Fast Macroblock Mode Selection Algorithm for B Frames in Multiview Video Coding

Mei Yu, Ping He, Zongju Peng, Yun Zhang, Yuehou Si and Gangyi Jiang
Faculty of Information Science and Engineering, Ningbo University
Ningbo, 315211 - China
[e-mail: jianggangyi@126.com]
*Corresponding author: Gangyi Jiang

Received November 29, 2010; revised January 11, 2011; revised January 20, 2011;
accepted January 23, 2011; published February 28, 2011

Abstract

Intensive computational complexity is an obstacle of enabling multiview video coding for real-time applications. In this paper, we present a fast macroblock (MB) mode selection algorithm for B frames which are based on the computational complexity analyses between the MB mode selection and reference frame selection. Three strategies are proposed to reduce the coding complexity jointly. First, the temporal correlation of MB modes between current MB and its temporal corresponding MBs is utilized to reduce computational complexity in determining the optimal MB mode. Secondly, Lagrangian cost of SKIP mode is compared with that of Inter16×16 modes to early terminate the mode selection process. Thirdly, reference frame correlation among different Inter modes is exploited to reduce the number of reference frames. Experimental results show that the proposed algorithm can promote the encoding speed by 3.71~7.22 times with 0.08dB PSNR degradation and 2.03% bitrate increase on average compared with the joint multiview video model.

Keywords: Multiview video coding, macroblock mode selection, low complexity, hierarchical B picture, Inter mode

---

This research was supported by the Natural Science Foundation of China (grant 60872094, 60832003, 61071120), the projects of Chinese Ministry of Education (grant 200816460003). It is also sponsored by K.C.Wong Magna Fund in Ningbo University.

DOI: 10.3837/tiis.2011.02.010

1. Introduction

The packet networks including internet offer an intrinsic diversity for media distribution, novel communication infrastructures use network diversity to extend their reach at low cost [1]. For media streaming transmission with network diversity, some technologies were developed such as distributed resource management [2], systematic scheduling [3][4], and so on, in which the video coding techniques with hierarchical layers of different importance had been used to reduce the redundancies of monoview video source. Different from traditional monoview video systems which only provide users with single view and passive video, multiview video is able to provide arbitrary views of scenes, interactivity view changing, realistic and unique three dimensional perception [5][6][7]. Multiview video can be applied to new generation consumer electronics applications, such as free viewpoint television, three dimensional television, virtual reality and digital video communication, thus it has attracted lots of interests from researchers to industrial communities. Since the amount of data is proportional to the number of views, multiview video have to be effectively compressed for its real-time and interactive functionalities.

The simplest method for multiview video coding (MVC) is to encode each view independently using a state-of-the-art video codec, such as H.264/AVC. However, except the temporal redundancies among monoview video, multiview video also contains strong inter-view redundancies since all cameras capture the same scene from slightly different view angles simultaneously. These inter-view redundancies, together with temporal redundancies, can be exploited by temporal/inter-view prediction techniques. Accordingly, various efficient MVC prediction structures [8][9][10][11][12] had been proposed. Hierarchical B Pictures (HBP) prediction structure is beneficial for media streaming with network diversity. Merkle et al. utilized HBP for exploiting spatio-temporal correlations in the same view and inter-view correlations among different views to achieve higher compression efficiency [9]. Chunga et al. proposed a novel prediction structure, called three-dimensional HBP, which can efficiently reduce horizontal inter-view redundancies, vertical inter-view redundancies, and temporal redundancies in multiview videos [10]. Zhang et al. proposed a multi-modal MVC scheme on the basis of dynamic correlation analysis to achieve not only better random accessibility, but also good performance in compression efficiency, low memory requirement, complexity and view scalability [11]. Park et al. proposed view-temporal prediction structures that can be adjusted to various characteristics of general multiview video by separating them into temporal and view prediction structures [12].

In order to reduce the coding complexity, many fast algorithms for inter-frame predication were proposed for monoview video coding based on H.264/AVC [13][14][15][16][17]. Lin et al. proposed a layer-adaptive intra/inter mode decision algorithm and a motion search scheme for the hierarchical B-frames in scalable video coding (SVC) with combined coarse-grain quality scalability and temporal scalability to speed up the H.264/SVC encoder [13]. Yeh et al also presented a fast mode decision algorithm that speeds up the SVC encoding process through probabilistic analysis [14]. The mode of the enhancement layer is first predicted by statistical analysis. Afterward, Bayesian theorem is utilized to detect whether the prediction mode of the current macroblock is the best or not. The mode is further predicted and refined by the Markov process. Shen et al. focus on adaptive and fast multi-frame selection algorithm to speed up the searching procedure for multiple reference frames [15]. Nisar et al. proposed a robust scheme for initial motion vector prediction based spatial and temporal predictors [16].

Grecos et al. utilized a set of skip mode conditions for P and B slices, two heuristics that reduce the cardinality of the Inter mode are set (Inter/Intra mode prediction and the monotonic property of Rate-Distortion (RD) cost functions) to save computational time [17]. Wang et al. early terminated mode decision and Motion Estimation (ME) by detecting all-zero blocks to lower complexity of mono-view video coding [18].

However, these monoview fast algorithms can hardly be used for MVC effectively, because the prediction structures of MVC additionally adopts disparity prediction compensation to eliminate inter-view redundancies in multiview video. In MVC standardization, a Joint Multiview Video Model (JMVM) [19] was developed by the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). In JMVM, HBP prediction structure [9] was adopted in the MVC standardization draft since it achieves superior compression efficiency. Moreover, as in H.264/AVC, variable macroblock (MB) size modes are also adopted in JMVM. For B frames, there are nine modes, including SKIP/DIRECT, Inter16×16, Inter16×8, Inter8×16, Inter8×8Frext and Inter8×8, Intra16×16, Intra8×8 and Intra4×4. In addition, each 8×8 block can further be spitted into smaller blocks, i.e. 8×8, 8×4, 4×8 and 4×4. These modes are probed among all temporal and inter-view reference frames to find the optimal MB mode with the best Rate Distortion (RD) performance. The mode with the minimal RD cost is then selected as the best coding mode. Accordingly, high compression efficiency is achieved at the expense of extremely large computational complexity, which is an obstacle for putting multiview video coding into practical real-time application.

To efficiently reduce heavy computational complexity of MVC, some fast MVC algorithms had been presented in [20][21][22][23][24]. Shen et al. proposed a fast Disparity Estimation (DE) and ME algorithm based on motion homogeneity to reduce MVC computational complexity [20]. The basic idea of the method is to utilize the spatial property of motion field in prediction where DE and variable size ME are needed, and only in these regions DE and variable size ME are enabled. Li et al. proposed a fast DE and ME algorithms which limit an estimable range and reduce the number of reference frames [21]. Peng et al. proposed a hybrid fast MB mode selection algorithm by utilizing the correlations of the MB modes among neighboring views [22]. However, the correlations of MB modes within view were not exploited. Cernigliaro et al. proposed a Fast Mode Decision (FMD) algorithm for MVC with the help of a depth map, which reduces computational complexity based on the analysis of the homogeneity of depth map [23]. In [24], a content-aware prediction algorithm with inter-view mode decision is proposed for MVC. The computation for ME in most views was reduced by the sharing and reusing of the coded information, such as rate-distortion cost, coding modes and motion vectors. In [25], a fast Inter mode decision scheme for the Skip mode and the Inter sub modes is proposed. Based on the RD cost correlation between neighboring views, and the RD cost of different textural segmentation regions, a predecision of the Skip mode is introduced to reduce other modes' estimation. In addition, the estimated direction of Inter sub modes is predicted based on the optimal direction of the Inter16×16 mode.

