Efficient Motion and Disparity Estimation Optimization for Low Complexity Multiview Video Coding

Zhaoqing Pan, Yun Zhang, *Member, IEEE*, and Sam Kwong, *Fellow, IEEE*

***Abstract*—The use of variable block-size motion estimation (ME), disparity estimation (DE), and multiple reference frames selection aims to improve the coding efficiency of multiview video coding (MVC), however, this is at the cost of high computational complexity of these advanced coding techniques, which are not suitable for real-time video broadcasting applications. In this paper, we propose an efficient ME and DE algorithm for reducing the computational complexity of MVC. Firstly, according to the characteristics of the coded block pattern and rate distortion (RD) cost, an early DIRECT mode decision algorithm is proposed. Then, based on the characteristics of the initial search point in the ME/DE process and the observation that the best point is center-biased, an early ME/DE termination strategy is proposed. If the ME/DE early termination is not satisfied, the ME/DE search window will be reduced by applying the optimal theory. At last, two block matching search strategies are proposed to predict the best point for the ME/DE. Experimental results show that the proposed algorithm can achieve 50.05% to 77.61%, 64.83% on average encoding time saving. Meanwhile, the RD performance degradation is negligible. Especially, the proposed algorithm can be applied to not only the odd views but also the even views.**

***Index Terms*—Early termination, optimal stopping theory, motion estimation (ME), disparity estimation (DE), multiview video coding (MVC).**

I. Introduction

MULTIVIEW video is a crucial data format for many multimedia applications, such as free-viewpoint television (FTV), three-dimensional television (3DTV) and so on, which is captured simultaneously by multiple cameras from different viewpoints. However, with the increased number of cameras, the volume of raw video data increases rapidly. Compared to traditional single-view video, multiview video needs more storage space, larger transmission bandwidth and larger computational power for coding. A straightforward method for encoding multiview video is to independently encode each view by using an existing video codec such as H.264/AVC. However, the coding efficiency of this coding method is limited because of the inter-view redundancies. For further improving the coding efficiency, the multiview video coding (MVC) [1], [2] is proposed to exploit inter-view and temporal redundancies by adopting a group of advanced coding tools, such as hierarchial B picture (HBP) prediction structure, variable block-size motion estimation (ME) and disparity estimation (DE). However, the achieved coding efficiency is at the cost of high computational complexity of these advanced tools which obstruct MVC to be used in real-time multimedia applications (e.g., 3D live broadcasting). Recently, the multiview and 3D extensions for the HEVC standard (MV/3D-HEVC) [3], [4] are attracting more and more people's attention. The high efficiency video coding is the newest video coding standard, which can achieve about 50% bit-rate saving and maintain the same image quality as compared with H.264/AVC. However, the achieved coding efficiency comes at the cost of huge computational complexity. Currently, the MVC and MV/3D-HEVC are facing the same problem - high computational complexity. Hence, reducing the computational complexity is vital for these video coding standards.

To address the high computational complexity of these advanced coding tools, a number of fast algorithms have been proposed for variable block-size ME in H.264/AVC. Zeng *et al.* [5] proposed a fast mode decision scheme for H.264/AVC, which is based on the rate distortion (RD) cost and macroblock (MB) motion activity. In [6], Zhao *et al.* proposed a fast mode decision for H.264/SVC by using the optimal stopping theory. However, the characteristics of the motion vector (MV) distribution and mode selection priority in MVC are different from these in H.264/AVC due to the HBP prediction structure and inter-views correlations in MVC. In [7], Lee proposed an efficient fast ME algorithm, which uses an adaptive search range reduction technique and a block matching error prediction. Po *et al.* [8] proposed a directional gradient descent search algorithm for fast ME, that multiple

Manuscript received May 14, 2014; revised December 11, 2014; accepted December 24, 2014. Date of publication April 27, 2015; date of current version June 4, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61232016, Grant 61272289, Grant 61471348, and Grant 61102088, 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 NUIST, and in part by the Shenzhen Overseas High-Caliber Personnel Innovation and Entrepreneurship Project under Grant KQCX20140520154115027.

Z. Pan is with the Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, China, also with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: zhaoqingpan@nuist.edu.cn).

Y. Zhang is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Department of Computer Science, City University of Hong Kong, Hong Kong.

S. Kwong is with the Department of Computer Science, City University of Hong Kong, Hong Kong, and also with the City University of Hong Kong Shenzhen Research Institute, Shenzhen 51800, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2015.2419824

one-at-a-time searches and gradient descent searches are performed on the error surface in eight directions. However, large search range (such as 64, 96, and larger) and high resolution video sequences (such as 640×480, 1024×768 and so on) are not considered by these fast ME algorithms. Especially, the DE needs larger search range to remove the inter-view redundancies. Thus, if MVC directly uses these small search range based fast ME algorithms, the RD performance will degrade dramatically [9]. By using neighboring reference-frame information and Bayes theory, a fast reference frame selection algorithm was proposed in [10]. Based on the spatial and temporal correlations between the reference frames and the MVs, an efficient reference frame selection scheme was proposed to reduce the computational complexity in H.264/AVC [11]. These algorithms can efficiently reduce the computational complexity in H.264/AVC. However, they are not so efficient for MVC, because the characteristics of hierarchial B frames and the correlations among inter-views are not being considered.

In order to deal with the limitations and make full use of the correlations among inter-views, many researchers have devoted their work in developing fast variable block-size ME/DE algorithms for MVC. In [12], Seo and Sohn proposed an early DE skipping for MVC, which is based on the characteristics of the MVs and disparity vectors. In [9], Pan *et al.* proposed a multiple hexagon search algorithm for the ME and DE in MVC. In [13], Purnachand *et al.* proposed an improved TZSearch ME algorithm by using a hexagon search pattern and a searching threshold. According to the characteristics of loop-epipolar constraint, a fast ME algorithm was proposed for MVC [14]. In [15], Wang *et al.* proposed an adaptive early DIRECT mode decision algorithm for low complexity MVC. By using the optimal stopping theory, [16] derived a fast mode decision for fast MVC. According to the quantization parameter, video content, and motion properties, a DIRECT mode early termination model was proposed for MVC fast mode decision [17]. In [18], Zhang *et al.* proposed an efficient multi-reference frame selection algorithm for the hierarchical B frame, which uses the spatial correlations within an MB in which smaller size block may select the same reference frame and direction as the 16×16 MB mode does. By using the RD cost property and inter-view correlation, they also proposed an early DIRECT mode decision algorithm for MVC [19]. These algorithms can be applied to not only the odd views but also the even views. However, the computational complexity of the variable block-size ME/DE is reduced individually in motion search, mode decision or reference frame selection, thus, the computational complexity can be further reduced by jointly optimizing the variable block-size ME/DE.

Based on the MB motion homogeneity, Shen *et al.* proposed a fast DE and ME algorithm to reduce the computational complexity of MVC [21]. In that algorithm, the motion homogeneity is measured by the MV of the spatial neighboring MBs and inter-view collocated MBs which are obtained by the global disparity vectors (GDVs). By jointly using the MB motion homogeneity and coding mode complexity, they also proposed a view-adaptive ME and DE for low computational complexity MVC [22]. In that scheme, the coding

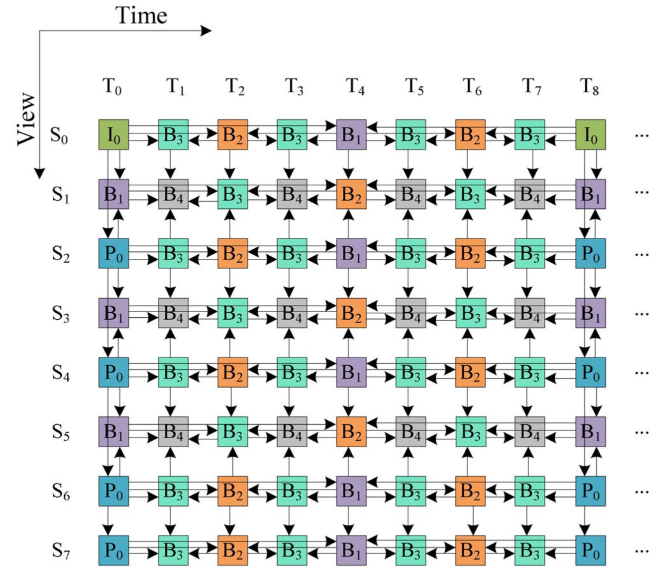


Fig. 1. MVC-HBP prediction structure in JMVC.

mode complexity and motion homogeneity are used for mode decision, ME/DE search window reduction and DE search, respectively. Because the GDV is derived by the DE, while there is no DE in the first view of MVC-HBP prediction structure, which results in the first view can not be optimized by these two algorithms.

