J. Vis. Commun. Image R. 40 (2016) 516–524

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Fast reference frame selection based on content similarity for low complexity HEVC encoder ☆

Zhaoqing Pan [a], Peng Jin [b], Jianjun Lei [b,]*, Yun Zhang [c], Xingming Sun [a], Sam Kwong [d]

[a] School of Computer and Software, Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[d] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 22 July 2015
Revised 12 July 2016
Accepted 21 July 2016
Available online 25 July 2016

Keywords:
Reference frame
Prediction unit
Early termination
HEVC
Video coding

ABSTRACT

The high efficiency video coding (HEVC) is the state-of-the-art video coding standard, which achieves about 50% bit rate saving while maintaining the same visual quality as compared to the H.264/AVC. This achieved coding efficiency benefits from a set of advanced coding tools, such as the multiple reference frames (MRF) based interframe prediction, which efficiently improves the coding efficiency of the HEVC encoder, while it also increases heavy computation into the HEVC encoder. The high encoding complexity becomes a bottleneck for the high definition videos and HEVC encoder to be widely used in real-time and low power multimedia applications. In this paper, we propose a content similarity based fast reference frame selection algorithm for reducing the computational complexity of the multiple reference frames based interframe prediction. Based the large content similarity between the parent prediction unit (Inter_2N × 2N) and the children prediction units (Inter_2N × N, Inter_N × 2N, Inter_N × N, Inter_2N × nU, Inter_2N × nD, Inter_nL × 2N, and Inter_nR × 2N), the reference frame selection information of the children prediction units are obtained by learning the results of their parent prediction unit. Experimental results show that the proposed algorithm can reduce about 54.29% and 43.46% MRF encoding time saving for the low-delay-main and random-access-main coding structures, respectively, while the rate distortion performance degradation is negligible.

1. Introduction

With the development of capture and display technologies, the full high definition (HD) and ultra HD videos are attracting more and more people's attention since they can provide higher perception/video quality. However, with the increased video resolution and frame rate, the data volume of the raw HD videos increases dramatically. It is highly desirable to develop high compression techniques due to the current memory and channel bandwidth are still limited. Under this kind of compression rate demand, the joint collaborative team on video coding (JCT-VC) of the ITU-T video coding experts group (VCEG) and ISO/IEC moving picture experts group (MPEG) has developed a state-of-the-art video coding standard named high efficiency video coding (HEVC) [1–3]. The HEVC can achieve the same subjective visual quality as the H.264/AVC [4] high profile while requiring only about 50% of the bit rate. This obtained coding efficiency benefits from a set of advanced coding tools, such as flexible size unit representation, intraframe prediction with 35 modes, multiple reference frames (MRF) interframe prediction, new in-loop filtering methods, and so on. Meanwhile, the computational complexity of the HEVC encoder increases dramatically as these used coding tools. The high computational complexity becomes a bottleneck for the HD videos and HEVC encoder to be widely used in real-time and low power multimedia applications, such as live video broadcasting, mobile video communication, and video surveillance. Thus, there is a pressing need to reduce the computational complexity of the HEVC encoder.

Recently, many researchers have devoted their efforts on reducing the computational complexity of the HEVC encoder [5–11]. Based on the Bayesian decision theory and rate distortion (RD) characteristics, Lee et al. proposed a fast coding unit (CU) size decision method for the HEVC [5]. In [6], Shen et al. proposed a CU depth decision method based on the depth selection correlation between the spatial-temporal neighboring CUs and the current CU. Besides, they also proposed an early termination for the motion estimation based on the motion homogeneity, RD cost and Skip

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: zqpan3-c@my.cityu.edu.hk (Z. Pan), qustjinpeng@163.com (P. Jin), jjlei@tju.edu.cn (J. Lei), yun.zhang@siat.ac.cn (Y. Zhang), sunnudt@163.com (X. Sun), cssamk@cityu.edu.hk (S. Kwong).

http://dx.doi.org/10.1016/j.jvcir.2016.07.018
1047-3203/

mode. By using the CU depth selection information of spatial neighboring CUs, Kim et al. proposed a CU depth range decision method for the HEVC [7]. In [8], the RD-complexity characteristics of the inter prediction was analyzed and derived an efficient inter mode decision method for the HEVC. Based on the CU motion activity and mode selection correlation among hierarchical depth CUs, Pan et al. proposed an early Merge mode decision method for the HEVC fast interframe prediction [9]. By utilizing the estimated optical flow of the downsampled frames, Xiong et al. proposed a fast inter CU selection method for the HEVC [10]. Based on the prediction mode and RD cost correlations among different quadtree depth levels and spatially neighboring CUs, Shen et al. proposed a fast CU size and intra mode decision method for the HEVC [11]. These methods mainly focus on reducing the computational complexity of the flexible size unit representation technique, the HEVC encoding complexity could be further reduced by optimizing the MRF based interframe prediction.

In the last decade, a number of methods have been proposed to reduce the encoding complexity of the MRF based interframe prediction for the H.264/AVC and its extensions [12–18]. By taking into account the correlation/continuity of motion vectors among different reference frames, Su et al. proposed a fast MRF based motion estimation for the H.264/AVC [12]. Based on the spatial and temporal correlation of the reference frame index and motion vectors, Jun et al. proposed an efficient priority-based MRF selection method for the H.264/AVC fast motion estimation [13]. In [14], Chen et al. proposed a fast MRF based motion estimation for the H.264/AVC by using the stored motion vectors to compose the motion vector without searching all active reference frames. In [15], Liu et al. proposed a fast MRF selection method for the H.264/AVC motion estimation by using the motion activity and Hadamard coefficients. Based on the reference frame selection of the 16×16 mode partition, Zhang et al. proposed an efficient MRF selection method for the H.264 based multiview video coding [16]. In [17], Yeh et al. proposed a fast mode decision based MRF selection for H.264 based multiview video coding system by using inter-view rate distortion prediction method. In [18], by using the inter-view and inter-component correlations based fast mode decision, Lei et al. proposed a low complexity MRF decision method for H.264 based multiview depth video coding. These methods can efficiently reduce the computational complexity, however, they were proposed for the H.264, and are not suitable for directly applying into the HEVC encoder due to the different statistical characteristics and different coding tools used in HEVC encoding system. In [19], according to the motion complexity which is computed by the distribution of the best reference frame, the motion vector difference and its associated average distortion, a fast reference frame selection was proposed. However, the MRF encoding time saving of that method is still limited and unstable for the HEVC with random-access-main coding structure due to using of the IBP prediction structure.

In this paper, we propose a fast MRF selection algorithm for the fast HEVC interframe prediction, which is based on the relationship between the content similarity and the reference frame selection. The rest of this paper is organized as follows. The review on the HEVC MRF encoding process is presented in Section 2. Then, the details of the proposed fast MRF selection algorithm are illustrated in Section 3. Section 4 shows the experimental results. Then, an algorithm discussion is given in Section 5. At last, Section 6 concludes this paper.