In JMVM, the encoding process of B frames in HBP prediction structure consumes most of the encoding time. In this paper, we present a fast mode selection algorithm for B frames to significantly reduce computational complexity. The rest of this paper is organized as follows: Section 2 depicts the characteristic of HBP prediction structure and strategy of mode selection. Section 3 analyses the computational complexity of MVC. Then, three mode decision techniques are presented in Section 4 to jointly reduce the MVC complexity. In Section 5, the performance of three proposed techniques is separately as well as jointly analyzed via MVC experiments. Finally, our work is concluded in Section 6.

2. HBP Prediction Structure and Mode Selection in MVC

2.1 HBP Prediction Structure in MVC

Fig. 1 shows an example of HBP prediction structure in JMVM with eight views. In the figure, the horizontal direction denotes the consecutive time and the vertical direction denotes the individual view. I frames are Intra coded; B and P frames are Inter frames. B frames are classified into four layers which are expressed by B1, B2, B3 and B4 in descendent order. In general, the higher layer B frames are referenced by the lower layer B frames. Naturally, B frames with higher layer are less important than the B frames in the lower layer in terms of RD performance. Hence, Quantization Parameter (QP) of B frames with higher layer is often higher than that of B frames in the lower layer during the encoding process. The frames of all views, from T_0 to T_7, constitute a Group-Of-Pictures (GOP) of multiview video sequence. The GOP-length in **Fig. 1**, the number of frames along the temporal axis, is 8. For the convenience of synchronization and random access, each group of pictures starts with I frame. View S_0 is the basic view in which the frames do not use any inter-view prediction. In views S_1, S_3 and S_5, inter-view correlation is exploited to encode B frames. In views S_2, S_4 and S_6, B frames do not use any inter-view prediction. The last view S_7 is similar to those of odd views, except only one neighboring view for inter-view prediction.

Most frames in HBP prediction structure are B frames, that is, the total number of B frames is far more than that of I and P frames. For example, B frames occupy 92.19%, 94.79% and 95.83% of all frames when the GOP-length is 8, 12 and 15, respectively. It can be seen that the percentage of B frames becomes even higher as the GOP-length increases. On the other hand, encoding B frame consumes more computational time than encoding I and P frames because of the bi-directional ME/DE search. Therefore, to reduce the total complexity of the MVC encoder efficiently, it is reasonable to focus on B frames.

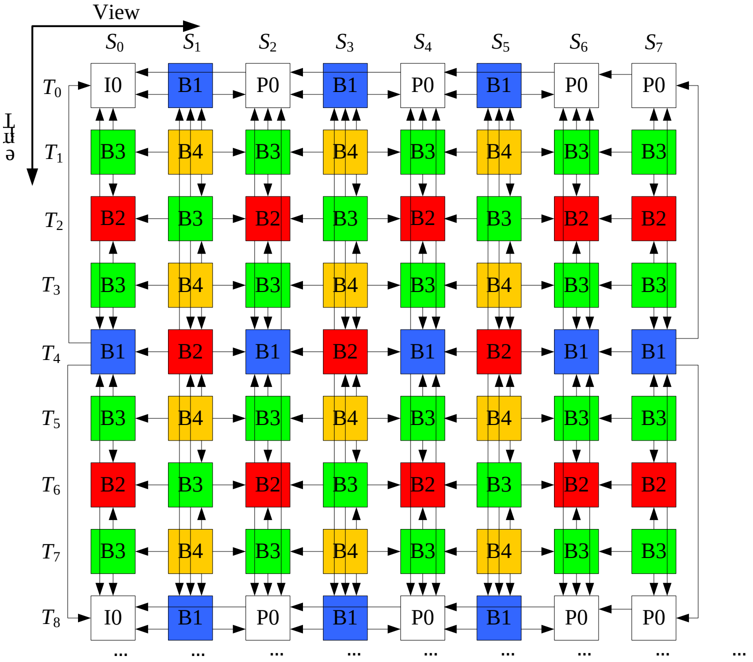


Fig. 1. HBP prediction structure [9].

2.2 The Strategy for The Best MB Mode Selection in MVC

In JMVM, MB mode selection is done by minimizing the Lagrangian cost function

$$J(s,c,MODE\,|\,QP,\lambda_{MODE}) = SSD(s,c,MODE\,|\,QP) + \lambda_{MODE} R(s,c,MODE\,|\,QP)\,, \qquad (1)$$

where $J(s,c,MODE|QP, \lambda_{MODE})$ denotes RD cost when candidate $MODE$ is selected as the mode of current MB, s and c are the original and reconstructed signals, respectively. QP is the macroblock quantization parameter. $R(s,c,MODE|QP)$ reflects the number of bits produced for header(s) (including MODE indicators), motion vector(s), and transform coefficients. $SSD(s,c,MODE|QP)$ is sum of square differences that measures the distortion between the original and the reconstructed macroblocks, and it is calculated by

$$SSD(s,c,MODE\,|\,QP) = \sum_{x=1}^{M}\sum_{y=1}^{N} \left| s[x,y] - c[x,y,MODE\,|\,QP] \right|^2\,, \qquad (2)$$

where M and N denote width and height of a block, respectively. λ_{MODE} is the Lagrangian multiplier for mode decision and given by

$$\lambda_{MODE} = 0.85 \times 2^{(QP-12)/3}\,. \qquad (3)$$

In addition, the best motion/disparity vectors (MV/DV) are computed by minimizing the following equation in the ME/DE search.

$$J(\mathbf{m},\lambda_{MOTION}) = SAD(s,c(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m}-\mathbf{p})\,, \qquad (4)$$

where $\mathbf{m}=(m_x,m_y)^T$ denotes motion/disparity vector, $\mathbf{p}=(p_x,p_y)^T$ denotes prediction for motion/disparity vector. The rate term $R(\mathbf{m}\text{-}\mathbf{p})$ represents the number of bits for coding MV/DV residuals. $SAD(\mathrm{s},\mathrm{c}(\mathbf{m}))$ is the sum of the absolute differences between the original and reconstructed signals, and λ_{MOTION} is the Lagrangian multiplier, they are defined as

$$\lambda_{MOTION} = \sqrt{\lambda_{MODE}}\,, \qquad (5)$$

$$SAD(s,c(\mathbf{m})) = \sum_{x=1}^{M}\sum_{y=1}^{N} \left| s(x,y) - c(x-m_x, y-m_y) \right|\,. \qquad (6)$$