In this paper, we propose an efficient ME/DE algorithm to further reduce the computational complexity in MVC, and maintain a comparable RD performance. The rest of this paper is organized as follows. The motivations and statistical analyses are given in Section II. Then, the details of the proposed efficient ME/DE algorithm are presented in Section III. Experiment results are shown in Section IV. Finally, Section V concludes this paper.

II. Motivations and Statistical Analyses

MVC-HBP prediction structure has been adopted in MVC standardization for its high coding efficiency by using hybrid inter-view and temporal prediction. Fig. 1 gives an example of the MVC-HBP prediction structure for an 8-view sequence (denoted by S_i, $i = 0, 1, \ldots, 7$) when the size of group of pictures equals to 8 (denoted by T_i, $i = 0, 1, \ldots, 7$). All views of the HBP prediction structure can be classified into two categories: even views (S_0, S_2, S_4, S_6) and odd views (S_1, S_3, S_5, S_7). In the even views, ME is employed to remove the temporal redundancies. In the odd views, both ME and DE are used, and the DE is performed to exploit the inter-view redundancies.

In MVC, there are seven different modes for the variable block-size mode selection and all these seven mode candidates are checked one by one sequentially. At last, the candidate mode with the minimum RD cost is selected as the optimal mode. We classify these candidate modes into two categories, inter mode set $\mathbf{S}$ and non-inter mode set $\mathbf{T}$, and they are defined as

$$\mathbf{S} = \{\text{B16×16, B16×8, B8×16, B8×8Frext, B8×8, SubB8×4, SubB4×8, SubB4×4}\}, \quad (1)$$

TABLE I
TEST CONDITIONS

Basis quantization parameter (BQP)	24,28,32,36
Size of group of pictures	12
Number of Reference Frames	2
Search Range	64
Max No. of Iterations for Bi-Prediction Search	4
Search Range for Iterations	8
Number of Frames to be Encoded	49

TABLE III
CPU TIME DISTRIBUTION IN ENCODING PROCESS (UNIT: %)

Views	BQP	Ballroom	Ballet	Exit	Average
		T/S	**T/S**	**T/S**	**T/S**
Even Views	24	3.20/96.80	2.73/97.27	0.04/99.96	1.99/98.01
	28	3.12/96.88	2.72/97.28	0.09/99.91	1.98/98.02
	32	3.01/96.99	2.71/97.29	0.07/99.93	1.93/98.07
	36	2.87/97.13	2.73/97.27	0.08/99.92	1.89/98.11
	Average	3.05/96.95	2.72/97.28	0.07/99.93	**1.95/98.05**
Odd Views	24	1.45/98.55	1.20/98.80	0.03/99.97	0.89/99.11
	28	1.38/98.62	1.23/98.77	0.08/99.92	0.90/99.10
	32	1.34/98.66	1.35/98.65	0.05/99.95	0.91/99.09
	36	1.39/98.61	1.46/98.54	0.06/99.94	0.97/99.03
	Average	1.34/98.66	1.31/98.69	0.05/99.95	**0.92/99.08**

$$\mathbf{T} = \{\text{DIRECT}, \text{SubDIRECT}, \text{I16MB}, \text{I8MB}, \text{I4MB,PCM}\}, \tag{2}$$

where the DIRECT mode in **T** represents either DIRECT mode in B frame or SKIP mode in P frame.

Let event **A** denote the candidate mode in **S** is selected as the optimal mode, **B** represents the candidate mode in **T** is selected as the optimal mode. $P(\mathbf{A})$, $P(\mathbf{B})$ are the probability of event **A** and event **B**, respectively.

We test four multiview video sequences with various motion activities to analyze the probabilities $P(\mathbf{A})$ and $P(\mathbf{B})$. "Ballroom" moves fast. "Ballet" is with medium motion. "Exit" and "Doorflowers" are with slow motion. The test conditions are listed in Table I. The statistical results of probabilities $P(\mathbf{A})$ and $P(\mathbf{B})$ are tabulated in Table II.

From Table II, it is observed that there is large number of MBs which select mode from **T** as their optimal mode. In the even views, the probability $P(\mathbf{B})$ is from 67.26% to 96.23%, and 84.17% on average. In the odd views, the $P(\mathbf{B})$ is from 65.86% to 95.01%, and 82.00% on average. Another observation on Table II is that there are 15.83% and 18.00% MBs encoded as the mode in set **S** in the even and odd views, respectively. These values demonstrate that most MBs (84.17%) are encoded as the mode in **T** and only a small number of MBs (15.83%) require the ME/DE.

The candidate modes in set **S** have a small probability to be selected as the optimal mode, however, it consumes the major proportion of total encoding time. According to the experimental results in Table III, the process of checking the candidate modes in **S** consumes 97.86% and 99.00% encoding time of total encoding time in the even and odd views, respectively. The encoding time is mainly consumed by the ME and DE, which are used to remove the temporal and inter-view redundancies, respectively. Hence, if the ME and DE process can be simplified, significant encoding time can be saved.

To further analyze the candidate modes in **T**, let event **C** represent DIRECT is selected as the optimal mode, **D** denotes that INTRA mode to be selected as the optimal mode. We give the conditional probabilities $P(\mathbf{C}|\mathbf{B})$ and $P(\mathbf{D}|\mathbf{B})$ in Table II. It means that events **C** and **D** are triggered given that **B** has occurred. From Table II, we can see that $P(\mathbf{C}|\mathbf{B})$ holds from 86.82% to 98.52%, 93.62% on average in the even views. In the odd views, $P(\mathbf{C}|\mathbf{B})$ is from 95.64% to 99.97%, 98.75% on average. On the contrary, $P(\mathbf{D}|\mathbf{B})$ holds a relatively small probability, 6.38% and 1.25% for the even and odd views, respectively.

Based on the conditional probability theory, we can obtain

$$P(\mathbf{BC}) = P(\mathbf{C}|\mathbf{B})P(\mathbf{B}), \tag{3}$$

and

$$P(\mathbf{BD}) = P(\mathbf{D}|\mathbf{B})P(\mathbf{B}). \tag{4}$$

From Eqs. (3), (4) and Table II, we can have that $P(\mathbf{BC})$ holds a larger probability than $P(\mathbf{BD})$ for about 70%. For instance, $P(\mathbf{B})$ and $P(\mathbf{C}|\mathbf{B})$ are 84.17% and 93.62%, 82.00% and 98.75%, in the even and odd views, respectively. Thus, according to Eq. (3), $P(\mathbf{BC})$ equals to 78.80% and 80.98% in the even and odd views, respectively. From Table II, $P(\mathbf{D}|\mathbf{B})$ equals to 6.38% and 1.25% on average in the even and odd views, respectively. Hence, based on Eq. (4), $P(\mathbf{BD})$ equals to 5.37% and 1.02% in the even and odd views, respectively. From Table III, it is observed that compared to the CPU time used in **S**, **T** consumes quite small encoding time, which are 2.14% and 1.00% in the even and odd views, respectively. Based on these values, we can conclude that INTRA modes in **T** have a little probability to be selected as the optimal mode and with quite low computational complexity. Therefore, we can have two findings. 1) $P(\mathbf{B})$ holds a large probability, but the process of encoding the candidate modes in **T** consumes little coding time. Thus, compared with early termination for the modes in **S**, it is more reasonable to perform early termination for the modes in **T**. 2) $P(\mathbf{BC})$ is much larger than $P(\mathbf{BD})$ and is the major part of the $P(\mathbf{B})$. Thus, we mainly do optimization for the DIRECT mode early termination.

