	-39.5	-0.03	0.59	-26.1	-0.10	1.23	-47.1
BQTerrace	-0.04	0.76	-42.4	-0.04	0.34	-28.4	-0.08	0.96	-53.2
BasketballDrill	-0.07	2.01	-31.7	-0.04	0.53	-24.2	-0.11	2.15	-41.4
BQMall	-0.01	0.85	-31.1	-0.05	0.78	-35.5	-0.06	1.48	-44.8
PartyScene	-0.06	0.87	-35.6	-0.03	0.86	-26.7	-0.09	1.56	-45.6
BasketballPass	-0.12	1.98	-33.7	-0.05	0.45	-20.5	-0.17	2.23	-38.9
BlowingBubbles	-0.07	0.85	-23.4	-0.04	0.37	-18.9	-0.11	1.06	-32.8
BQSquare	-0.12	2.64	-29.6	-0.08	1.02	-29.0	-0.20	3.34	-35.6
SlideShow	-0.11	1.98	-41.2	-0.07	0.80	-32.3	-0.18	2.45	-55.1
SlideEditing	-0.16	2.01	-45.4	-0.10	1.03	-36.4	-0.26	2.87	-59.7
ChinaSpeed	-0.29	1.36	-32.6	-0.08	0.83	-20.2	-0.37	2.01	-40.4
BasketballDrillText	-0.33	1.75	-32.2	-0.06	0.98	-32.5	-0.39	2.34	-41.3
Average	-0.11	1.34	-34.6	-0.05	0.74	-26.6	-0.16	1.88	-44.3

**Table 5.** Comparison results of the proposed algorithm and HM reference software under LD condition

Test sequence	FCDRD			FPPMD			The overall algorithm		
	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)
Traffic	-	-	-	-	-	-	-	-	-
PeopleOnStreet	-	-	-	-	-	-	-	-	-
Kimonol	-0.05	0.24	-28.3	-0.00	0.70	-22.9	-0.06	0.90	-38.7
ParkScene	-0.01	0.98	-33.5	-0.01	0.62	-28.4	-0.02	1.45	-45.6
Cactus	-0.03	0.54	-35.2	-0.00	0.53	-23.3	-0.03	1.03	-39.4
BQTerrace	-0.02	0.34	-38.7	-0.01	0.21	-27.5	-0.03	0.46	-49.4
BasketballDrill	-0.03	1.61	-26.6	-0.02	0.64	-23.1	-0.04	1.95	-42.7
BQMall	-0.01	0.67	-28.5	-0.02	0.76	-31.9	-0.03	1.28	-43.7
PartyScene	-0.03	0.45	-29.7	-0.00	0.73	-24.2	-0.02	1.06	-41.5
BasketballPass	-0.07	1.08	-31.2	-0.02	0.34	-19.6	-0.09	1.23	-36.2
BlowingBubbles	-0.02	0.65	-19.8	-0.02	0.36	-16.9	-0.03	0.86	-30.6
BQSquare	-0.04	1.97	-22.4	-0.04	0.91	-26.3	-0.06	2.54	-34.4
SlideShow	-0.05	1.76	-39.5	-0.05	0.50	-29.2	-0.09	2.14	-49.3
SlideEditing	-0.18	1.47	-42.7	-0.08	1.00	-33.7	-0.26	2.36	-54.8
ChinaSpeed	-0.24	0.96	-31.1	-0.03	0.90	-18.4	-0.34	1.67	-39.5
BasketballDrillText	-0.23	1.29	-30.4	-0.02	0.99	-28.1	-0.24	1.94	-40.2
Average	-0.07	1.00	-31.3	-0.02	0.66	-25.2	-0.10	1.49	-41.9

Figs. 5 and 6 show the coding performance of Kimonol and BQMall using different QPs under both RA and LD conditions. The RD curves between the proposed algorithms and the HM14.0 are approximated. However, the proposed algorithms can save plenty time.
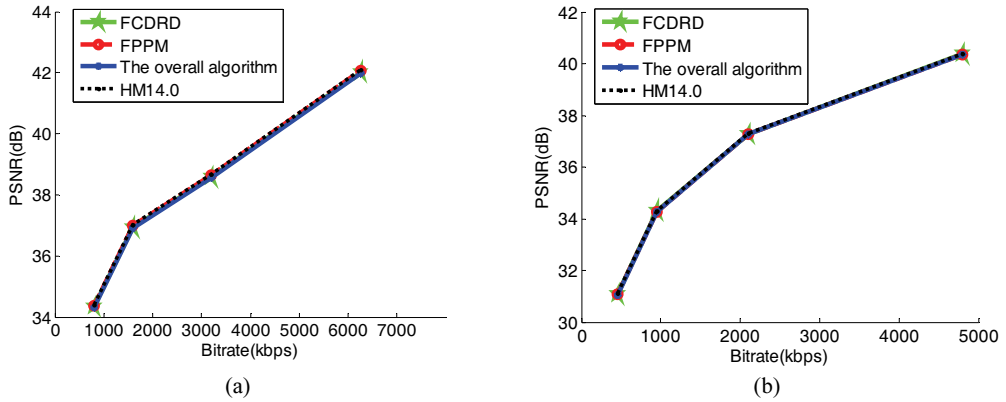
Fig. 5. Result of RD curves of Kimonol and BQMall: (a) Kimonol under RA condition and (b) BQMall under LD condition.
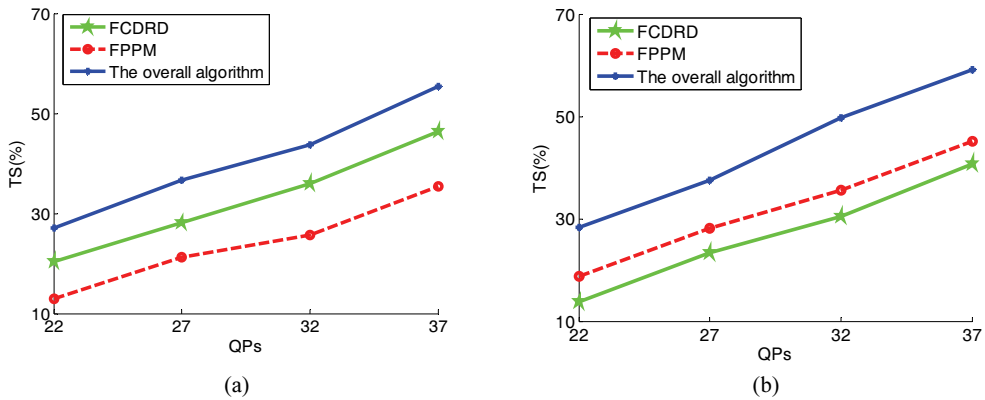
Fig. 6. Time saving curves of Kimonol and BQMall: (a) Kimonol under RA condition and (b) BQMall under LD condition.