To obtain an optimal MB mode, a full mode search method is adopted in JMVM. **Fig. 2** shows the available MB modes of B frames which include a SKIP mode (SKIP mode of B frame is encoded like SKIP mode of P frame), three Intra modes, and four MB-level Inter modes. An 8×8 block can further be split into three kinds of sub-MB blocks. To achieve the best RD performance, each MB mode should be tested and the RD cost of each candidate mode is calculated by Eq. (1). In Intra coding process, an Intra 16×16 mode has four prediction directions, while Intra4×4 and Intra8×8 modes have nine prediction directions. In Inter coding process, to obtain the best MV/DV, $J(\mathbf{m},\lambda_{MOTION})$ in Eq.(4) of each Inter mode has to be calculated, and $SAD(s, c(\mathbf{m}))$ with respect to each mode is calculated in all reference frames. **Fig. 3** gives the pseudo codes of Inter frame encoding process in JMVM, in which there are three loop levels: ME/DE, frame selection and mode decision. Additionally, the ME/DE can be fractional-pel accuracy which enlarges the computational complexity even more.

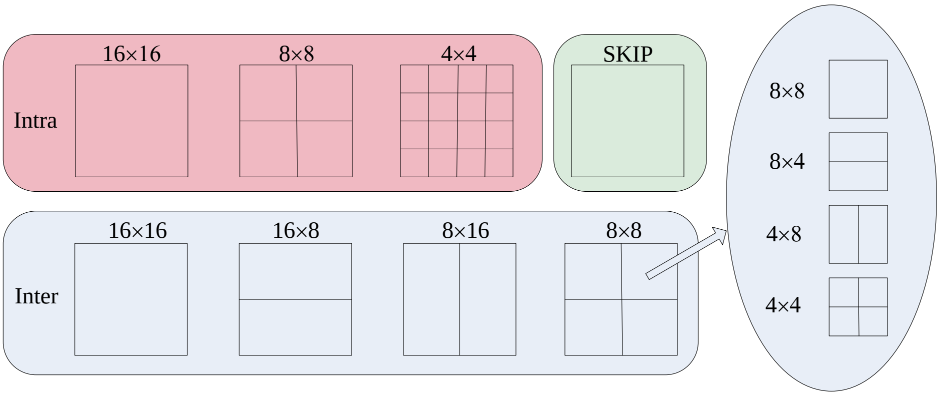


Fig. 2. MB modes of B frames in JMVM.

```
FOR all different Inter modes
{       FOR all reference frames
        {       DO motion estimation/disparity estimation
                {     FOR all possible points in ME/DE window
                      {       Calculate $J(\mathbf{m},\lambda_{MOTION})$ for each point;
                              Store the corresponding MV/DV;
                      }
                      Select the best MV/DV with minimal $J(\mathbf{m},\lambda_{MOTION})$;
                }
                Calculate RD cost with respect to current reference frame;
        }
        Select reference frame associated with minimal $J(s,c,MODE|QP,\lambda_{MODE})$;
}
Select the optimal Inter mode with minimal RD cost;
```

Fig. 3. Pseudo codes of multiple reference ME/DE.

3. Analyses on Computational Complexity of MVC

In this section, computational complexity of MVC is discussed. The computational complexity of MB mode selection and reference frame selection is analyzed in subsection 3.1 and 3.2.

3.1 Analyses on Computational Complexity of MB Mode Selection

To analyze the MB mode selection complexity, four multiview video sequences are utilized to perform statistical exploration experiments on JMVM 7.0 [19]. The experiments in this paper are performed on PC with 3.2GHz CPU, 3.25GB DDR2 memory. During the encoding process, the optimal mode of each MB and the computational time of each MB mode are recorded.

Fig. 4 shows the statistical results of the optimal MB modes selection of B frames. Obviously, MB mode distribution has an uneven characteristic in frames. Most of MBs are encoded with SKIP mode. Next to the SKIP mode, Inter16×16 mode is the second most [26]. The proportions of other modes, such as Inter16×8, Inter8×16, Inter8×8, Inter8×8Frext and Intra, are much less than those of SKIP and Inter16×16 modes. **Table 1** shows the computational time percentage of each mode used in one MB. SKIP mode only needs 0.52% computational time of the full mode search. For other Inter modes, such as Inter16×16, Inter16×8, Inter8×16, Inter8×8 and Inter8×8Frext, consume most of computational time. Especially, Inter8×8 mode consumes 54.37% computational time of the full mode search.

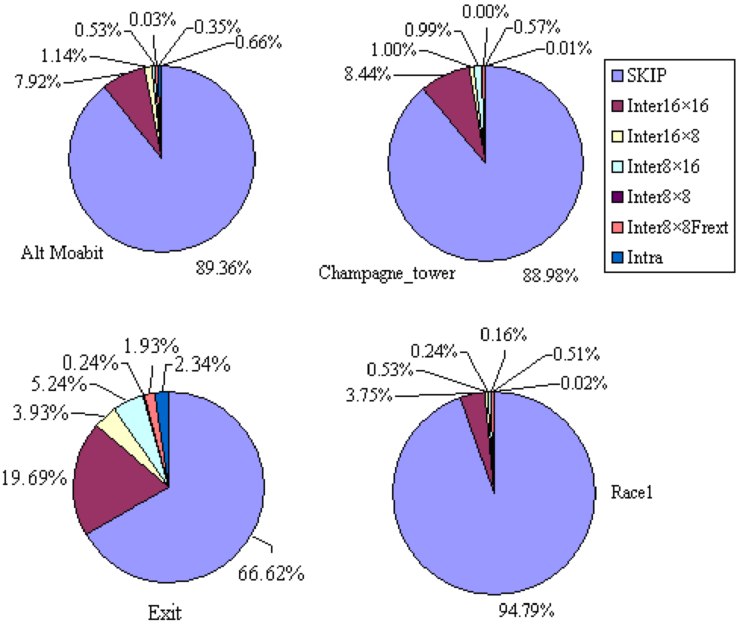


Fig. 4. Statistical results of MB mode selection for B frames.