III. PROPOSED EFFICIENT ME/DE ALGORITHM

A. Early DIRECT Mode Decision

Coded block pattern (CBP) is a syntax element in the encoded MB header that specifies six 8×8 blocks, including four luma blocks and two chroma blocks, for 4:2:0 subsampling [23]. If the CBP value equals to 0, it represents that all six 8×8 blocks don't have non-zero quantized transform coefficients. Hence, if the CBP value of DIRECT mode is equal to 0, the current MB is suitable for being encoded as DIRECT mode. As a result, the other candidate mode will be skipped and significant encoding time will be saved.

Four multiview video sequences ("Ballroom", "Exit", "Vassar" and "Doorflowers") are used to analyze the CBP values when the MB is encoded as DIRECT mode. The experimental conditions are tabulated in Table I. The statistical results are listed in Table IV. It is observed that when one MB is encoded as DIRECT mode, the probability of the CBP value

TABLE II
STATISTICAL RESULTS OF PROBABILITIES $P(\mathbf{A})$, $P(\mathbf{B})$, $P(\mathbf{C}|\mathbf{B})$, AND $P(\mathbf{D}|\mathbf{B})$ (UNIT: %)

	bQP	Even Views					Odd Views				
		Ballroom	Ballet	Exit	Doorflowers	Average	Ballroom	Ballet	Exit	Doorflowers	Average
$P(\mathbf{A})$	24	32.74	16.55	24.07	11.75	21.28	34.14	26.34	21.19	13.10	23.69
	28	27.86	13.41	18.11	7.44	16.71	29.75	20.40	17.64	8.78	19.14
	32	24.20	11.02	14.80	5.07	13.77	25.61	17.16	14.65	6.47	15.97
	36	21.01	8.74	12.77	3.67	11.55	21.32	14.55	11.95	4.99	13.20
	Average	26.45	12.43	17.44	6.98	**15.83**	27.71	19.61	16.36	8.34	**18.00**
$P(\mathbf{B})$	24	67.26	83.45	75.93	88.25	78.72	65.86	73.66	78.81	86.90	76.31
	28	72.14	86.59	81.89	92.56	83.30	70.25	79.60	82.36	91.22	80.86
	32	75.80	88.98	85.20	94.93	86.23	74.39	82.84	85.35	93.53	84.03
	36	78.99	91.26	87.23	96.33	88.45	78.68	85.45	88.05	95.01	86.80
	Average	73.55	87.57	82.56	93.02	**84.17**	72.30	80.39	83.64	91.67	**82.00**
$P(\mathbf{C}\|\mathbf{B})$	24	87.12	88.95	86.82	95.70	89.65	97.05	96.42	95.64	99.54	97.16
	28	88.87	93.50	91.41	97.23	92.75	98.31	98.74	98.05	99.84	98.74
	32	94.53	95.95	93.80	98.06	95.59	99.07	99.58	99.03	99.92	99.40
	36	94.27	97.57	95.69	98.52	96.51	99.67	99.85	99.35	99.97	99.71
	Average	91.20	93.99	91.93	97.38	**93.62**	98.53	98.65	98.02	99.82	**98.75**
$P(\mathbf{D}\|\mathbf{B})$	24	12.88	11.05	13.18	4.30	10.35	2.95	3.58	4.36	0.46	2.84
	28	11.13	6.50	8.59	2.77	7.25	1.69	1.26	1.95	0.16	1.27
	32	5.47	4.05	6.20	1.94	4.42	0.93	0.42	0.97	0.08	0.60
	36	5.73	2.43	4.31	1.48	3.49	0.33	0.15	0.65	0.03	0.29
	Average	8.80	6.01	8.07	2.62	**6.38**	1.48	1.35	1.98	0.18	**1.25**

TABLE IV
PROBABILITIES OF CBP EQUALS TO ZERO WHEN THE MB IS ENCODED AS DIRECT MODE (UNIT: %)

Views	Sequence	BQP=24	BQP=28	BQP=32	BQP=36	Average
Even Views	Ballroom	94.73	96.92	93.96	99.82	96.36
	Exit	97.31	98.60	99.34	99.91	98.79
	Vassar	98.32	99.17	99.82	99.93	99.31
	Doorflowers	98.52	99.31	99.92	99.94	99.42
	Average	97.22	98.50	98.26	99.90	**98.47**
Even Views	Ballroom	90.23	95.43	93.67	97.02	94.09
	Exit	92.17	97.33	97.15	97.73	96.10
	Vassar	93.56	98.05	98.07	98.12	96.95
	Doorflowers	95.22	98.64	98.69	98.59	97.79
	Average	92.80	97.36	96.90	97.87	**96.23**

equals to 0 is quite large, about 98.47% and 96.23% in the even and odd views, respectively. Based on this observation, we set $CBP_{DIRECT} = 0$ as the early DIRECT mode decision condition. This means that after encoding one MB as DIRECT mode, if its CBP value is equal to 0, then the following variable block-size ME/DE and intra prediction could be skipped. Two multiview video sequences ("Ballroom" and "Breakdancer") are used to verify the RD performance of this assumption. The results of RD performance are presented in Fig. 2.

Fig. 2 shows the RD curves of "Ballroom" and "Breakdancer" between the original JMVC and the assumption that the early DIRECT mode decision condition with $CBP_{DIRECT} = 0$, respectively. In the legend, the solid lines denote the RD curves of the original JMVC; and the dash lines represent the RD curves of the CBP based early DIRECT mode decision; the first item represents the algorithm; the second item denotes the multiview video test sequence and the last represents the view of the test sequence. For example, "JMVC&Ballroom&Even" indicates the even views of the "Ballroom" sequence are encoded by the original JMVC algorithm. When compared to the original JMVC, the RD performance of the assumption degrades dramatically. This indicates that no matter how much computational complexity is saved by the $CBP_{DIRECT} = 0$, it is not suitable to be used

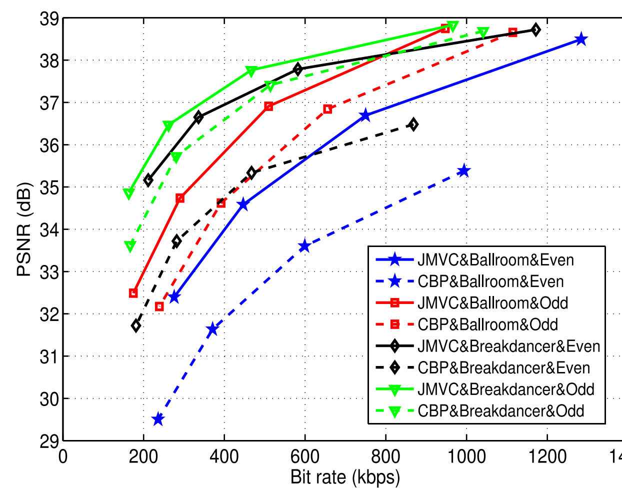


Fig. 2. RD comparison between the early DIRECT mode decision with $\mathrm{CBP_{DIRECT}} = 0$ and the original JMVC.

as the condition of early DIRECT mode decision. Therefore, we need to design a stricter criteria which combines CBP with other conditions in order to achieve a better RD performance.