2. Review on the HEVC MRF encoding process

As previously video coding standards such as H.264/AVC, the HEVC standard is also a hybrid video encoder. In the HEVC encoding process, each frame is partitioned into a sequence of coding tree units (CTUs), which is the basic unit of coding, and consists of a luma coding tree block (CTB), two chroma CTBs and associated syntax elements of 4:2:0 color sampling. According to the quadtree syntax, the CTU is further split into one or multiple CUs. Then, based on the prediction-type, the CU can be split into one, two, or four prediction units (PUs). The PU is the basic unit of intraframe and interframe prediction. In the HEVC prediction structures, such as IPP and HBP prediction structures, the first frame of each group-of-picture (GOP) is encoded with the intraframe prediction. For the remaining frames of the GOP are encoded by using the interframe prediction. The encoding process of the interframe prediction is to predict the samples of each block using the reference frame and motion vector.

In the HEVC encoder, the MRF based variable block-size interframe prediction highly improves the coding efficiency, while it also extremely increases the coding complexity. Fig. 1 gives an illustration of the variable block-size CU and PU mode decision as well as the MRF selection in the HEVC. There are three difference loop processes for encoding a CTU. The first loop is the CTU quadtree depth decision, which is adopted to determine the size of the CU. In HEVC, it supports the quadtree depth from 0 to 3, and it means the best CU size is chosen by sequential encoding the CU from 64×64 to 8×8. The second loop is the PU partition type decision. The HEVC supports eleven PU modes including Merge/Skip, eight Inter modes (Inter_2N × 2N, Inter_2N × N, Inter_N × 2N, Inter_N × N, Inter_2N × nU, Inter_2N × nD, Inter_nL × 2N, and Inter_nR × 2N), two Intra modes (Intra_N × N, and Intra_2N × 2N). These modes will be checked sequentially, and the mode with the minimum RD cost is selected as the best PU mode. The third loop is the best reference frame selection, which will check all active reference directions and reference frame indexes when encoding one PU mode. The best reference frame decision process consists of two sub-loop processes. The first sub-loop is the reference direction checking over the Forward (List 0), Backward (List 1), and Bi-iterative (Bi) direction. The Lists 0 and 1 store the forward and backward reference frames, respectively. For the Bi-iterative search process, all the reference frames in the Lists 0 and 1 will be checked. The second sub-loop is the reference frame checking over all active reference frames in each reference direction. Finally, the best CU depth, best PU mode, best reference direction and reference frame index, $\{d^*, p^*, \varphi^*, \gamma^*\}$ are determined according to the minimization of the Lagrangian RD cost function [20],

$$\{d^*, p^*, \varphi^*, \gamma^*\} = \arg\min_{d \in \mathbf{D}} \left(\arg\min_{p \in \mathbf{P}} \left(\arg\min_{\varphi \in \{\mathrm{List0,List1,Bi}\}} \times \arg\min_{\gamma} J(\mathbf{O}, \mathbf{R}(d, p, \varphi, \gamma)) \right) \right), \tag{1}$$

where $\mathbf{D}$ is the set of the CU quadtree depths, and $\mathbf{D} = \{0,1,2,3\}$; $\mathbf{P}$ denotes the set of the eleven candidate PU modes, $\mathbf{P}$ = {Merge/Skip, Inter_2N × 2N, Inter_2N × N, Inter_N × 2N, Inter_N × N, Inter_2N × nU, Inter_2N × nD, Inter_nL × 2N, and Inter_nR × N}; $J(\cdot)$ is the Lagrangian RD cost function [20]; φ is the reference direction, $\varphi \in$ {List 0, List 1, Bi}; γ represents the reference frame indexes in the List 0 and List 1; $\mathbf{O}$ is the original CU, and $\mathbf{R}$ denotes the reconstructed CU which is obtained by encoding the original CTU with the depth d, PU mode p, reference direction φ, and reference frame index γ. This "try all then select the best" CU size, PU mode, and reference frame decision method could significantly improve the coding efficiency of the HEVC encoder, while it also could increase the encoding complexity of the HEVC encoder, especially the MRF based motion estimation.

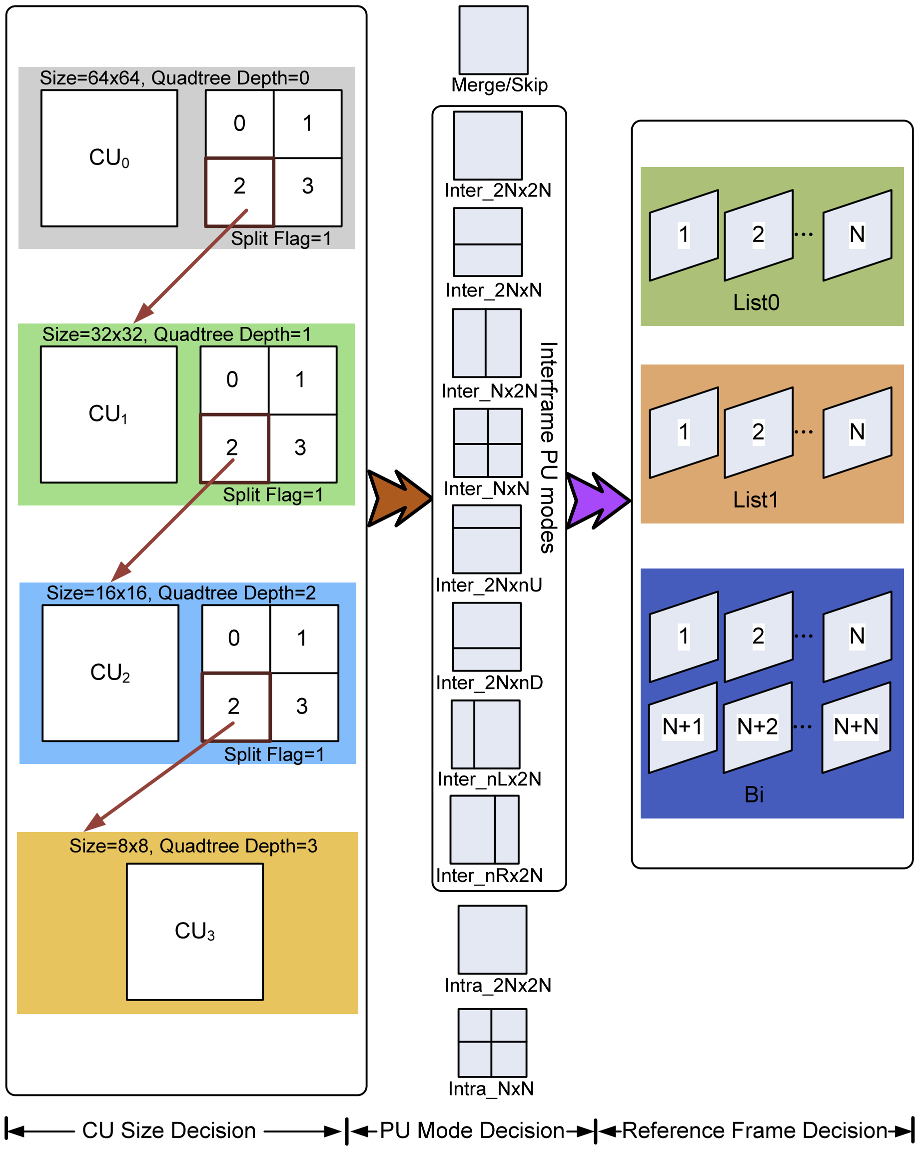

Fig. 1. Illustration of the CU size decision, PU mode decision and reference frame selection in the HEVC.