Table 1. Computational time allocation of MB modes in each MB for 1000 times

	SKIP	**16×16**	**16×8**	**8×16**	**8×8**	**8×8 Frext**	**Intra**	**Total**
T (μs)	23	402	411	450	2384	496	219	4384
Ratio(%)	0.52	9.18	9.37	10.27	54.37	11.30	4.99	100

It is not balance in proportions of MB modes as well as the consuming time. The SKIP mode which has the most percentages in quantitative distribution takes negligible computational time in comparison with other inter modes. According to **Fig. 4** and **Table 1**, much unnecessary computational load can be lessened during the mode selection process. Let t_{fs} be the time of coding an MB by using full mode search, s be the total number of MB in a frame, the maximum saving ratio β_{max} can be calculated by

$$\beta_{max} = \frac{st_{fs} - (s_1 t_{\text{SKIP}} + s_2 t_{16\times16} + s_3 t_{16\times8} + s_4 t_{8\times16} + s_5 t_{8\times8} + s_6 t_{8\times8\text{Frext}} + s_7 t_{\text{Intra}})}{st_{fs}}, \tag{7}$$

$$s = s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7, \tag{8}$$

$$t_{fs} = t_{\text{SKIP}} + t_{16\times16} + t_{16\times8} + t_{8\times16} + t_{8\times8} + t_{8\times8Frext} + t_{Intra}, \tag{9}$$

s_1, s_2, s_3, s_4, s_5, s_6 and s_7 are the numbers of MBs encoded with SKIP, Inter16×16, Inter16×8, Inter8×16, Inter8×8, Inter8×8Frext and Intra modes, respectively. t_{SKIP}, $t_{16\times16}$, $t_{16\times8}$, $t_{8\times16}$, $t_{8\times8}$,

$t_{8\times8\text{Frext}}$ and t_{Intra} denote computational time with respect to SKIP, Inter16×16, Inter16×8, Inter8×16, Inter8×8, Inter8×8Frext and Intra modes.

Compared with the full mode search method, for the MBs encoded with SKIP mode, if SKIP mode is directly selected, the time saving ratio β_{SKIP} will be

$$\beta_{\text{SKIP}} = \frac{s_1(t_{fs} - t_{\text{SKIP}})}{st_{fs}}. \tag{10}$$

Similarly, for MBs encoded with SKIP or Inter16×16 mode, the time saving ratio β_{DI} will be

$$\beta_{\text{DI}} = \frac{s_1(t_{fs} - t_{\text{SKIP}}) + s_2(t_{fs} - t_{16\times16})}{st_{fs}}. \tag{11}$$

Since dominant distributional proportion and negligible encoding time proportion of SKIP mode, β_{max} may approaches to 100%. **Table 2** shows β_{max}, β_{SKIP} and β_{DI} of different test sequences. For Race1, β_{max} is 98.94%. It is clear that most of the encoding time can be saved at the same encoding RD performance. β_{SKIP} and β_{DI} are lower than β_{max}. However, while β_{DI} is nearly equal to β_{max}, the most saving time is mainly contributed by SKIP and Inter16×16 modes. Thus, it is necessary to select quickly SKIP and Inter16×16 modes.

Table 2. β_{max}, β_{SKIP} and β_{DI} of different test sequences.

	Race1	**Exit**	**Alt Moabit**	**Champagne tower**
β_{max} (%)	98.94	96.48	98.56	98.50
β_{SKIP} (%)	94.30	66.28	88.91	88.53
β_{DI} (%)	97.70	84.17	96.10	96.19

Table 3. Coding time allocation of reference frames in each inter mode.

	16×16	**16×8**	**8×16**	**8×8**	**8×8Frext**
T(4refs) (μs)	402	411	450	2384	496
T(1ref) (μs)	88	85	96	483	122

Table 4. Time saving ratio from MB mode and reference frames optimization in each sequence.

	Race1	**Exit**	**Alt Moabit**	**Champagne tower**
ΔT_{save} (%)	99.38	98.87	99.30	99.31
$\Delta T'_{save}$ (%)	98.14	90.91	96.75	96.65
$\Delta T''_{save}$ (%)	99.05	95.67	98.67	98.69

3.2 Analyses on Computational Complexity of Reference Frames Selection

MVC employs multiple reference frames technique for inter modes. It can achieve the best visual quality and coding efficiency for videos with repetitive motions, uncovered backgrounds, and textural areas. Unfortunately, it resulted in intensive computational complexity. **Table 3** shows the encoding time of inter modes with different number of reference frames, where T(4refs) is the searching time for different modes in all four reference frames, and T(1ref) is the searching time for different modes in the most suited one of four reference frames. If we adopt the full mode search method and each inter mode is only using one reference frame, near 80% coding time can be saved.

Furthermore, more computational time can be saved by combining with the result in section III.A. Let ΔT_{save} , $\Delta T'_{save}$ and $\Delta T''_{save}$ be time saving ratio of combing **Table 3** and β_{max}, β_{SKIP} and β_{DI} . They are calculated in **Table 4**. Thus, it is significant to decrease the number of reference frames while keeping the similar video coding performances.

According to the aforementioned analyses of complexity of the encoding process, we can summarize it as follows

1) The MB mode distribution in MVC is uneven. Generally, for B frame, the proportion of MBs encoded with SKIP mode has the largest ratio and more than 50%, Inter16×16 mode ranks the second. The proportions of other modes, such as Inter16×8, Inter8×16, and Inter8×8, are often less than 20%.
2) The coding time for each mode is quite different. Compared with total encoding time, computational time of SKIP mode is only 0.52% which is negliable, while the computational time of Inter8×8 mode is more than 50% because its four sub-MB modes of Inter8×8 are probed one by one. Additionally, the computational time of Inter16×16, Inter16×8, Inter8×16, and Inter8×8Frext are similar.
3) Combining 1) with 2), if MBs whose best modes is SKIP or Inter16×16 mode are early terminated, most of the coding time could be saved.
4) The computational time is mainly used in the Inter modes searching due to ME and DE using multiple reference frames. It is possible to speedup the encoding process by reducing the number of reference frames for Inter modes.

4. The Proposed Fast MB Mode Selection Algorithm

The proposed fast mode selection algorithm includes three strategies, Fast Decision of SKIP mode and Inter16×16 mode (FDSI), Fast Decision for SKIP mode (FDS) and Correlation based Reference Frames Selection (CRFS).

4.1 FDSI strategy

In H.264/AVC CODEC, B frames have three prediction directions, that is, the forward, backward and bi-direction. H.264/AVC uses two queues to store the two reference frames lists, list0 and list1. For the frame S_0T_6 in **Fig. 5**, its store-order is 0 and 12 in list0, but 12 and 0 in list1. As mentioned above, there are strong temporal/inter-view correlations in multiview videos. For the convenience of narration, we denote the temporal corresponding MBs as TMBs. Based on our observations, it is found that the MB modes of TMBs are similar especially in the still and background areas among temporal consecutive frames. Let *ref*0 and *ref*1 be the nearest temporal reference frames in list0 and list1, $\mathbf{m}$ be a MB mode set, $f(\mathbf{m})$ be the number of MBs in current frame whose optimal modes and the optimal modes of the TMBs in *ref*0 and *ref*1 all belong to $\mathbf{m}$, and $g(\mathbf{m})$ be the number of MBs in current frame whose TMBs in *ref*0 and *ref*1 are with optimal modes in $\mathbf{m}$. Then, a mode correlation factor between the current frame, *ref*0 and *ref*1, $k(\mathbf{m})$, can be defined and calculated by

$$k(\mathbf{m}) = f(\mathbf{m}) / g(\mathbf{m}) . \qquad (12)$$

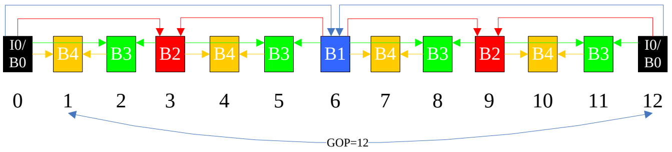


Fig. 5. Basic structures, GOP-length=12(I0/B0).