In the RD based mode decision process, the mode with the minimum RD cost will be selected as the optimal mode. If the RD cost of one mode is quite different from others, then it can be classified into one particular category. In [24], the linear correlation of RD cost between DIRECT mode and the candidate modes in $\mathbf{S}$ have been achieved. It also concluded that the RD cost of DIRECT mode is quite different from the modes in $\mathbf{S}$. However, the RD costs of modes in $\mathbf{S}$ are very similar to each other. Based on this characteristic, all candidate modes can be classified into two categories, DIRECT and $\mathbf{S}$, respectively. Therefore, DIRECT mode can be early determined by

$$J_{cur} \leq \eta \cdot J_{Aveg}, \tag{5}$$

where J_{cur} represents the RD cost of DIRECT mode, J_{Aveg} denotes the average RD cost of DIRECT which is selected

TABLE V
PROBABILITY OF THE INITIAL SEARCH POINT SELECTED AS THE BEST POINT (UNIT: %)

ME/DE	Sequence	BQP=24	BQP=28	BQP=32	BQP=36	Average
ME	Ballroom	89.04	88.85	88.86	89.01	88.94
	DooFlowers	94.30	93.97	93.91	94.15	94.08
	Exit	89.06	88.51	88.49	88.80	88.72
	Vassar	92.29	92.22	91.86	91.94	92.08
	Average	91.17	90.89	90.78	90.98	**90.95**
DE	Ballroom	66.67	66.68	68.21	69.48	67.76
	DooFlowers	91.46	92.29	93.16	95.33	93.06
	Exit	66.93	66.10	67.37	69.88	67.57
	Vassar	82.00	85.51	85.42	86.68	84.90
	Average	76.77	77.65	78.54	80.34	**78.32**

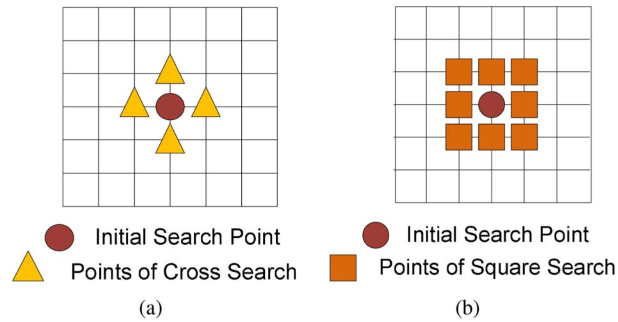


Fig. 3. Early ME/DE termination strategy. (a) Cross search. (b) Square search.

as the optimal mode in the previous frame. If J_{Aveg} is not available, which means the previous frame is encoded as intra frame, the DIRECT mode early termination is not activated and the original full search algorithm is used.

In order to achieve a good trade-off between the RD performance and computational complexity reduction, CBP and RD cost are jointly used to early determine the DIRECT mode. Hence, the current MB is encoded as DIRECT mode if

$$CBP_{DIRECT} = 0 \ \&\& \ J_{cur} \leq \eta \cdot J_{Aveg}, \tag{6}$$

where $\&\&$ is the logical AND operation. Eq. (6) means the current MB is encoded as DIRECT mode when these two conditions $CBP_{DIRECT} = 0$ and $J_{cur} \leq \eta \cdot J_{Aveg}$ are satisfied. η is a regulating parameter, which is used to trade off the RD performance and computational complexity reduction. The determination of η will be discussed in the experimental section.

B. Initial Search Point-Based ME/DE Early Termination

In multiview video sequences, there are a large number of static and slow motion activity MBs. For these MBs, they have a very large probability to select the initial search point (ISP) as the best point in ME/DE process. Four multiview video sequences ("Ballroom", "Exit", "Vassar" and "Doorflowers") are used to analyze the probability of the ISP to be selected as the best point. The test conditions are listed in Table I. The full search is used for the ME/DE search. Table V shows the statistical results for the inter MBs that the ISP is the best point.

From Table V, we can see that there are 88.49% to 94.30% ISPs selected as the best point in the ME process. In the DE process, there are 66.10% to 95.33% ISPs selected as the best point. There are 90.95% and 78.32% on average ISPs selected as the best search point in the ME and DE process, respectively. Another observation on Table V is that the probabilities of ISPs selected as the best search point are different between ME and DE process. Compared to the ME, the probability decreases about 12% in the DE process. This is because that 1) the similarity among temporally consecutive frames is larger than the similarity among inter-view frames; 2) the DE usually has a global disparity. Hence, we should design different early termination conditions for the ME and DE by considering the trade-off between the RD performance and computational complexity reduction.

Based on above analyses and the best point is center-biased [9], a cross search with four points is used to early terminate the ME process. Fig. 3(a) gives an example of the cross search. The candidates of the cross search are defined as

$$\overrightarrow{MV_{cross}} = \{(MV_x,MV_y)|(MV_x,MV_y) = (x \pm 1,y),\ (x,y \pm 1)\}, \tag{7}$$

where (x, y) denotes the MV of the ISP. If the ISP is the best point among the five checked points, the ME process will be early terminated.

In order to early terminate the DE process, an eight points square search is employed to detect whether the ISP is the best search point. Fig. 3(b) gives an example of the square search. The candidates of the square search are defined as

$$\overrightarrow{MV_{square}} = \big\{(MV_x,MV_y)|(MV_x,MV_y) = (x \pm 1,y), (x,y \pm 1),(x \pm 1,y \pm 1)\big\}, \tag{8}$$

where (x, y) denotes the MV of the ISP. If the ISP is the best point among the nine checked points, the DE process will be early terminated.

To evaluate the efficiency of the proposed early ME/DE termination algorithm, the Hit Rate (HR) is adopted, which is defined as

$$HR(\mathbf{A}|\mathbf{B}) = \frac{N(\mathbf{A}|\mathbf{B})}{N(\mathbf{B})} \times 100\%, \tag{9}$$

where $HR(\mathbf{A}|\mathbf{B})$ denotes the HR; the event $\mathbf{A}$ represents the ISP is selected as the final best search point; $\mathbf{B}$ denotes the early termination condition; $N(\cdot)$ represents the number of blocks in corresponding event. Four multiview video sequences ("Ballroom", "Breakdancer", "Exit" and "Vassar") are tested, and the results are tabulated in Table VI.

From Table VI, it is observed that the proposed early ME/DE termination algorithm can achieve the HR from 93.74% to 99.47%, 97.63% on average in the even views. In the odd views, the HR of the proposed method is from 89.30% to 95.22%, 91.73% on average. Compared to the other three sequences ("Ballroom", "Exit" and "Vassar"), the HR of "Breakdancer" has a little decrease, because its video content moves very fast. These values demonstrate that the proposed algorithm can early terminate the ME and DE efficiently.

TABLE VI
HIT RATE OF THE PROPOSED EARLY ME/DE TERMINATION (UNIT: %)

ME/DE	Sequence	BQP=24	BQP=28	BQP=32	BQP=36	Average
ME	Ballroom	97.77	97.71	97.69	97.65	97.71
	Breakdancer	93.74	94.58	95.27	95.87	94.87
	Exit	98.34	98.73	98.83	98.72	98.66
	Vassar	98.96	99.34	99.45	99.47	99.31
	Average	97.20	97.59	97.81	97.93	**97.63**
DE	Ballroom	92.29	91.38	90.76	90.88	91.33
	Breakdancer	89.30	90.91	92.24	93.44	91.47
	Exit	90.48	90.48	90.25	89.98	90.30
	Vassar	92.03	93.53	94.57	95.22	93.84
	Average	91.03	91.58	91.96	92.38	**91.73**

C. Optimal Stopping Theory-Based Search Range Reduction Strategy

Optimal stopping problems are defined by two objects, a sequence of random variables and a sequence of reward functions. The problem is how to maximize the expected reward based on these values of the sequentially checked variables [25]. In video coding, some processes (such as mode decision, ME/DE and so on) can be optimized by the optimal stopping theory to achieve a better computational complexity reduction and RD performance. Therefore, optimal stopping theory is an ideal choice for optimizing these encoding processes.