3. Proposed content similarity based fast MRF selection algorithm

3.1. Encoding complexity analysis on the MRF encoding process

In order to analyze the encoding complexity of the MRF selection process, eight HEVC test sequences (*BQSquare*, *BasketballPass*, *BQMall*, *BasketballDrill*, *FourPeople*, *Johnny*, *Cactus*, and *ParkScene*) with various resolutions and motion activities are encoded by the HM12.0 [21] under the HEVC common test conditions [22]. Four quantization parameters (QPs) (22, 27, 32, and 37) are used. The motion estimation search range and method are 64 and TZSearch, respectively. The low-delay-main and random-access-main coding structures are adopted. The encoding complexity of the MRF selection is computed as

$$MRF_{complexity} = T_{MRF}/T_{All}, \tag{2}$$

where T_{MRF} represents the encoding time consumed by the MRF based motion estimation process; T_{All} denotes the total encoding time. If the value of $MRF_{complexity}$ is large, it means the complexity is high. The statistical results of MRF complexity are shown in Table 1. From Table 1, it can be observed that (1) the complexity of MRF increases as the QP increases, this is because when the transformation and quantization time decrease as the QP increases. (2) The $MRF_{complexity}$ varies as the video content changes. Usually, the complexity of MRF increases as the texture of the video content is simple. For these simple content video sequences, a large amount of potential complexity could be removed. (3) The complexity of MRF is from 0.527 to 0.742, 0.684 on average, for the low-delay-main coding structure; and from 0.542 to 0.720, 0.666 on average, for the random-access-main coding structure. These values indicate that the MRF selection method which tries all reference directions and reference frames then selects the best is with high computational complexity. Hence, if the MRF selection process can be simplified, significant encoding complexity could be removed.

3.2. Proposed fast MRF selection algorithm

In the HEVC encoding process, the CU can be split into one, two, or four PUs based on the prediction-type. For example, a CU with size of $2N \times 2N$ can be partitioned into a PU with size of

Table 1
Statistical results of the MRF complexity.

Class	Sequence	Resolution	Low-delay-main coding structure					Random-access-main coding structure				
			QP = 22	QP = 27	QP = 32	QP = 37	Average	QP = 22	QP = 27	QP = 32	QP = 37	Average
D	BQSquare	416 × 240	0.527	0.620	0.673	0.709	0.632	0.542	0.621	0.668	0.695	0.632
	BasketballPass		0.639	0.681	0.717	0.736	0.693	0.626	0.668	0.697	0.715	0.677
C	BQMall	832 × 480	0.605	0.666	0.704	0.726	0.675	0.604	0.655	0.688	0.708	0.664
	BasketballDrill		0.612	0.668	0.711	0.735	0.682	0.614	0.664	0.698	0.717	0.673
E	FourPeople	1280 × 720	0.662	0.705	0.721	0.728	0.704	n/a[a]	n/a	n/a	n/a	n/a
	Johnny		0.677	0.717	0.731	0.737	0.716	n/a	n/a	n/a	n/a	n/a
B	Cactus	1920 × 1080	0.583	0.669	0.721	0.728	0.675	0.597	0.667	0.696	0.709	0.667
	ParkScene		0.624	0.688	0.724	0.742	0.695	0.625	0.679	0.706	0.720	0.683
	Average		0.627	0.686	0.719	0.733	0.684	0.601	0.659	0.692	0.711	0.666

[a] The "n/a" means the result is not available, because the sequences "FourPeople" and "Johnny" with the random-access-main coding structure are not tested according to the HEVC common test conditions [22].

2N × 2N, two PUs with size of 2N × N, N × 2N, 2N × nU, 2N × nD, nL × 2N, or nR × 2N, or four PUs with size of N × N. For HEVC interframe prediction, all PU partition types are supported, and these interprediction PU modes are Inter_2N × 2N, Inter_2N × N, Inter_N × 2N, Inter_N × N, Inter_2N × nU, Inter_2N × nD, Inter_nL × 2N, and Inter_nR × 2N. Therefore, according to the PU size, the Inter_2N × 2N can be regarded as the parent of the other interprediction PU modes, as shown in Fig. 2. Since high spatial correlation and similar properties of the pixels within a CU, the children PU partition modes, such as Inter_2N × N, Inter_N × 2N, Inter_N × N and so on, may have a large probability to select the same reference direction and reference frame as their parent PU (Inter_2N × 2N) does. In other words, there may have a large probability that $\varphi^* = \varphi_{Inter\_2N\times 2N}$ and $\gamma^* = \gamma_{Inter\_2N\times 2N}$ are used for encoding the children PUs.

In order to analyze the reference frame selection correlation between the parent PU and the children PUs, the event **A** represents both reference direction and reference frame index of the children PUs are equal to the reference direction and reference frame of their parent PU. Six HEVC test video sequences with various resolutions and motion activities are analyzed with different coding configurations. The statistical results of $P(\mathbf{A})$ are tabulated in Table 2.

From Table 2, it can be seen that for the random-access-main coding structure, the probability of $P(\mathbf{A})$ is from 60.25% to 66.28%, 63.26% on average. For the video sequence with slow motion such as *BQSquare*, $P(\mathbf{A})$ is 64.02% on average; while for the video sequence with medium or violent motion such as *BasketballDrill*, the average probability of $P(\mathbf{A})$ is approximate to 61.76%. Another observation from Table 2 is that for the low-delay-main coding structure, the probability of $P(\mathbf{A})$ is from 52.31% to 89.27%, 67.74% on average. For the video sequence with fast motion activities such as *BasketballDrill*, the probability of $P(\mathbf{A})$ is 63.51% on average; for the video sequence with slow motion such as *FourPeople* and *Johnny*, the probabilities of $P(\mathbf{A})$ are 79.30% and 76.22%, respectively. From Table 2, we can see that the probabilities $P(\mathbf{A})$ of the slow motion sequences are larger than that of the fast motion sequences, this is because that the content similarity of temporal successive frames of slow motion sequences is higher than that of the fast motion sequences. In addition, the probability $P(\mathbf{A})$ is large as the QP increases, the reason is when the QP increases, the prediction residuals will be transformed and quantized into small values, which results in large reference frame information selection correlation between the parent PU and the children PUs. From these values in Table 2, we can figure out that (1) the children PUs have a large probability to select the same reference direction as their parent PU does; (2) the children PU modes have a large probability to select the same reference frame index as their parent PU does.