Since most MBs are encoded with SKIP and Inter16×16 modes, we set **m** as {SKIP, Inter16×16} to analyze the MB correlation. **Fig. 6** shows the analytical results of five B frames in view 0 of the four test sequences at different time instants. In the figure, the x-axis denotes current frame, and the y-axis is $k(\mathbf{m})$. $k(\mathbf{m})$ basically determine the MB correlation because of these two modes' majority quantitative distribution. $k(\mathbf{m})$ approaches 1. Although there are mass movement areas in Exit sequence, $k(\mathbf{m})$ is still larger than 96%. That is, if the MB modes of the TMBs are SKIP or Inter16×16, MB mode of the current MB may be SKIP or Inter16×16.

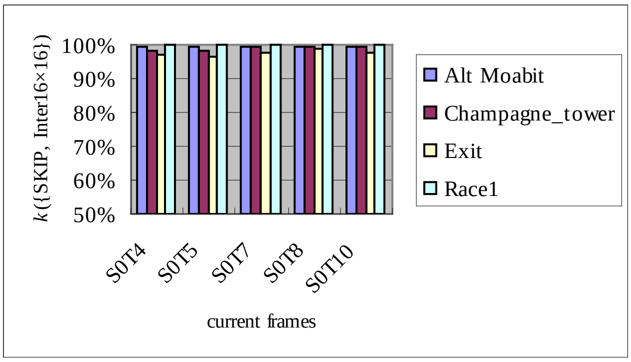


Fig. 6. Analyses for correlation of MB modes in TMBs.

Based on the analyses, we propose a strategy of FDSI, which is described as following

> If *ref*0 and *ref*1 are both B frames, and the optimal modes of the corresponding MBs in *ref*0 and *ref*1 are SKIP or Inter16×16, only SKIP or Inter16×16 modes are probed during the encoding process of current MB.

The FDSI strategy makes use of the mode correlation between the current MB and MBs at corresponding locations in *ref*0 and *ref*1 to determine the optimal mode of current MB. It can then avoid unnecessary mode searching process. For example, if there are 800 MBs of B frame in Exit sequence (640×480, 1200MBs in total) satisfying the above condition, theoretically, 60.19% of computational time can be saved.

However, FDSI strategy is only suitable for encoding the B frames which *ref*0 and *ref*1 are non-anchor B frames. Here we define these non-anchor frames as α-B frames. If *ref*0 or *ref*1 is a non-α-B frame, we present another approach to reduce computational complexity in next subsection.

4.2 FDS strategy in non-α-B frames

As described in subsection 2.2, Lagrangian cost of each mode is determined by SSD and R. Since both SKIP and Inter16×16 modes are with the same block size, Lagrangian cost difference between these two modes is mainly decided by R when the difference of $SSD(s,c,\text{SKIP}|QP)$ and $SSD(s,c,\text{Inter16×16}|QP)$ are very small. Especially, SKIP mode does not consume the bits of motion/disparity vectors, so $R(s,c,\mathrm{SKIP}|QP)$ is less than $R(s,c,\mathrm{Inter}16\times16|QP)$. Thus, Lagrangian cost of SKIP mode is usually less than that of Inter16×16 mode for MB with small *SSD*. Let $Rd(\mathrm{SKIP})$ and $Rd(\mathrm{Interl}6\times16)$ be Lagrangian costs of SKIP mode and Interl6×16 mode, respectively. A strategy of FDS is proposed for the SKIP mode in non-α-B frames, which is described as following.

> Search SKIP and Inter16×16, if the condition $Rd(\mathrm{SKIP})\leq Rd(\mathrm{Interl}6\times16)$ is satisfied, SKIP mode is regarded as the best coding mode of current MB and the searching process is early terminated.

In the following exploration experiments, we analyze the accuracy of FDS strategy. **Table 5** lists the statistical results of SKIP mode in 264 B frames of eight views in various test sequences. The second column in the table denotes the total number of MBs in which SKIP mode is selected as their best mode with FDS strategy. The third and fourth columns list the number (percentage) of MBs with correct and incorrect mode selection, respectively. As for Race1, Exit, Alt Moabit and Champagne tower test sequences, the correct percentages of SKIP mode selection under the above condition are 99.75%, 95.67%, 99.26% and 99.08%, respectively. Especially for Champagne tower, it almost approaches 100%, since Champagne tower has a great deal of still areas.

Most computational time can be saved if FDS strategy is adopted. For example, if there are 65% MBs eventually encoded as SKIP mode, the time saving ratio is 65%×(1-0.52% -9.18%)=58.70%.

Table 5. Statistical results of SKIP mode in B frames.

Sequences	SKIP mode MBs	Correct selection	Wrong selection
Alt Moabit	730059	724671 (99.26%)	5388 (0.74%)
Champagne tower	1138084	1127594 (99.08%)	10490 (0.92%)
Exit	221315	211734 (95.67%)	9581 (4.33%)
Race1	75263	75075 (99.75%)	188 (0.25%)

Table 6. Statistical results of the best reference frames among inter modes.

	Ballroom	Race1	Exit	Alt Moabit	Champagne tower
P(A)	72.29%	93.57%	79.72%	89.45%	89.56%
P(B|A)	95.83%	98.41%	96.05%	97.70%	97.97%
P(C|A)	92.38%	93.11%	92.16%	90.35%	94.64%

4.3 CRFS strategy

The strategies mentioned in subsections 4.1 and 4.2 reduce coding time based on mode correlation. In addition, faster determination of best reference frame for Inter modes can also be used to improve the encoding speed of MVC.

In MVC, multiple reference frame technology is adopted. It means that multiple reference frames is searched to obtain the best matching block of each Inter mode. However, the best reference frames of various Inter modes may have the same reference frame. Let $Best\_Ref_{8\times8}$, $Best\_Ref_{8\times4}$, $Best\_Ref_{4\times8}$ and $Best\_Ref_{4\times4}$ be the best reference frames for 8×8, 8×4, 4×8 and 4×4 blocks, respectively, **A** be the set of MBs satisfying the condition of $Best\_Ref_{8\times8}= Best\_Ref_{4\times4}$, **B** be the set of MBs satisfying the condition of $Best\_Ref_{8\times4}= Best\_Ref_{8\times8}$ and $Ref_{8\times4}= Best\_Ref_{4\times4}$, and **C** be the set of MBs satisfying the condition of $Best\_Ref_{4\times8}= Best\_Ref_{8\times8}$ and $Ref_{4\times8}= Best\_Ref_{4\times4}$, we define the flowing parameters to denote the reference frame correlation of various Inter mode.

$$P(\mathbf{A}) = |\mathbf{A}| / \phi\,, \tag{13}$$

$$P(\mathbf{B} \mid \mathbf{A}) = |\mathbf{B}| / |\mathbf{A}|\,, \tag{14}$$

$$P(\mathbf{C} \mid \mathbf{A}) = |\mathbf{C}| / |\mathbf{A}|\,. \tag{15}$$

In (13), ϕ is the number of total MBs. **Table 6** tabulates the statistical results of five test sequences in terms of $P(\mathbf{A})$, $P(\mathbf{B}|\mathbf{A})$ and $P(\mathbf{C}|\mathbf{A})$. Obviously, $P(\mathbf{B}|\mathbf{A})$ and $P(\mathbf{C}|\mathbf{A})$ exceed 90%. That is, if the reference frames of 8×8 and 4×4 block mode are the same, the reference frames of 8×4 and 4×8 block modes are most likely to be the same as reference frame of 8×8 block mode. Similarly, it is also true for Inter16×8, Inter8×16 modes when reference frames of Inter16×16 and Inter8×8 are the same.