Based on the optimal stopping problem named duration problem [26], Zhao *et al.* modeled a duration problem for mode decision in H.264/SVC [6]. In this model, according to the probabilities of each mode, all candidate modes are sorted in a descending order,

$$p_i > p_j, \ \forall i, j \in [1, N], \ \ i < j, \tag{10}$$

where p_n, $n = 1, 2, \ldots, N$, denote the sorted probabilities of each candidate. N denotes the number of total candidates. With the sorted probabilities in Eq. (10), the optimal stopping position K is at

$$K = \min\left\{ k \geq 1 : \ \sum_{i=1}^{k} p_i \sum_{j=k}^{N} \frac{1}{\sum_{m=1}^{j} p_m} > N - k + \frac{1}{2} \right\}. \tag{11}$$

This duration model obtains a good performance in decision accuracy and encoding time saving. By regarding sub-search windows as the random variables and encoding time saving as the reward functions, the mode decision duration problem model can be extended to search window reduction in the ME and DE process.

The size of search window is directly correlated with the computational complexity of the ME and DE process. If the search window size can be reduced by using some kinds of strategies, the computational complexity of ME and DE will be saved. In this paper, based on [6], a duration problem is modeled for the ME and DE search window size reduction. We assume that there are N candidate sub-search windows, denoted as X_n, $n = 1, 2, \ldots, N$; and the corresponding probabilities of each sub-search window are predicted as p_n, $n = 1, 2, \ldots, N$. Based on Eq. (10), all candidate sub-search windows are sorted in a descending order. Ultimately, the optimal stopping position K is computed according to Eq. (11).

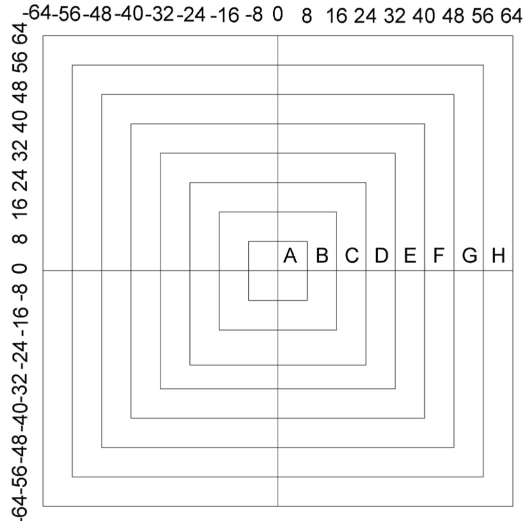


Fig. 4. Illustration of the eight sub-search windows, $R = 64$ and $W = 8$.

Generally, two questions need to be considered in the duration problem of the ME/DE search window size reduction. Firstly, how to arrange the sub-search windows; Secondly, when to stop. In this paper, we split the original search range R into equal-sized sub-search windows with interval W. Then, the probabilities of each sub-search window are computed adaptively based on the best search points distribution in the previous frame. Fig. 4 gives an example of sub-search windows splitting where R and W are equal to 64 and 8, respectively. According to Eq. (10), all sub-search windows are sorted in a descending order based on their corresponding probabilities, and make the optimal stopping decision using Eq. (11). We test the multiview video sequence "Ballroom" with S_0 as an example, the BQP and ME/DE search mode are set to 24 and full search, respectively. R and W are set to 64 and 8, respectively. We obtain that the probabilities of the eight sub-search windows are 0.4406, 0.1509, 0.0937, 0.1123, 0.0818, 0.0376, 0.0357 and 0.0474; the optimal stopping is $K = 5$ after sorting all the sub-search windows in a descending order. This means that block matching search will be performed in sub-search windows A, B, D, C, E one by one to locate the best search point. Note that the probability of each sub-search window in current frame will be adaptively updated according to the best search points distribution in the pervious frame.

D. Search Strategies of Locating the Best Matching Block

Based on the sorted sub-search windows and the optimal stopping position K, the block matching search strategies will be performed sequentially in the sorted sub-search windows. In conventional ME algorithms (such as three step search, four step search and so on), the square search pattern is a basic search pattern to locate the best matching block. In this paper, the square search pattern is also used as the basic search pattern. A multiple eight points square search is performed in the sub-search windows to locate the best matching block. The eight points square search pattern is defined as

$$\Omega_{EPS} = \{(i,j) | (i,j) = (\pm 2,0), \ (0, \pm 2), \ (\pm 2, \pm 2)\}, \tag{12}$$

The multiple eight points square search is performed by extending the eight points square search pattern with a scale factor, and it is defined as

$$\begin{aligned}\overrightarrow{MV_{MEPS}} = \Big\{\big(MV_x, MV_y\big)|MV_x &= x + \varphi \cdot x' + W \cdot (k-1),\\ MV_y &= y + \varphi \cdot y' + W \cdot (k-1),\ (x', y') \in \Omega_{EPS},\\ \varphi &= 1,\ 2,\ 3,\ 4\Big\},\end{aligned} \tag{13}$$

where $(x,\ y)$ denotes the MV of the ISP; W represents the interval value of sub-search window; k denotes the n-th sub-search window.

At last, in order to achieve the best RD performance, a four points cross search is used to refine the search result. The cross search is performed on the best point of the multiple eight points square search, this process will be recursively repeated until the best point is located at the center of the cross search pattern.

E. Overall Algorithm

Based on above analysis, the overall algorithm is summarized as follows.

Step 1. Compute the average RD cost of DIRECT mode which is selected as the best mode in the previous encoded frame, then record this average RD cost as J_{aveg}.

Step 2. Encode the current MB as DIRECT mode, and compute its CBP and RD cost, denote them as CBP_{DIRECT} and J_{cur}, respectively. If Eq. (6) is satisfied, go to Step 9; else go to Step 3.

Step 3. Check the current MB whether to do ME or DE, if it belongs to ME, go to Step 4; otherwise go to Step 5.

Step 4. Perform a four points cross search on the ISP, the cross search points candidates are defined in Eq. (7). If the best search point which has the minimum RD cost among the five checked points is the ISP, go to Step 9; else go to Step 6.

Step 5. Perform an eight points square search on the ISP, the candidate search point set is defined in Eq. (8). If the best search point which has the minimum RD cost among the nine checked points is the ISP, go to Step 9; else go to Step 6.

Step 6. According to the best points distribution in previous frame, the probability of each sub-search window is computed adaptively. Then, based on Eqs. (10) and (11), an optimal stopping position K is computed.

Step 7. Based on the sorted sub-search windows and the optimal stopping position K, a multiple eight points square search is performed in the sub-search windows according to Eq. (13).

Step 8. Perform a recursively four points cross search on the best point which is obtained from Step 7 to refine the search result.

Step 9. Return the MV information. Go back to Step 2 to encode the next MB.

IV. Experimental Results

A. Regulating Parameter η Determination

In order to trade off the computational complexity saving (CCS) and RD performance, a regulating parameter η is used in Eq. (6). Larger η will lead to larger CCS at the cost of more RD degradation. On contrary, smaller η leads to lower RD degradation, however, smaller CCS can be achieved. With the purpose of achieving the best overall RD performance, the early DIRECT mode decision associates with the early ME/DE termination, and the ME/DE search window reduction are jointly considered to obtain the value of η. To explore the best value of η, η equals to 0.5, 0.6, 0.7, 0.9, and 1.0 are tested respectively. Two multiview video sequences ("Ballet" and "Doorflowers") are used to determine η. Table VII shows the average encoding results in terms of BDPSNR, BDBR and CCS. BDPSNR and BDBR are computed according to [27]. CCS is computed as

$$CCS = \frac{T_\eta - T_j}{T_j} \times 100\%, \tag{14}$$

where T_η represents the total encoding time of the proposed early DIRECT mode decision with different η; T_j denotes the total encoding time of the original JMVC8.0.

In order to control RD performance, we set the upper bounds for BDPSNR degradation and BDBR increase to 0.1 dB and 3.0%, respectively. From Table VII, we can see that in the even views, when η is equal to 1.0, the average encoding results of BDPSNR, BDBR and CCS are -0.084 dB, 2.50% and -55.61%, respectively. These three values are quite acceptable. Therefore, in even views, η is set to 1.0. In the odd views, when η is equal to 0.7, it achieves the best RD performance and computational complexity. The average BDPSNR, BDBR and CCS are -0.088 dB, 2.86% and -62.43%, respectively. Hence, in the odd views, η is set to 0.7.