Based on the probability of $\varphi^* = \varphi_{Inter\_2N\times 2N}$ and $\gamma^* = \gamma_{Inter\_2N\times 2N}$, the optimal problem of Eq. (1) can be solved by the following two steps. Firstly, encode the parent PU, Inter_2N × 2N, with all active reference frames and reference directions, and the best reference direction and reference frame are obtained as

$$\{\varphi_{Inter\_2N\times 2N}, \gamma_{Inter\_2N\times 2N}\} = \arg\min_{\varphi}\left(\arg\min_{\gamma} J(\mathbf{O}, \mathbf{R}(d, p, \varphi, \gamma))\right)\Big|_{p=Inter\_2N\times 2N}, \quad (3)$$

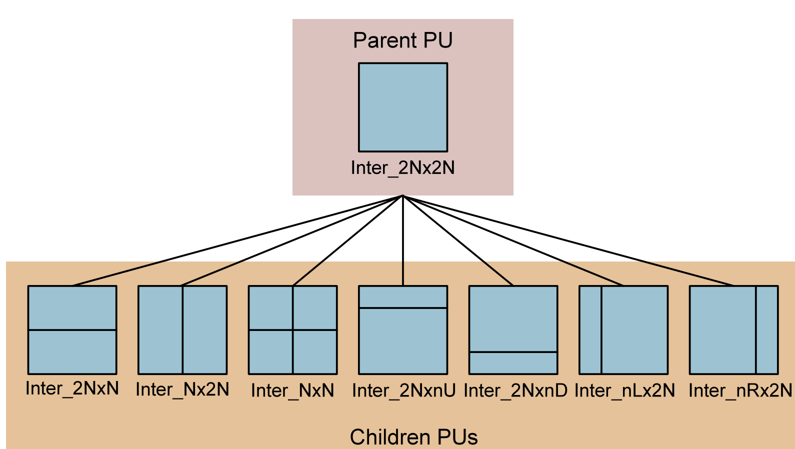

Fig. 2. An illustration on the parent PU and children PUs.

Table 2
Statistical analyses on the probability $P(\mathbf{A})$ (%).

Sequence	Low-delay-main coding structure					Random-access-main coding structure				
	QP = 22	QP = 27	QP = 32	QP = 37	Average	QP = 22	QP = 27	QP = 32	QP = 37	Average
BQMall	60.36	64.64	68.54	72.43	66.49	64.22	63.22	62.43	61.91	62.95
BasketballDrill	57.50	60.93	65.30	70.30	63.51	63.67	62.27	60.79	60.32	61.76
BQSquare	52.31	52.93	57.11	60.72	55.77	66.28	65.17	63.63	60.99	64.02
BasketballPass	61.35	65.13	68.75	72.05	66.82	65.24	64.1	62.2	60.28	62.96
FourPeople	65.24	78.41	84.29	89.27	79.3	n/a	n/a	n/a	n/a	n/a
Johnny	62.88	75.05	80.84	86.10	76.22	n/a	n/a	n/a	n/a	n/a
Cactus	58.68	68.72	79.06	72.43	69.72	65.71	65.38	65.82	65.82	65.68
ParkScene	56.69	60.61	65.80	73.17	64.07	64.74	62.83	61.05	60.25	62.22
Average	60.23	68.06	73.97	77.28	67.74	64.98	63.83	62.65	61.60	63.26

it means that the best reference direction and the best reference frame index of the parent PU mode are chosen from all active reference directions and reference frame indexes, and the best reference frame information is the one with the minimum RD cost. After encoding the CU with the parent PU mode, Inter_2N × 2N, the best reference direction, $\varphi_{Inter\_2N\times 2N}$, and the best reference frame index, $\gamma_{Inter\_2N\times 2N}$ are stored for the fast reference frame decision of the children PU modes.

Then, the CU is encoded with the children PU modes by using the reference direction, $\varphi_{Inter\_2N\times 2N}$, and reference frame index, $\gamma_{Inter\_2N\times 2N}$. That is

$$\{d^*, p^*\} = \arg\min_{d\in\mathbf{D}}\left(\arg\min_{p\in\mathbf{P}} J(\mathbf{O}, \mathbf{R}(d, p, \varphi, \gamma))\right)\Bigg|_{\varphi=\varphi_{Inter\_2N\times 2N}, \gamma=\gamma_{Inter\_2N\times 2N}}, \tag{4}$$

where $\mathbf{P}$ is the set of the children PU modes, $\mathbf{P}$ = {Inter_2N × N, Inter_N × 2N, Inter_N × N, Inter_2N × nU, Inter_2N × nD, Inter_nL × 2N, Inter_nR × 2N}. And, the best reference direction and the best reference frame index of the children PU modes are obtained by learning the results of their parent PU mode, then the reference frame information is used for encoding the CU with the children PU modes. As a result, the encoding complexity of the MRF selection and HEVC encoder could be removed significantly.

3.3. The overall algorithm

Based on the above analyses, the proposed fast MRF selection algorithm is summarized and illustrated step-by-step as follows.

Step 1. Encode the current CU with the Merge/Skip mode. Go to Step 2;
Step 2. Encode the current CU with the parent PU mode (Inter_2N × 2N) according to the Eq. (3), all active reference directions and reference frame indexes are used for encoding the parent PU mode, then the best reference frame information $(\varphi_{Inter\_2N\times 2N}, \gamma_{Inter\_2N\times 2N})$ which is with the minimum RD cost is stored for encoding the children PU modes. Go to Step 3;
Step 3. Encode the current CU with the children PU modes according to the Eq. (4), the best reference frame information of the children PU modes are obtained by learning the results of their parent PU mode, and the achieved reference frame information $(\varphi_{Inter\_2N\times 2N}, \gamma_{Inter\_2N\times 2N})$ is applied for encoding the current CU with the children PU modes. Go to Step 4;
Step 4. Encode the current CU with the Intra PU modes. Go to Step 5;
Step 5. Store the coding information and write the encoded bit stream. Go back to Step 1 to process the next CU.