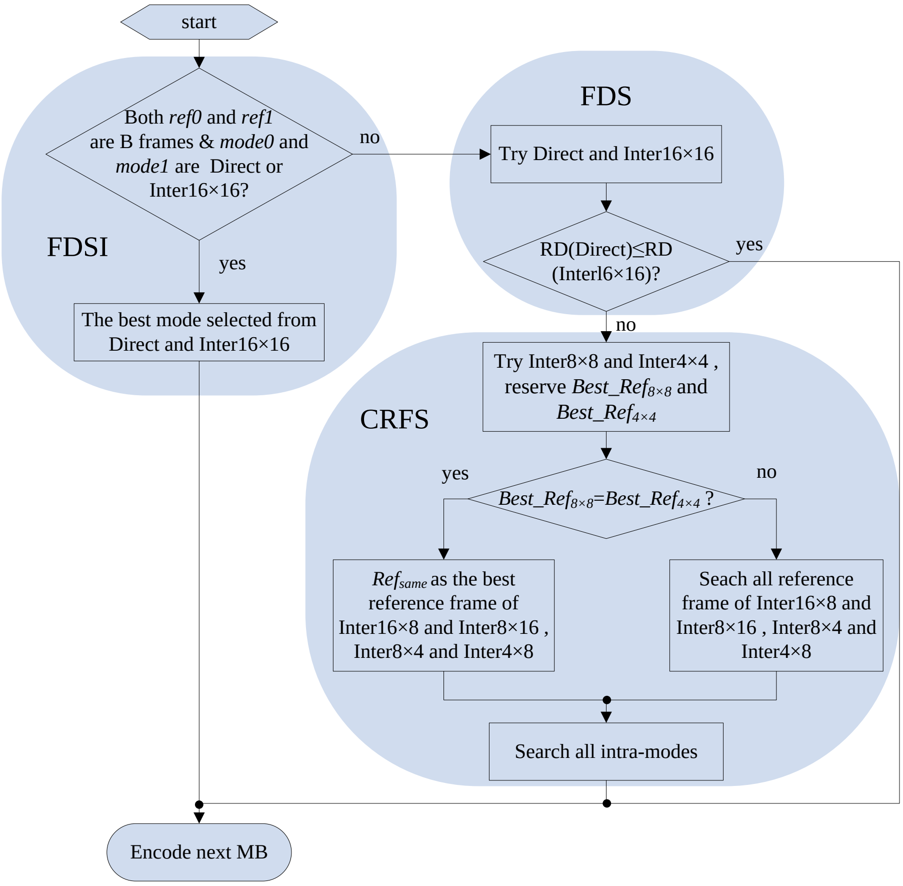


Fig. 7. Flow chart of the proposed fast MB mode selection algorithm.

Thus, a strategy of CRFS is proposed and described as following.

Search in all reference frames for Inter8×8 and Inter4×4 modes, obtain $Best\_Ref_{8\times8}$ and $Best\_Ref_{4\times4}$, if $Best\_Ref_{8\times8}$ and $Best\_Ref_{4\times4}$ are the same frame, only this reference is searched to obtain the motion/disparity vector for Inter8×4 and Inter4×8, Inter16×8, Inter8×16 modes. Otherwise, all reference frames is searched.

Hence, the coding time is saved through reducing the reference frames of Inter16×8, Inter8×16, Inter8×4 and Inter4×8 modes.

4.4 The Proposed fast MB mode selection algorithm

By combining the three strategies, FDSI, FDS and CRFS, we propose a fast MB mode selection algorithm for B frames in MVC, as shown in **Fig. 7**. The algorithm is as follow:

Step 1) If *ref*0 and *ref*1 are B frames, and modes of TMBs in *ref*0 and *ref*1 are SKIP or Interl6×16, only SKIP and Interl6×16 modes are searched for optimal mode when current MB is encoded, then go to step 6). Otherwise, directly go to step 2).

Step 2) Search SKIP mode and Inter16×16 mode and compute RD cost of SKIP and Inter16x16 mode, denoted as *Rd*(SKIP) and *Rd*(Interl6×16). If *RD*(SKIP) is smaller than *RD*(Interl6×16), SKIP mode is selected as the optimal mode, then go to step 6). Otherwise, go to step 3).

(a) Eight views of Race1 sequence

(b) Eight views of Ballroom sequence

(c) Eight views of Exit sequence

(d) Eight views of Alt Moabit sequence

(e) Eight views of Leaving Laptop sequence

(f) Eight views of Breakdancers sequence

(g) Eight views of Champagne tower sequence

(h) Eight views of Pantomime sequence

Fig. 8. Multiview video test sequences.

Step 3) Search all reference frames for Inter8×8 and Inter4×4 modes, and obtain the best reference frames of the two modes. If the best reference frames of Inter8×8 Inter4×4 modes are the same frame (denoted by Ref_{same}). ME/DE is only performed in Ref_{same} for Inter16×8 and Inter8×16, Inter8×4 and Inter4×8 modes, then go to step 5). Otherwise, go to step 4).

Step 4) All reference frames is searched for motion/disparity vector for Inter16×8 and Inter8×16, Inter8×4 and Inter 4×8 modes, then go to step 5).

Step 5) Search all Intra modes, and obtain the best mode based on the minimum RD cost, then go to step 6).

Step 6) Encode next MB.