B. HR and RD Performance of the Proposed Early DIRECT Mode Decision

In order to evaluate the HR and RD performance of the proposed early DIRECT mode decision algorithm, three multiview video sequences ("Ballroom", "Breakdancer" and "Exit") are tested. We compare the HR of early DIRECT mode decision with single CBP and CBP combined with RD cost. The results of HR and RD performance are presented in Table VIII and Fig. 5, respectively. In Table VIII, the HR is computed according to Eq. (9), where the event **A** represents the MB is encoded as DIRECT mode; **B** denotes the early termination condition. The CBP and CBP&RD represent the early DIRECT mode decision with single CBP and CBP combined with RD cost, respectively.

From Table VIII, we can see that the proposed combined early DIRECT mode decision algorithm can achieve the HR from 92.27% to 99.81% in the even views. In the odd views, it can achieve the HR from 97.78% to 99.84%. Compared to the cases that BQP with 28, 32 or 36, the HR of BQP with 24 has a little decrease. This is because more MBs are encoded as inter modes when BQP equals to 24. From these results, we can see that the HR of the proposed combined algorithm

TABLE VII
ENCODING PERFORMANCE WITH DIFFERENT η

η	Sequence	Even Views			Odd Views		
		BDPSNR (dB)	BDBR (%)	CCS (%)	BDPSNR (dB)	BDBR (%)	CCS (%)
0.5	Ballet	-0.088	2.22	-39.05	-0.074	1.94	-54.35
	Doorflowers	-0.069	2.45	-37.49	-0.100	3.78	-60.15
	Average	-0.078	2.33	-38.27	-0.087	2.86	-57.25
0.6	Ballet	-0.087	2.20	-40.55	-0.072	1.86	-56.25
	Doorflowers	-0.068	2.43	-42.50	-0.098	3.72	-62.34
	Average	-0.078	2.32	-41.52	-0.085	2.79	-59.30
0.7	Ballet	-0.088	2.22	-44.48	-0.072	1.88	-59.09
	Doorflowers	-0.069	2.44	-44.10	-0.103	3.85	-65.78
	Average	-0.078	2.33	-44.29	**-0.088**	**2.86**	**-62.43**
0.8	Ballet	-0.090	2.29	-48.09	-0.076	2.00	-62.26
	Doorflowers	-0.070	2.49	-49.02	-0.111	4.15	-67.87
	Average	-0.080	2.39	-48.55	-0.093	3.07	-65.07
0.9	Ballet	-0.092	2.33	-52.30	-0.074	1.92	-65.81
	Doorflowers	-0.071	2.54	-51.90	-0.119	4.50	-70.89
	Average	-0.082	2.43	-52.10	-0.096	3.21	-68.35
1.0	Ballet	-0.094	2.38	-54.83	-0.081	2.13	-68.34
	Doorflowers	-0.073	2.62	-56.39	-0.122	4.63	-73.87
	Average	**-0.084**	**2.50**	**-55.61**	-0.101	3.38	-71.10

TABLE VIII
HIT RATE OF THE SINGLE CBP-BASED EARLY DIRECT MODE DECISION AND THE CBP COMBINED WITH RD COST-BASED EARLY DIRECT MODE DECISION (UNIT: %)

Views	BQP	Ballroom	Ballroom	Ballroom	Average
		CBP/CBP&RD	CBP/CBP&RD	CBP/CBP&RD	CBP/CBP&RD
Even Views	24	78.16/93.92	75.59/92.27	80.35/95.67	78.03/93.95
	28	84.05/97.98	80.77/96.67	86.63/98.86	83.82/97.84
	32	86.77/99.02	83.95/98.79	88.58/99.60	86.43/99.14
	36	88.55/99.18	87.52/99.77	89.69/99.81	88.59/99.59
	Average	84.38/97.53	81.96/96.88	86.31/98.49	**84.22/97.63**
Odd Views	24	92.40/97.78	86.81/98.30	91.96/97.80	90.32/97.96
	28	95.46/98.89	87.43/99.10	95.06/99.26	92.65/99.08
	32	95.10/99.35	88.22/99.59	95.18/99.75	92.83/99.56
	36	94.83/99.71	90.34/99.84	95.12/99.84	93.43/99.80
	Average	94.45/98.93	88.15/99.21	94.33/99.16	**92.31/99.10**

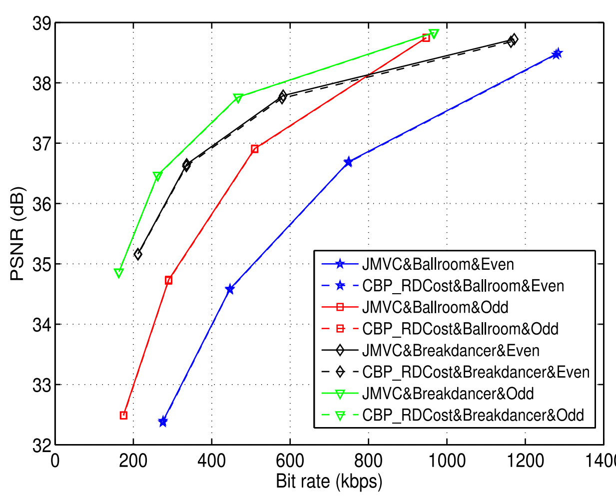


Fig. 5. RD performance between the proposed CBP combined with RD cost-based early DIRECT mode decision and the original JMVC.

can reach up to 97.63% and 99.10% in the even and odd views, respectively. For the single CBP based early DIRECT mode decision, it achieves the HR 84.22% and 92.31% in the even and odd views, respectively. Hence, compared to single CBP based algorithm, the HR of the proposed combined algorithm increases 13.41% and 6.79% in the even and odd views, respectively. Fig. 5 shows the RD curves of the proposed early DIRECT mode decision method and the original JMVC. In the legend, the solid lines represent the RD curves of the original JMVC, and the dash lines denote the proposed CBP combined with RD cost based early DIRECT mode decision; the first item represents the algorithm, where JMVC and CBP_RDCost denote the original JMVC algorithm and the proposed CBP combined with RD cost based early DIRECT mode decision algorithm, respectively; the second item denotes the test sequence; the last represents the view of the test sequence. For example, "JMVC&Ballroom&Even" indicates the even views of the "Ballroom" are encoded by the original JMVC algorithm. It is observed that the proposed algorithm achieves a quite similar RD performance with the original JMVC. From the HR and RD performance, we can conclude that the proposed CBP combined with RD cost algorithm can early determine the DIRECT mode efficiently.

C. Comparison of BDPSNR, BDBR, and CPU Time

In order to evaluate the efficiency of the proposed overall algorithm, MVC reference software JMVC8.0 [20] is