4. Experimental results

To evaluate the coding performance of the proposed fast MRF selection algorithm, the HEVC reference software HM12.0 is used as the software platform. The hardware platform is Intel Xeon CPU E5-1620 v2 @ 3.70 GHz, 16.0 GB RAM with the Microsoft Windows 7 64-bit operating system. To compare the coding performance in terms of BDPSNR, BDBR, total encoding time saving (TS for short), total reference frame encoding time saving (RTS for short), three 416 × 240 sequences with Class D (*BQSquare*, *BlowingBubblus*, *BasketballPass*), four 832 × 480 sequences with Class C (*RaceHorses*, *BQMall*, *PartyScene*, *BasketballDrill*), six 1280 × 720 sequences with Class E (*FourPeople*, *Johnny*, *KristenAndSara*, *Vidyo1*, *Vidyo3*, *Vidyo4*), four 1920 × 1080 sequences with Class B (*ParkScene*, *Cactus*, *BQTerrace*, *BasketballDrive*), two 2560 × 1600 sequences with Class A (*Traffic*, *PeopleOnStreet*) are tested according to the HEVC common test conditions [22]. The number of coding frames of 416 × 240, 832 × 480, 1280 × 720, 1920 × 1080, and 2560 × 1080 sequences are 97, 81, 65, 41, and 33, respectively. Four QPs (22, 27, 32, and 37) are used. The ME search method is the TZSearch, and the search range equals to 64. The proposed algorithm is compared with Wang's algorithm [19], the HM12.0 is used as the benchmark. The test results are compared and summarized in Table 3. In the table, the Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and Bjontegaard delta bit rate (BDBR) are computed according to [24,25], that the BDPSNR indicates the average peak signal-to-noise ratio (PSNR) differences in dB while they are with the same bit rates, and the BDBR denotes the average bit rate differences in percent while they are with the same PSNR [26–28]. The TS and RTS are computed as

$$\begin{cases} TS = \frac{1}{4}\sum_{i=1}^{4}\frac{T_\phi(QP_i)-T_o(QP_i)}{T_o(QP)} \times 100\%, \\ RTS = \frac{1}{4}\sum_{i=1}^{4}\frac{TR_\phi(QP_i)-TR_o(QP_i)}{TR_o(QP_i)} \times 100\%, \end{cases} \tag{5}$$

where $T_\phi(QP_i)$ and $TR_\phi(QP_i)$ indicate the total encoding time of the HM12.0 with the fast MRF selection algorithm ϕ with QP_i, and the total reference frame selection time of the fast MRF selection algorithm ϕ with $QP_i, \phi \in$ {Wang [19], Proposed}, respectively; $T_o(QP_i)$ and $TR_o(QP_i)$ denote the total encoding time of the HM12.0 with the original full MRF selection with QP_i, and the total reference selection time of the original full MRF selection with QP_i, respectively; QP_i = {22, 27, 32, 37}.

From Table 3, it can be observed that for the low-delay-main coding structure, the Wang's algorithm reduces the total encoding time from 19.50% to 27.83%, 23.83% on average; and removes the MRF encoding complexity from 32.64% to 38.46%, 34.58% on

Table 3
Summary of encoding results.

Method	Class	Sequence	Low-delay-main coding structure				Random-access-main coding structure			
			BDPSNR (dB)	BDBR (%)	TS (%)	RTS (%)	BDPSNR (dB)	BDBR (%)	TS (%)	RTS (%)
Wang [19] VS Original HM12.0	A	Traffic	n/a*	n/a	n/a	n/a	−0.014	0.39	−5.45	−7.71
		PeopleOnStreet	n/a	n/a	n/a	n/a	−0.013	0.29	−2.79	−3.96
	B	ParkScene	−0.161	3.64	−23.08	−33.24	−0.003	0.09	−2.57	−3.38
		Cactus	−0.034	1.45	−22.02	−32.64	−0.003	0.10	−1.76	−2.56
		BQTerrace	−0.191	9.85	−21.98	−33.44	−0.013	0.65	−3.56	−5.41
		BasketballDrive	−0.125	3.23	−22.00	−32.45	0.001	−0.11	−1.94	−2.70
	C	RaceHorses	−0.099	2.37	−22.96	−32.99	−0.019	0.48	−3.00	−4.44
		BQMall	−0.086	2.07	−22.42	−33.20	−0.012	0.30	−2.70	−4.16
		PartyScene	−0.161	3.64	−20.43	−32.30	−0.039	0.83	−2.26	−3.40
		BasketballDrill	−0.125	3.23	−22.00	−32.45	−0.031	0.77	−2.24	−3.27
	D	BQSquare	−0.366	9.26	−19.50	−31.61	−0.067	1.46	−2.40	−3.57
		BlowingBubbles	−0.074	1.92	−22.20	−33.21	−0.018	0.46	−2.30	−3.57
		BasketballPass	−0.025	0.53	−23.96	−34.29	−0.010	0.19	−4.00	−5.84
	E	FourPeople	−0.052	1.53	−25.75	−36.60	n/a	n/a	n/a	n/a
		Johnny	−0.040	1.53	−26.77	−37.24	n/a	n/a	n/a	n/a
		KristenAndSara	−0.024	0.83	−27.53	−37.96	n/a	n/a	n/a	n/a
		Vidyo1	−0.029	0.86	−27.67	−38.46	n/a	n/a	n/a	n/a
		Vidyo3	−0.020	0.60	−26.98	−37.78	n/a	n/a	n/a	n/a
		Vidyo4	−0.016	0.59	−27.87	−38.03	n/a	n/a	n/a	n/a
	Average		−0.096	2.77	−23.83	−34.58	−0.019	0.45	−2.84	−4.15
ProposedVSOriginal HM12.0	A	Traffic	n/a	n/a	n/a	n/a	−0.023	0.62	−31.42	−44.74
		PeopleOnStreet	n/a	n/a	n/a	n/a	−0.052	1.16	−32.50	−45.71
	B	ParkScene	−0.032	1.04	−36.73	−53.05	−0.019	0.59	−30.58	−43.27
		Cactus	−0.018	0.73	−36.86	−54.78	−0.013	0.55	−31.56	−45.84
		BQTerrace	−0.023	1.11	−33.84	−50.73	−0.013	0.64	−28.03	−41.06
		BasketballDrive	−0.021	0.97	−39.42	−54.37	−0.017	0.78	−31.12	−43.87
	C	RaceHorses	−0.062	1.49	−40.89	−58.74	−0.069	1.75	−30.44	−45.21
		BQMall	−0.052	1.24	−37.73	−56.29	−0.046	1.10	−29.49	−44.54
		PartyScene	−0.062	1.38	−34.31	−54.76	−0.044	0.95	−25.61	−40.71
		BasketballDrill	−0.034	0.88	−37.92	−55.98	−0.028	0.68	−30.57	−45.59
	D	BQSquare	−0.095	2.29	−31.84	−51.64	−0.041	0.90	−24.35	−38.29
		BlowingBubbles	−0.064	1.68	−36.15	−55.83	−0.041	1.02	−27.13	−42.04
		BasketballPass	−0.055	1.14	−39.22	−56.87	−0.045	0.95	−30.04	−44.17
	E	FourPeople	−0.019	0.53	−37.55	−53.31	n/a	n/a	n/a	n/a
		Johnny	−0.035	1.16	−37.64	−52.74	n/a	n/a	n/a	n/a
		KristenAndSara	−0.027	0.92	−38.25	−53.00	n/a	n/a	n/a	n/a
		Vidyo1	−0.016	0.44	−37.86	−53.02	n/a	n/a	n/a	n/a
		Vidyo3	−0.034	0.94	−38.00	−53.49	n/a	n/a	n/a	n/a
		Vidyo4	−0.008	0.33	−39.25	−54.39	n/a	n/a	n/a	n/a
	Average		−0.039	1.07	−37.26	−54.29	−0.035	0.90	−29.45	−43.46

* The "n/a" denotes the result is not available, because the sequences of Class A with the random-access-main coding structure, and the sequences of Class E with the low-delay-main coding structure are not tested according to the HEVC common test conditions [22,23].

average. Meanwhile, the BDPSNR between the Wang's algorithm and the original HM12.0 is from −0.366 dB to −0.016 dB, −0.096 dB on average; the BDBR between the Wang's algorithm and the original HM12.0 is from 0.53% to 9.26%, 2.77% on average. For the sequence with large global motion, *BQSqure*, the RD performance degrades significantly. The proposed algorithm reduces the total encoding time from 31.84% to 40.89%, 37.26% on average; and saves the MRF encoding time from 50.73% to 58.74%, 54.29% on average; meanwhile, the BDPSNR between the original HM12.0 and the proposed algorithm is from −0.008 dB to −0.095 dB, −0.039 dB on average; and the BDBR between the original HM12.0 and the proposed algorithm is from 0.33% to 2.29%, 1.07% on average. From these values, we can see that the proposed method obtains a better performance in terms of RD performance and encoding time saving than the Wang's method.