5. Experimental Results and Discussions

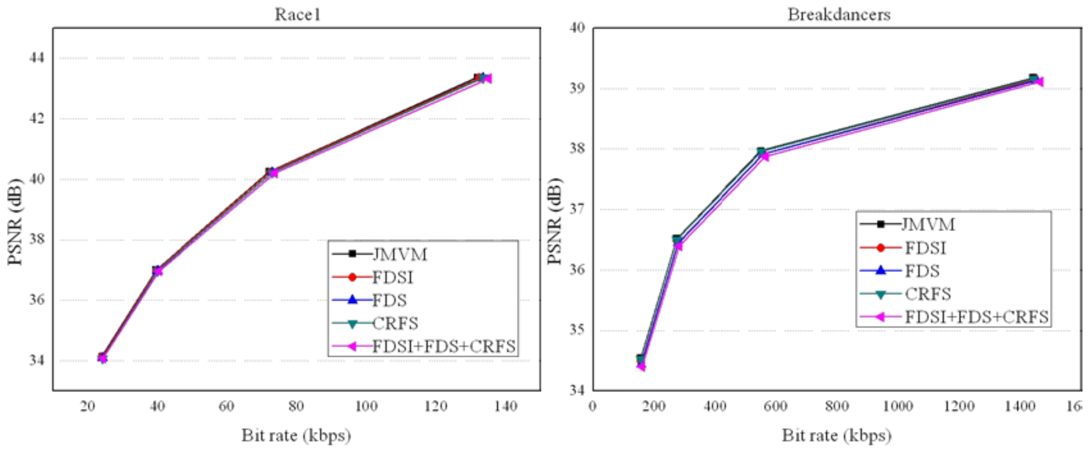
The proposed algorithm is implemented in JMVM 7.0 software and all experiments are performed on a PC with Inter(R) Core(TM)2 Duo 3.0GHz CPU, 3.25GB DDRII memory. The information on test multiview sequences is listed in **Table 7**, and the coding conditions are also given in **Table 8**. The test sequences include Race1, Ballroom, Exit, Alt Moabit, Leaving Laptop, Breakdancers, Champagne tower and Pantomime [27][28][29]. These sequences have different characteristics of motion, content, disparity, resolution, property of camera array and baseline distance, frame rate and GOP-length. **Fig. 8** show the first frames of eight views with respect to the test sequences.

Table 7. Information on test multiview sequences.

Test sequence	**size**	**GOP length**	**Frame rate (fps)**	**Camera array and baseline distance (cm)**	**Features**
Race1	320×240	15	30	1D/ parallel, 20	Fast motion
Ballroom	640×480	12	25	1D/ parallel, 19.5	Great disparity, rotated motion
Exit	640×480	12	25	1D/ parallel, 19.5	Great disparity
Alt Moabit	1024×768	15	15	1D/ parallel, 6.5	Outdoor scene
Leaving Laptop	1024×768	15	15	1D/ parallel, 6.5	Close shot
Breakdancers	1024×768	15	15	1D/ arc, 20	Fast motion
Champagne tower	1280×960	15	15	1D/ parallel, 5	Slow motion
Pantomime	1280×960	15	15	1D/ parallel, 5	Slow motion

Table 8. Coding conditions.

Encoder	JMVM 7.0	MaxRefIdxActiveBL0	2					
Prediction structure	HBP	MaxRefIdxActiveBL1	2					
Number of views	8	Basis QP	22, 27, 32, 37					
Search range	±64	DeltaLayerXQuant	0	1	2	3	4	5
Encoded frames	61	Delta QP Values	0	3	4	5	6	7



(a) Race1 **(b)** Breakdancers

Fig. 9. Rate distortion of the proposed algorithms compared with JMVM 7.0.

The performances of the proposed algorithm are evaluated by comparing the coding time, PSNR and bit rates of the proposed algorithm and JMVM7.0. Since the proposed algorithm incorporates the three strategies FDSI, FDS and CRFS, we also independently perform experiments for FDSI, FDS and CRFS algorithm to analyze their performances and contributions to the proposed joint fast algorithm. **Fig. 9** shows the rate distortion curves of Race1 and Breakdancers with respect to JMVM 7.0, FDSI, FDS, CRFS and the proposed joint algorithm. The four points of each curve are with respect to four different basic QPs, that is, 22, 27, 32, 37 as shown in **Table 8**. It is seen that the five curves in **Fig. 9-(a)** and **Fig. 9-(b)** are very close or even overlaped with each other, which means that the rate distortion performances of these five algorithms are comparative. The rate distortion curves of the other six test multiview sequences with respect to the five different coding algorithms show similar results as in **Fig. 9**.

Table 9 shows experimental comparison results of all the eight test multiview sequences in which $\Delta PSNR$, ΔBR and *Speedup* indicate RD performance and speedup performances and they are defined by

$$\Delta PSNR = PSNR_{pro} - PSNR_{JMVM}, \tag{16}$$

$$\Delta BR = \frac{BR_{pro} - BR_{JMVM}}{BR_{JMVM}} \times 100\%, \tag{17}$$

$$Speedup = \frac{T_{JMVM}}{T_{pro}}. \tag{18}$$

where $PSNR_{pro}$, BR_{pro} and T_{pro} are the Peak Signal-to-Noise Ratio (PNSR), bit rate and encoding time of FDSI, FDS, CRFS or the proposed algorithm, $PSNR_{JMVM}$, BR_{JMVM} and T_{JMVM} are the PNSR, bit rate and encoding time of JMVM 7.0. The data of each test multiview sequence are the average results with respect to the four basic QPs, and the last row of **Table 9** gives the results of the eight sequences on averge. Negative value of $\Delta PSNR$ means degradation in PSNR and positive value of ΔBR indicates bit rate increase. *Speedup* reflects the time saved by the proposed algorithm compared with JMVM 7.0 benchmark.

Table 9 show RD performance comparison among different schemes. We can see that FDSI, FDS, CRFS and the proposed joint algorithm have good adaptability for all test sequences. The four algorithms speed up the coding speed, ranging from 1.61 to 5.42 on average. Simultaneously, they can control increased bit rate from 0.27% to 2.03% and decreased PSNR from 0.01dB to 0.08dB on average. CRFS gains the least saving time, but it almost keeps the same RD performance as JMVM. FDSI takes the second place in keeping original PSNR and bit rate. It also can be seen that the proposed algorithm can reduce coding time significantly. Since there is some computational redundancy among FDSI, FDS and CRFS, *Speedup* of the proposed joint algorithm is less than multiplication of *speedup* ratio of FDSI, FDS and CRFS.

The speedup ratio and RD performances of the proposed algorithm are mainly determined by the distributional proportion of SKIP mode according to the former analyses of FDSI and FDS. Compared with other test sequences, Race1 and Champagne tower sequences contain more areas of static background and more MBs are encoded with SKIP mode. Hence, the best speedup and RD performances are achieved. The *Speedup* is larger than 7. As for Ballroom and Breakdancers test sequences are with fast motion or abundant texture, the percentage of MBs encoded as SKIP mode is less than other sequences. Hence, only 3.71-3.83 times of speedup ratio is achieved with 0.11-0.16dB PSNR decrease and 1.39%-2.38% bit rate increase.

Table 10 gives performance comparison results between the proposed joint algorithm and another fast inter mode decision algorithm named Zhu's algorithm here [25]. The data of Zhu's algorithm come from literature [25], in which the used *QP*s are also 22, 27, 32 and 37. From **Table 10**, it is seen that the speedup ratio of the proposed joint algorithm are higher than that of Zhu's algorithm. This means that the proposed joint algorithm encodes faster than Zhu's algorithm, especially for sequences like Race1, for which more MBs are encoded with SKIP mode.

Table 9. Performance comparisons among JMVM, FDSI, FDS, CRFS and the proposed joint algorithm.