TABLE IX
SUMMARY OF ENCODING RESULTS

Views	Seuqence	Resolution	YangH.264 VS JMVC			ShenCSVT VS JMVC			ShenTB VS JMVC			Proposed VS JMVC		
			BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)	BDPSNR (dB)	BDBR (%)	TS (%)
Even Views	Ballroom	640×480	-0.063	1.59	-50.38	-0.022	0.57	-27.44	0.000	0.00	0.00	-0.109	2.77	-64.65
	Exit	640×480	-0.083	2.87	-62.18	-0.034	1.27	-25.56	0.000	0.00	0.00	-0.070	2.53	-63.19
	Vassar	640×480	-0.017	0.74	-70.78	-0.007	0.31	-19.28	0.000	0.00	0.00	-0.029	1.39	-65.88
	Breakdancer	1024×768	-0.051	2.29	-54.84	-0.015	0.69	-17.89	0.000	0.00	0.00	-0.092	4.33	-69.92
	Lovebird1	1024×768	-0.109	3.02	-71.67	-0.001	0.04	-9.46	0.000	0.00	0.00	-0.025	0.68	-58.75
	Champanger	1280×960	-0.176	3.98	-77.04	-0.042	0.95	-26.93	0.000	0.00	0.00	-0.045	1.00	-68.91
	Dog	1280×960	-0.227	5.89	-64.53	-0.011	0.28	-16.05	0.000	0.00	0.00	-0.051	1.28	-65.26
	PoznanHall2	1920×1080	-0.216	7.47	-68.99	-0.005	0.31	-6.68	0.000	0.00	0.00	-0.038	1.89	-62.72
	PoznanStreet	1920×1080	-0.091	3.11	-67.48	-0.002	0.07	-14.05	0.000	0.00	0.00	-0.081	2.83	-69.55
	UndoDancer	1920×1080	-0.172	4.87	-66.02	-0.001	0.03	-8.09	0.000	0.00	0.00	-0.080	2.33	-51.78
	Average		**-0.121**	**3.58**	**-65.39**	**-0.014**	**0.45**	**-17.14**	**0.000**	**0.00**	**0.00**	**-0.062**	**2.10**	**-64.06**
Odd Views	Ballroom	640×480	-0.148	3.81	-63.09	-0.050	1.35	-50.18	-0.056	1.49	-41.45	-0.087	2.26	-65.22
	Exit	640×480	-0.140	5.11	-70.29	-0.040	1.49	-52.84	-0.047	1.76	-54.35	-0.091	3.45	-68.40
	Vassar	640×480	-0.048	2.35	-73.33	-0.007	0.36	-55.46	-0.025	1.21	-67.83	-0.022	1.07	-67.92
	Breakdancer	1024×768	-0.120	5.17	-61.74	-0.012	0.64	-16.70	-0.046	2.09	-28.22	-0.104	4.68	-68.28
	Lovebird1	1024×768	-0.133	3.61	-75.06	-0.006	0.18	-40.99	-0.016	0.44	-60.24	-0.026	0.69	-60.24
	Champanger	1280×960	-0.279	7.18	-76.94	-0.010	0.27	-42.70	-0.015	0.41	-59.04	-0.071	1.79	-75.50
	Dog	1280×960	-0.299	9.27	-69.52	0.011	-0.33	-30.56	-0.044	1.37	-44.26	-0.015	0.43	-57.97
	PoznanHall2	1920×1080	-0.358	11.32	-70.01	-0.003	0.02	-23.18	0.005	-0.60	-19.47	-0.031	1.27	-65.75
	PoznanStreet	1920×1080	-0.218	8.46	-72.70	-0.004	0.08	-16.70	-0.038	1.58	-56.45	-0.031	1.25	-70.37
	UndoDancer	1920×1080	-0.308	7.05	-70.99	0.009	-0.17	-23.53	0.031	-0.78	-17.27	-0.078	2.26	-56.44
	Average		**-0.205**	**6.33**	**-70.37**	**-0.011**	**0.39**	**-35.28**	**-0.025**	**0.90**	**-44.86**	**-0.056**	**1.92**	**-65.61**

used as the software platform. The test conditions are listed in Table I. The interval size W of our proposed algorithm is set to 8. The hardware platform is Intel Core 2 Duo CPU E5800 @ 3.16GHz and 3.17GHz, 4.00GB RAM with Microsoft Windows 7 64-bit operating system.

We compare the coding performance of the the proposed algorithm with three recent algorithms, YangH.264 [28], ShenTB [21] and ShenCSVT [22] in terms of BDPSNR, BDBR, and total encoding CPU time saving. The ME/DE search method in YangH.264, ShenTB, ShenCSVT and JMVC8.0 is TZSearch [20]. Ten multiview video test sequences (“Ballroom”, “Exit”, “Vassar”, “Breakdancer”, “Champagne”, “Dog”, “Lovebird1”, “PoznanHall2”, “PoznanStreet”, and “UndoDancer”.) are used to compare the encoding performance. The experimental results are compared and summarized in Table IX. In this table, BDPSNR and BDBR are calculated according to [27]. TS denotes the total encoding CPU time saving. They are defined as

$$TS = \frac{T_p - T_o}{T_o} \times 100\%, \qquad (15)$$

where T_p represents the total encoding CPU time of algorithm p, $p \in \{$YangH.264, ShenCSVT, ShenTB, Proposed$\}$; T_o denotes the total encoding CPU time of original JMVC8.0.

From Table IX, for the even views, it can be observed that the YangH.264 can reduce the encoding time from 38.17% to 77.70%, 65.39% on average; meanwhile, the BDPSNR between the YangH.264 and the original JMVC is from −0.227 dB to −0.017 dB, −0.121 dB on average, and the BDBR between the YangH.264 and the original JMVC is from 0.74% to 7.47%, 3.58% on average. Even though the YangH.264 can reduce the encoding time significantly, however, the bit-rate increases dramatically. These values reflect that the YangH.264 can not make the mode decision efficiently in the even views. The ShenCSVT can reduce the total encoding time from 6.11% to 29.95%, 17.14% on average; the BDPSNR between the ShenCSVT and the original JMVC is from −0.042 dB to −0.001 dB, −0.014 dB on average, and the BDBR between the ShenCSVT and the original JMVC is from 0.03% to 1.27%, 0.45% on average. It can be seen that the total encoding time reduction of the ShenCSVT is quite limited, 17.14% on average. This is because the first view of MVC-HBP prediction structure can not be optimized by the ShenCSVT due to the fact that no DE is in the first view, the inter-view collocated MB can not be located. The ShenTB is only applied to the odd views, there is no optimization for the even views. The proposed algorithm can reduce the total encoding time from 44.93% to 73.69%, 63.66% on average; the BDPSNR between the proposed algorithm and the original JMVC is from −0.109 dB to −0.029 dB, −0.062 dB on average, and the BDBR between the proposed algorithm and the original JMVC is from 0.68% to 4.33%, 2.10% on average. Compared with the ShenCSVT, the proposed algorithm achieves a similar RD performance, while more than 56% encoding time can be saved by the proposed algorithm. These values demonstrate that the proposed algorithm can efficiently reduce the computational complexity in the even views while maintaining a comparable RD performance.

For the odd views, the YangH.264 can reduce the computational complexity from 53.91% to 78.90%, 70.37% on average; the BDPSNR between the YangH.264 and the original JMVC is from −0.358 dB to −0.048 dB, −0.205 dB on average; the BDBR between the YangH.264 and the original JMVC is from 2.35% to 11.32%, 6.33% on average. We can see that the bit-rate of YangH.264 increases significantly, this is because the inter-view prediction and properties of disparity field are not considered. The ShenCSVT can reduce the computational complexity from 9.00% to 59.84%, 41.35% on average; meanwhile, the BDPSNR and BDBR between the ShenCSVT and the original JMVC change from −0.050 dB to 0.011 dB, −0.011 dB on average and from −0.33% to 1.49%, 0.39%

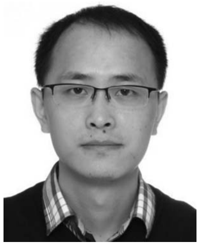
on average, respectively. The ShenTB can reduce the total encoding time from 11.78% to 69.23%, 44.86% on average, meanwhile, the BDPSNR and BDBR between the ShenTB and the original JMVC change from −0.056 dB to 0.031 dB, −0.025 dB on average, and from −0.78% to 2.09%, 0.90% on average, respectively. However, the computational complexity reduction for video sequence "Breakdancer" is quite limited, about 28.22% on average. This is because "Breakdancer" moves quite fast, most of MBs need to do variable block-size ME and DE. From the results of BDPSNR and BDBR, we can see that the RD performance degradation of the ShenTB is negligible. For the proposed algorithm, we can see that the computational complexity can be reduced from 55.18% to 77.61%, 65.61% on average; meanwhile, the BDPSNR between the proposed algorithm and the original JMVC is from −0.104 dB to −0.015 dB, −0.056 dB on average; the BDBR between the proposed algorithm and the original JMVC is from 0.43% to 4.68%, 1.92% on average. Comprehensively evaluating the encoding performance of the proposed algorithm, we can conclude that the proposed algorithm achieves quite better RD performance than the YangH.264. Compared to the ShenCSVT and the ShenTB, the proposed algorithm achieves a similar RD performance to them; meanwhile, there are more than 46.86% and 37.63% computational complexity saved by the proposed algorithm.