For the random-access-main coding structure, the Wang's algorithm reduces the total encoding time from 1.76% to 5.45%, 2.84% on average; and reduces the MRF encoding time from 2.56% to 7.71%, 4.15% on average. While, the BDPSNR between the Wang's algorithm and the original HM12.0 is from −0.067 dB to 0.001 dB, −0.019 dB on average; and the BDBR between the Wang's algorithm and the original HM12.0 is from −0.11% to 1.46%, 0.45% on average. From these values, it can be observed that the Wang's algorithm can not efficiently reduce the encoding complexity of the HEVC encoder with the random-access-main coding structure. The proposed algorithm saves the total encoding time from 24.35% to 31.56%, 29.45% on average; and reduces the MRF encoding time from 38.29% to 45.84%, 43.46% on average; meanwhile, the BDPSNR between the original HM12.0 the proposed algorithm is from −0.013 dB to −0.069 dB, −0.035 dB on average; and the BDBR between the original HM12.0 and the proposed algorithm is from 0.55% to 1.75%, 0.90% on average. For the video sequences *PeopleOnStreet* and *RaceHorses*, the BDBR of the proposed method increases a little, this is because these two sequences are with violent motion, and results in the relationship between the parent PU (Inter_2N × 2N) and the children PUs (Inter_2N × N, Inter_N × 2N, and so on) has a little decrease. In summary, the proposed method achieves a better coding performance than the Wang's method in not only the coding complexity reduction but also the RD performance.

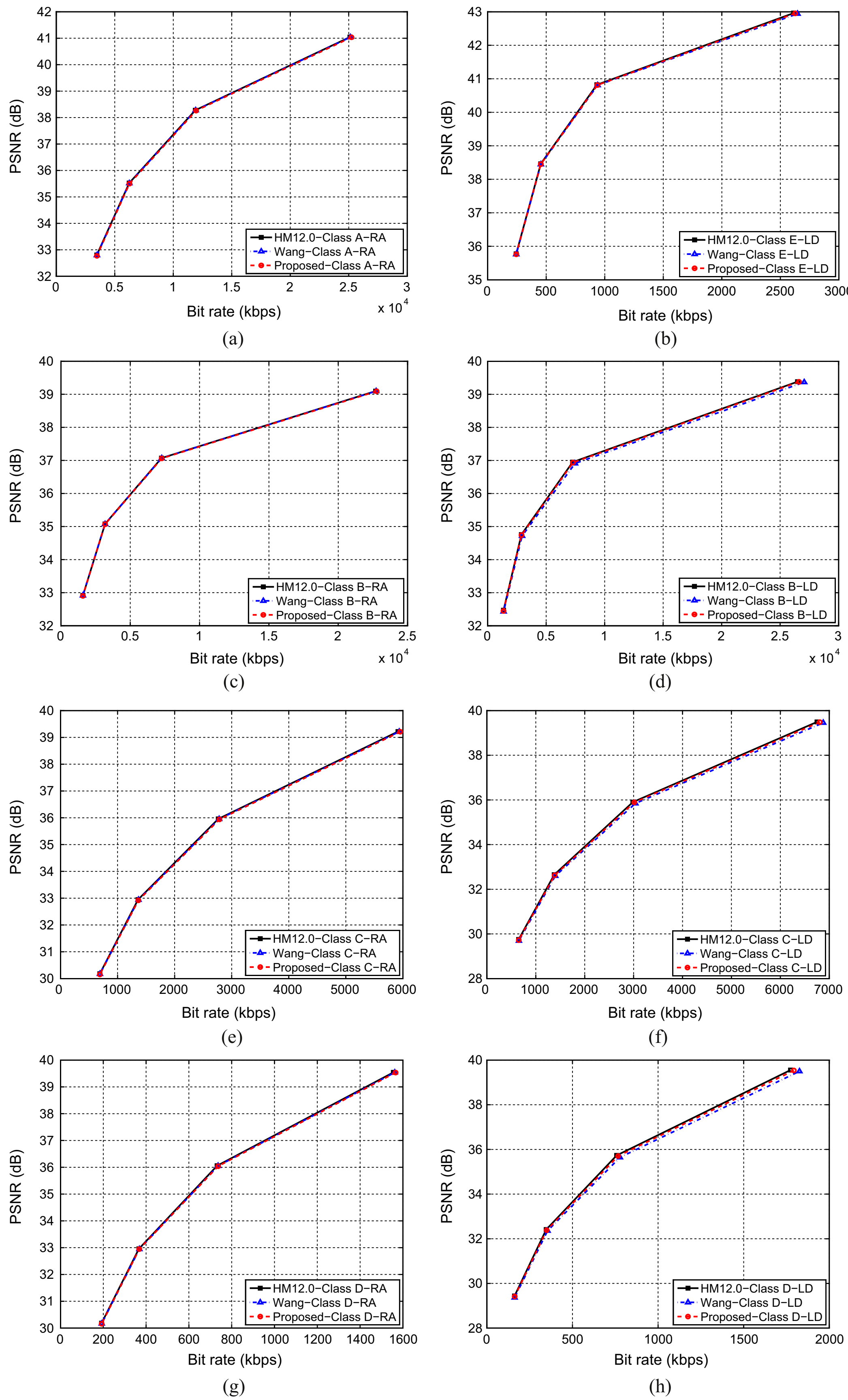

Fig. 3. RD performance comparison.