6. Conclusion

To reduce computational complexity, the proposed whole algorithm which consists of FCDRD and FPPMD is performed on HM14.0. FCDRD making use of depth level of spatially and temporally adjacent CU to predict depth level range of the current CU. FPPMD skips unnecessary mode selection according to the depth relationship of the corresponding position CU and the current CU. The proposed whole algorithm reduces 41.9%–44.3% of encoding time on average, while RD performance is approximately unchanged. In this paper, we only use the spatial-temporal correlation to obtain prediction mode of CU and PU. In the future work, other motion parameters are considered, so that CU and PU can be predicted more accurately.

Acknowledgement

References

[1] G. J. Sullivan and J. R. Ohm, "Recent developments in standardization of high efficiency video coding (HEVC)," in *Proceedings of SPIE 7798: Applications of Digital Image Processing XXXIII*. Bellingham, WA: International Society for Optics and Photonics, 2010.

[2] G. J. Sullivan and T. Wiegand, "Video compression-from concepts to the H. 264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18-31, 2005.

[3] L. Shen, Z. Zhang, and Z. Liu, "Adaptive inter-mode decision for HEVC jointly utilizing inter-level and spatiotemporal correlations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 10, pp. 1709-1722, 2014.

[4] A. Lee, D. Jun, J. Kim, J. S. Choi, and J. Kim, "Efficient inter prediction mode decision method for fast motion estimation in high efficiency video coding," *ETRI Journal*, vol. 36, no. 4, pp. 528-536, 2014.

[5] J. Lee, S. Kim, K. Lim, and S. Lee, "A fast CU size decision algorithm for HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 411-421, 2014.

[6] S. Ahn, B. Lee, and M. Kim, "A novel fast CU encoding scheme based on spatiotemporal encoding parameters for HEVC inter coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 422-435, 2015.

[7] X. Huang, Q. Zhang, X. Zhao, W. Zhang, Y. Zhang, and Y. Gan, "Fast inter-prediction mode decision algorithm for HEVC," *Signal, Image and Video Processing*, vol. 11, pp. 33-40, 2017.

[8] Z. Pan, P. Jin, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast reference frame selection based on content similarity for low complexity HEVC encoder," *Journal of Visual Communication and Image Representation*, vol. 40, pp. 516-524, 2016.

[9] S. H. Jung and H. W. Park, "A fast mode decision method in HEVC using adaptive ordering of modes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 10, pp. 1846-1858, 2016.

[10] G. Correa, P. A. Assuncao, L. V. Agostini, and L. A. da Silva Cruz, "Pareto-based method for high efficiency video coding with limited encoding time," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1734-1745, 2016.

[11] J. Vanne, M. Viitanen, and T. D. Hamalainen, "Efficient mode decision schemes for HEVC inter prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 9, pp. 1579-1593, 2014.

[12] T. Mallikarachchi, A. Fernando, and H. K. Arachchi, "Fast coding unit size selection for HEVC inter prediction," in *Proceedings of 2015 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2015, pp. 457-458.

[13] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 465-470, 2013.

[14] X. Li, J. An, X. Guo, and S. Lei, "Adaptive CU depth range (JCTVC-E090)," in *the Meeting report of the Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T WP3/16 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, Switzerland, 2011 [Online]. Available: https://www.itu.int/wftp3/av-arch/jctvc-site/2011_03_E_Geneva/JCTVC-E_Notes_d8.doc.

Weixin Yao https://orcid.org/0000-0001-5602-2143

She received the M.S. degree in communication and information system from Gui Zhou University, Guiyang, china, in 2007. She is currently a lecturer with School of Electrical Engineering, Anhui Polytechnic University. Her research interests include video coding and image processing.

Dan Yang https://orcid.org/0000-0003-3898-8627

He received the M.S. degree in computer application technology from Gui Zhou University, Guiyang, china, in 2007. He is currently associate professor with School of computer and information, Anhui Polytechnic University. His research interests include video coding and image processing.