V. Conclusion

In this paper, an efficient ME and DE algorithm is proposed by jointly using the CBP, RD cost, ISP, and optimal stopping theory. Firstly, based on the characteristics of the CBP and RD cost, an early DIRECT mode decision algorithm is proposed. Then, according to the characteristics of the ISP, an early ME/DE termination algorithm is proposed. If the ME/DE early termination is not satisfied, the ME/DE search window will be reduced by applying the optimal theory. At last, two block matching search strategies are proposed to predict the best search point for the ME/DE. Experimental results show that our proposed algorithm can yield a quite promising coding performance in terms of RD performance and computational complexity. Especially, the proposed algorithm can be applied to not only the odd views but also the even views. In addition, the proposed algorithm can be applied into MV/3D-HEVC for reducing the computational complexity.

References

[1] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Document H.264 and ISO/IEC Standard 14496-10, Apr. 2013.
[2] J. Lei, K. Feng, M. Wu, S. Li, and C. Hou, "Rate control of hierarchical B prediction structure for multi-view video coding," *Multimedia Tools Appl.*, vol. 72, no. 1, pp. 825–842, Sep. 2014.
[3] G. Tech, K. Wegner, Y. Chen, M. Hannuksela, and J. Boyce, "MV-HEVC draft text 9," in *Proc. 9th Meeting Joint Coll. Team 3D Video Coding Ext. Develop. (JCT-3V) Doc. JTC3V-I1004*, Sapporo, Japan, Jul. 2014, pp. 3–9.
[4] G. Tech, K. Wegner, Y. Chen, and S. Yea, "3D-HEVC draft text 5," in *Proc. 9th Meeting Joint Coll. Team 3D Video Coding Ext. Develop. (JCT-3V) Doc. JTC3V-I1001*, Sapporo, Japan, Jul. 2014, pp. 3–9.
[5] H. Zeng, C. Cai, and K.-K. Ma, "Fast mode decision for H.264/AVC based on macroblock motion activity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 491–499, Apr. 2009.
[6] T. Zhao, S. Kwong, H. Wang, and C.-C. J. Kuo, "H.264/SVC mode decision based on optimal stopping theory," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2607–2618, May 2012.
[7] S. Lee, "Fast motion estimation based on adaptive search range adjustment and matching error prediction," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 805–811, May 2009.
[8] L.-M. Po *et al.*, "Novel directional gradient descent searches for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1189–1195, Aug. 2009.
[9] Z. Pan, S. Kwong, and Y. Zhang, "A multiple hexagon search algorithm for motion and disparity estimation in multiview video coding," in *The Era of Interactive Media*, J. Jin, C. Xu, and M. Xu, Eds. New York, NY, USA: Springer, 2012, pp. 113–122.
[10] D. Jeon and H. Park, "An adaptive reference frame selection method for multiple reference frame motion estimation in the H.264/AVC," in *Proc. 16th Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 629–632.
[11] D. Jun and H. Park, "An efficient priority-based reference frame selection method for fast motion estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, pp. 1156–1161, Aug. 2010.
[12] J. Seo and K. Sohn, "Early disparity estimation skipping for multi-view video coding," *EURASIP J. Wireless Commun. Netw.*, vol. 32, pp. 1–12, Feb. 2012.
[13] N. Purnachand, L. N. Alves, and A. Navarro, "Improvements to TZ search motion estimation algorithm for multiview video coding," in *Proc. 19th Int. Conf. Syst. Signals Image Process. (IWSSIP)*, Vienna, Austria, Apr. 2012, pp. 388–391.
[14] Z. Cui, G. Jiang, S. Yang, and C. Wu, "A new fast motion estimation algorithm based on the loop-epipolar constraint for multiview video coding," *Signal Process. Image Commun.*, vol. 27, no. 2, pp. 172–179, Feb. 2012.
[15] F. Wang, H. Zeng, Q. Shen, and S. Du, "Efficient early direct mode decision for multi-view video coding," *Signal Process. Image Commun.*, vol. 28, no. 7, pp. 736–744, Aug. 2013.
[16] H. Wang, Y. Heng, T. Zhao, and B. Xiao, "Fast mode decision based on optimal stopping theory for multiview video coding," in *Proc. Adv. Multimedia Model. (MMM)*, Huangshan, China, Jan. 2013, pp. 176–186.
[17] Y. Zhang, S. Kwong, G. Jiang, X. Wang, and M. Yu, "Statistical early termination model for fast mode decision and reference frame selection in multiview video coding," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 10–23, Mar. 2012.
[18] Y. Zhang, S. Kwong, G. Jiang, H. Wang, and M. Yu, "Efficient multi-reference frame selection algorithm for hierarchical B frames in multiview video coding," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 15–23, Mar. 2011.
[19] Y. Zhang, S. Kwong, L. Xu, and G. Jiang, "DIRECT mode early decision optimization based on rate distortion cost property and interview correlation," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 390–398, Jun. 2013.
[20] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, *Joint Draft 8.0 on Multiview Video Coding*, ISO/IEC Standard JTC1/SC29/WG11, ITU-T Document SG16 Q.6, and JVT Document JVT-AB204, Jul. 2008.
[21] L. Shen, Z. Liu, S. Liu, Z. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. Broadcast.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.
[22] L. Shen, Z. Liu, T. Yao, Z. Zhang, and P. An, "View-adaptive motion estimation and disparity estimation for low complexity multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 925–930, Jun. 2010.
[23] B.-Y. Chen and S.-H. Yang, "Using H.264 coded block patterns for fast inter-mode selection," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Hannover, Germany, Apr. 2008, pp. 721–724.
[24] S. Hu, T. Zhao, H. Wang, and S. Kwong, "Fast inter-mode decision based on rate-distortion cost characteristics," in *Proc. 11th Pac. Rim Conf. Multimedia (PCM)*, vol. 2. Shanghai, China, Sep. 2010, pp. 145–155.
[25] T. S. Ferguson. (Mar. 12, 2014). *Optimal Stopping and Applications*. [Online]. Available: http://www.math.ucla.edu/ tom/Stopping/Contents.html
[26] T. S. Ferguson, J. P. Hardwick, and M. Tamaki, "Maximizing the duration of owning a relatively best object in strategies for sequential search and selection in real time," *Contemp. Math.*, vol. 125, pp. 37–57, Mar. 1992.
[27] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves*, ITU-T VCEG Document VCEG-M33, Apr. 2001.
[28] S.-H. Yang, B.-Y. Chen, and K.-H. Wang, "H.264 fast inter-mode selection based on coded block patterns," *IEICE Trans. Inf. Syst.*, vol. E92-D, no. 6, pp. 1324–1327, Jun. 2009.



Zhaoqing Pan received the B.S. degree in computer science and technology from Yancheng Normal University, Yancheng, China, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2009 and 2014, respectively. In 2013, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA, for six months. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include video compression and 3-D video processing.

Yun Zhang (M'12) received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2014, he was a Post-Doctoral Research Associate and a Visiting Scholar with the Department of Computer Science, City University of Hong Kong, Hong Kong. In 2010, he became an Assistant Professor with the Shenzhen Institutes of Advanced Technology, CAS, where he has been an Associate Professor since 2012. His research interests include video compression, 3-D video processing, and visual perception.

Sam Kwong (M'93–SM'04–F'13) received the B.S. and M.S. degrees in electrical engineering from the State University of New York at Buffalo and the University of Waterloo, Waterloo, ON, Canada, in 1983 and 1985, respectively, and the Ph.D. degree from the University of Hagen, Germany, in 1996. From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada. He joined Bell Northern Research Canada, as a Scientific Staff Member. In 1990, he became a Lecturer with the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor with the Department of Computer Science. His research interests are video and image coding and evolutionary algorithms.