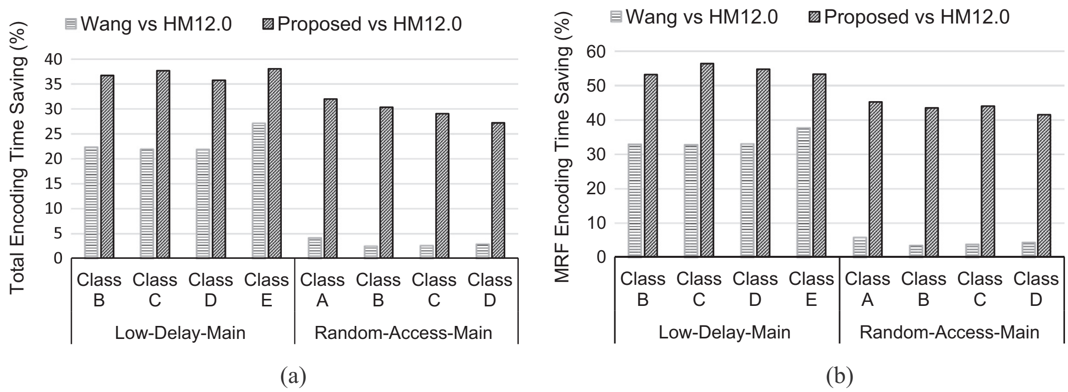

Fig. 4. Summary of encoding time saving. (a) Total encoding time saving. (b) MRF encoding time saving.

To show the RD performance intuitively, we perform a RD-curves comparison among the Wang's algorithm, the proposed algorithm and the original HM12.0, the results are shown in Fig. 3. In the legend, the first item indicates the algorithm; the second item denotes the test sequences; and the last item is the coding structure, the LD represents the low-delay-main coding structure, and the RA indicates the random-access-main coding structure. From Fig. 3, it can be observed that the proposed algorithm achieves almost the same RD performance as the original HM12.0. Moreover, the proposed algorithm achieves an obviously better RD performance than the Wang's algorithm for the low-delay-main coding structure. Fig. 4 shows the summary of encoding time saving, Fig. 4(a) is the total encoding time saving comparison between the Wang's method and proposed method, it can be observed that the proposed method achieves a better encoding time saving performance than the wang's algorithm. Compared to the Wang's algorithm, the proposed algorithm achieves more than 17.92% and 27.44% encoding time saving for the HEVC encoder with the low-delay-main and random-access-main coding structures, respectively. Fig. 4(b) is the MRF encoding time saving comparison between the Wang's method and the proposed method, it can be seen that the proposed method gains a significantly better MRF encoding time saving performance than the Wang's algorithm. Compared to the Wang's algorithm, the proposed algorithm obtains more than 30.82% and 40.98% MRF encoding time saving for the HEVC encoder with the low-delay-main and random-access-main coding structures, respectively. From these values, we can see that the proposed method reduces the encoding complexity of the HEVC encoder more efficient than the Wang's method. From Figs. 3 and 4, we can figure out that the proposed algorithm efficiently reduces the encoding complexity of the HEVC encoder and maintains a comparable RD performance.

5. Algorithm discussion

It is well known that when designing the fast algorithms for video coding, the decision accuracy of the proposed algorithm is highly correlated with the coding efficiency. In other words, if the decision accuracy is large and close to 100%, there would be no RD performance degraded; on the contrary, if the decision accuracy is small, the RD performance would be degraded. From Table 2, we can see that the average decision accuracy of the proposed algorithm is 63.26% and 67.74% for the random-access-main and low-delay-main coding structures, respectively. From Table 3, it can be observed that the average BDBR between the original HM12.0 and the proposed algorithm is 0.90% and 1.07% for the random-access-main and low-delay-main coding structures, respectively. We can see that the decision accuracy of the proposed algorithm is not large and less than 80%, while the RD performance degradation is quite acceptable. In order to exploit this causation, the best PU mode distribution in the HEVC encoding process is analyzed, the test conditions are same as in Section 4, the results are summarized and tabulated in Table 4.

Table 4
The best mode distribution in HEVC encoding process (%).

Sequence	Low-delay-main		Random-access-main	
	2N × 2N	Others	2N × 2N	Others
BQMall	90.52	9.48	91.56	8.45
BasketballDrill	93.51	6.49	93.87	6.13
BQSquare	88.40	11.60	92.90	7.10
BasketballPass	90.48	9.52	92.52	7.48
FourPeople	98.09	1.91	n/a	n/a
Johnny	98.14	1.86	n/a	n/a
Cactus	94.38	5.62	95.04	4.96
ParkScene	93.48	6.52	94.76	5.24
Average	93.38	6.62	93.44	6.56

In Table 4, the "$2N \times 2N$" indicates the $2N \times 2N$ partition, which includes two modes, Merge/Skip and Inter_$2N \times 2N$; "Others" means the remaining modes, including Inter_$2N \times N$, Inter_$N \times 2N$, Inter_$N \times N$, Inter_$2N \times nU$, Inter_$2N \times nD$, Inter_$nL \times 2N$, Inter_$nR \times N$, Intra_$N \times N$, and Intra_$2N \times 2N$. From Table 4, we can see that most of CUs select the $2N \times 2N$ as their best mode in the HEVC CU encoding process. The percentage for the "$2N \times 2N$" is from 88.40% to 98.14%, 93.38% on average, for the low-delay-main coding structure; and from 91.56% to 95.04%, 93.44% on average, for the random-access-main coding structure. While the percentage of the "Other" modes is only 6.62% and 6.56% for the low-delay-main and random-access-main coding structures, respectively. It can be observed that more than 90% CUs select the "$2N \times 2N$" as their best PU mode, and this is the reason why the probability of $P(\mathbf{A})$ is less than 80% and the RD degradation of the proposed MRF selection is negligible.

There is a small number of CUs selecting the "Other" modes as their best mode, does that mean the MRF selection process of the "Other" modes is not necessary for the HEVC CU encoding process? In order to demonstrate the importance of the MRF selection process of the "Other" modes, an experiment is performed. In the experiment, the "Other" modes are skipped directly, and only the Merge/Skip and Inter_$2N \times 2N$ are adopted. The experimental results are summarized in terms of BDPSNR and BDBR, and tabulated in Table 5.

From Table 5, it can be seen that when the MRF selection process of the "Other" modes is skipped in the HEVC CU encoding

Table 5
Encoding results of the other modes skipped scheme.

Sequence	Low-delay-main		Random-access-main	
	BDPSNR(dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)
BQMall	−0.415	10.33	−0.492	12.28
BasketballDrill	−0.304	8.05	−0.421	10.80
BQSquare	−0.230	5.62	−0.161	3.53
BasketballPass	−0.424	9.16	−0.543	11.88
FourPeople	−0.123	3.51	n/a	n/a
Johnny	−0.105	4.38	n/a	n/a
Cactus	−0.157	6.99	−0.157	6.99
ParkScene	−0.094	3.04	−0.105	3.39
Average	−0.232	6.38	−0.313	8.15

process, the RD performance degrades dramatically. For the low-delay-main coding structure, the BDPSNR between the the original HM12.0 and the MRF selection of the "Other" modes skipped scheme is from −0.123 dB to −0.424 dB, −0.232 dB on average; and the BDBR between the the original HM12.0 and the MRF selection of the "Other" modes skipped scheme is from 3.04% to 10.33%, 6.38% on average. For the random-access-main coding structure, the BDPSNR between the the original HM12.0 and the MRF selection of the "Other" modes skipped scheme is from −0.105 dB to −0.492 dB, −0.313 dB on average; and the BDBR between the the original HM12.0 and the MRF selection of the "Other" modes skipped scheme is from 3.39% to 12.28%, 8.15% on average. These values reflect that the MRF selection process of the "Other" modes is extremely important in the HEVC CU encoding process, which can efficiently remove the data redundancies among video sequences. Thus, the MRF selection process of the "Other" modes can not be skipped in the HEVC CU encoding process.

6. Conclusion

The MRF encoding process consumes about 70% of total encoding time of an HEVC encoder. To reduce the computational complexity of the MRF encoding process, an early reference frame decision algorithm is proposed in this paper. Since there is high video content similarity between the parent PU and children PUs, the reference frame information including inference frame index and reference frame direction of the children PUs is set according to the parent PU has. Experimental results show that the proposed algorithm obtains a promising encoding performance in terms of encoding time saving and RD performance.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61501246, Grant 61271324, Grant 61471348, Grant 61232016, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20150930, in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 15KJB510019, in part by the Natural Science Foundation of Hebei Province of China under Grant F2015202311, in part by the Project through the Priority Academic Program Development of Jiangsu Higher Education Institutions, in part by the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology under Grant 2015r012, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2016A030306022.

References

[1] Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 2: High efficiency video coding, ISO/IEC 23008-2:2013 ITU-T Rec. H.265, 2013.
[2] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, IEEE Trans. Circ. Syst. Video Technol. 22 (12) (2012) 1649–1668.
[3] H. Wang, H. Du, W. Lin, S. Kwong, Oscar C. Au, J. Wu, Z. Wei, Early detection of all-zero 4 × 4 blocks in high efficiency video coding, J. Visual Commun. Image Represent. 25 (7) (2014) 1784–1790.
[4] J. Chen, J. Villasenor, Y. He, G. Luo, Parallel fast inter mode decision for H.264/AVC encoding, J. Visual Commun. Image Represent. 24 (8) (2013) 1443–1449.
[5] J. Lee, S. Kim, K. Lim, S. Lee, A fast CU size decision algorithm for HEVC, IEEE Trans. Circ. Syst. Video Technol. 25 (3) (2014) 411–421.
[6] L. Shen, Z. Liu, X. Zhang, W. Zhao, Z. Zhang, An effective CU size decision method for HEVC encoders, IEEE Trans. Multimedia 15 (2) (2013) 465–470.
[7] D.-H. Kim, Y.-H. Kim, W.-C. Park, Selective CU depth range decision algorithm for HEVC encoder, in: IEEE International Symposium on Consumer Electronics, 2014 (ISCE 2014), JeJu Island, Korea, 2014, pp. 1–2.
[8] J. Vanne, M. Viitanen, T.D. Hamalainen, Efficient mode decision schemes for HEVC inter prediction, IEEE Trans. Circ. Syst. Video Technol. 24 (9) (2014) 1579–1593.
[9] Z. Pan, S. Kwong. M.-T. Sun, J. Lei, Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC, IEEE Trans. Broadcasting 60 (2) (2014) 405–412.
[10] J. Xiong, H. Li, Q. Wu, F. Meng, A fast HEVC inter CU selection method based on pyramid motion divergence, IEEE Trans. Multimedia 16 (2) (2014) 559–564.
[11] L. Shen, Z. Zhang, P. An, Fast CU size decision and mode decision algorithm for HEVC intra coding, IEEE Trans. Consumer Electron. 9 (1) (2013) 207–213.
[12] Y. Su, M.-T. Sun, Fast multiple reference frame motion estimation for H.264/AVC, IEEE Trans. Circ. Syst. Video Technol. 16 (3) (2006) 447–452.
[13] D. Jun, H. Park, An efficient priority-based reference frame selection method for fast motion estimation in H.264/AVC, IEEE Trans. Circ. Syst. Video Technol. 20 (8) (2010) 1156–1161.
[14] M.-J. Chen, G.-L. Li, Y.-Y. Chiang, C.-T. Hsu, Fast multiframe motion estimation algorithms by motion vector composition for the MPEG-4/AVC/H.264 standard, IEEE Trans. Multimedia 8 (3) (2006) 478–487.
[15] Z. Liu, L. Li, Y. Song, S. Li, S. Goto, T. Ikenaga, Motion feature, Motion feature and hadamard coefficient-based fast multiple reference frame motion estimation for H.264, IEEE Trans. Circ. Syst. Video Technol. 18 (5) (2008) 620–632.
[16] Y. Zhang, S. Kwong, G. Jiang, H. Wang, Efficient multi-reference frame selection algorithm for hierarchical B pictures in multiview video coding, IEEE Trans. Broadcasting 57 (1) (2011) 15–23.
[17] C.-H. Yeh, M.-F. Li, M.-J. Chen, M.-C. Chi, X.-X. Huang, C.-W. Chi, Fast mode decision algorithm through inter-view rate-distortion prediction for multiview video coding system, IEEE Trans. Indust. Inform. 10 (1) (2014) 594–603.
[18] J. Lei, J. Sun, Z. Pan, S. Kwong, C. Hou, Fast mode decision using inter-view and inter-component correlations for multiview depth video coding, IEEE Trans. Indust. Inform. 11 (4) (2015) 978–986.
[19] S. Wang, S. Ma, S. Wang, D. Zhao, W. Gao, Fast multi reference frame motion estimation for high efficiency video coding, in: Proc. ICIP, Melbourne, Australia, 2014, pp. 2005–2009.
[20] J.-R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, T. Wiegand, Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC), IEEE Trans. Circ. Syst. Video Technol. 22 (12) (2012) 1669–1684.
[21] F. Bossen, D. Flynn, K. Suehring, HEVC HM12 Reference Software, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) Document JCTVC-N1010, Jul. 2013.
[22] F. Bossen, Common test conditions and software reference configurations, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) Document JCTVC-J1100, Mar. 2012.
[23] F. Bossen, Common test conditions and software reference configurations, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) Document JCTVC-D600, Jan. 2011.
[24] G. Bjontegaard, Calculation of average PSNR differences between RD curves, ITU-T VCEG, Document VCEG-M33, Apr. 2001.
[25] Z. Pan, Y. Zhang, J. Lei, L. Xu, X. Sun, Early DIRECT mode decision based on all-zero block and rate distortion cost for multiview video coding, IET Image Process. 10 (1) (2016) 9–15.
[26] P. Hanhart, T. Ebrahimi, Calculation of average coding efficiency based on subjective quality scores, J. Visual Commun. Image Represent. 25 (3) (2014) 555–564.
[27] Z. Pan, Y. Zhang, S. Kwong, Efficient motion and disparity estimation optimization for low complexity multiview video coding, IEEE Trans. Broadcasting 61 (2) (2015) 166–176.
[28] Z. Pan, J. Lei, Y. Zhang, X. Sun, S. Kwong, Fast motion estimation based on content property for low-complexity H.265/HEVC encoder, IEEE Trans. Broadcasting (2016) 1–10, http://dx.doi.org/10.1109/TBC.2016.2580920.