Sequences	**FDSI**			**FDS**			**CRFS**			**Joint algorithm (FDSI+FDS+CRFS)**		
	Δ*PSNR* **(dB)**	Δ*BR* **(%)**	***Speedup***	Δ*PSNR* **(dB)**	Δ*BR* **(%)**	***Speedup***	Δ*PSNR* **(dB)**	Δ*BR* **(%)**	***Speedup***	Δ*PSNR* **(dB)**	Δ*BR* **(%)**	***Speedup***
Race1	-0.01	+0.44	2.59	-0.03	+0.99	6.18	0.01	+0.52	1.76	0.05	+1.91	7.22
Ballroom	-0.03	+0.74	1.66	-0.04	+1.52	2.29	-0.02	+0.37	1.46	-0.08	+2.70	3.71
Exit	-0.02	+0.21	1.84	-0.04	+1.19	3.02	-0.02	+0.15	1.58	-0.06	+1.81	4.60
Alt Moabit	-0.02	+0.35	2.30	-0.04	+2.70	3.99	-0.01	+0.20	1.61	-0.05	+3.47	5.96
Leaving Laptop	-0.01	+0.50	2.49	-0.02	+1.69	4.35	-0.01	+0.33	1.65	-0.04	+2.38	6.06
Breakdancers	-0.05	+1.16	1.90	-0.07	+1.16	2.25	-0.02	+0.04	1.46	-0.11	+2.38	3.83

Champagne tower	-0.03	+0.16	2.82	-0.04	+0.16	5.29	-0.01	+0.23	1.75	-0.06	+0.28	7.01
Pantomime	-0.09	+0.24	2.34	-0.09	+0.67	2.83	-0.03	+0.31	1.59	-0.16	+1.29	4.99
Average	-0.03	+0.48	2.24	-0.05	+1.26	3.78	-0.01	+0.27	1.61	-0.08	+2.03	5.42

Table 10. Performance comparisons among the proposed joint algorithm and Zhu's algorithm [25].

	Performance	**Zhu's algorithm**			**Joint algorithm (FDSI+FDS+CRFS)**		
		Race1	**Ballroom**	**Exit**	**Race1**	**Ballroom**	**Exit**
Average	$\Delta PSNR$(dB)	-0.0028	-0.0052	-0.0012	0.05	-0.08	-0.06
	ΔBR(%)	0.12	0.45	0.10	1.91	2.70	1.81
	Speedup	2.23	2.47	3.39	7.22	3.71	4.60

6. Conclusion

Hierarchical B picture based on MVC prediction structure is adopted in the MVC standard for high compression efficiency, but the prediction structure results in intensive computational complexity. Meanwhile, the variable macroblock mode and motion/disparity estimation also significantly increase the computational complexity. To reduce MVC computational complexity and enable real-time multiview video applications, we propose a fast MB mode selection algorithm for B frames in MVC in this paper. It uses the uneven features of MB mode distribution and the correlations among the reference frames of Inter modes. The proposed algorithm includes two aspects. First, we use the correlations of modes in corresponding location of reference frames to reduce the number of candidate modes. Secondly, we propose the strategy that quickly selects SKIP mode. Moreover, we exploit a fast strategy to select the best reference frame under the correlation among reference frames of Inter modes. Experimental results show that the proposed algorithm can speed up 3.71~7.22 times coding speed over JMVM7.0 while maintaining the similar coding quality. Moreover, the proposed scheme can be combined with other fast motion and disparity estimation algorithms to further reduce the computational complexity.

References

[1] Pascal Frossard, Juan Carlos de Martin and Reha Civanlar, "Media Streaming with Network Diversity," in *Proc. of the IEEE*, vol. 96, no. 1, pp. 39-53, January 2008. Article (CrossRef Link).

[2] Hsien-Po Shiang and Mihaela van der Schaar, "Distributed Resource Management in Multi-hop Cognitive Radio Networks for Delay Sensitive Transmission," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 941-953, February 2009. Article (CrossRef Link).

[3] L. Zhou, X. Wang, W. Tu, G. Mutean and B. Geller, "Distributed Scheduling Scheme for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 409-419, April 2010. Article (CrossRef Link).

[4] L. Zhou, B. Geller, B. Zheng, A. Wei and J. Cui, "System Scheduling for Multi-Description Video Streaming Over Wireless Multi-Hop Networks," *IEEE Transactions on Broadcasting*, vol. 55, no. 4, pp. 731-741, December 2009. Article (CrossRef Link).

[5] Y.M. Feng, D.X. Li, K. Luo and M. Zhang, “Asymmetric bidirectional view synthesis for free viewpoint and three-dimensional video,” *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2349-2355, November 2009. Article (CrossRef Link).

[6] K.-J. Oh et al., “Multi-view video and multi-channel audio broadcasting system,” in *Proc. of 3DTV-CON*, no. 4379437, May 2007. Article (CrossRef Link).

[7] P. Merkle, K Müller and T. Wiegand, “3D Video: Acquisition, Coding, and Display,” *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 946-950, July 2010. Article (CrossRef Link).

[8] P. Merkle et al., “Coding efficiency and complexity analysis of MVC prediction structures,” in *Proc. of 15th European Signal Conference*, pp. 5-9, September 3-7, 2007. Article (CrossRef Link).

[9] P. Merkle et al., “Efficient prediction structures for multi-view video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461-1473, November 2007. Article (CrossRef Link).

[10] Tae-Young Chunga, II-Lyong Junga, Kwanwoong Songa and Chang-Su Kim, “Multi-view video coding with view interpolation prediction for 2D camera arrays,” *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 474-486, July-August 2010. Article (CrossRef Link).

[11] Y. Zhang, M. Yu and G. Jiang, “New approach to multi-modal multi-view video coding,” *Chinese Journal of Electronics*, vol. 18, no. 2, pp. 338-342, April 2009.

[12] P.K. Park and Y. S. Ho, “Prediction structure and quantization parameter selection for efficient multiview video coding,” *Optical Engineering*, vol. 47, no. 4, 047401, April 2008. Article (CrossRef Link).

[13] Hung-Chih Lin, Wen-Hsiao Peng and Hsueh-Ming Hang, “Fast context-adaptive mode decision algorithm for scalable video coding with combined coarse-grain quality scalability (CGS) and temporal scalability,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 5, pp. 732-748, May 2010. Article (CrossRef Link).

[14] Chia-Hung Yeh et al., “Fast mode decision algorithm for scalable video coding using Bayesian theorem detection and Markov process,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 563-574, April, 2010. Article (CrossRef Link).

[15] L. Shen, Z. Liu, Z. Zhang and G. Wang, “An adaptive and fast multi-frame selection algorithm for H.264 video coding,” *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 836-839, November 2007. Article (CrossRef Link).

[16] H. Nisar and T. S. Choi, “Multiple initial point prediction based search pattern selection for fast motion estimation,” *Pattern Recognition*, vol. 42, no. 3, pp. 475-486, March 2009. Article (CrossRef Link).

[17] Grecos and M. Yang, “A framework for fast mode decision in the H.264 video coding standard,” *Digital Signal Processing: A Review Journal*, vol. 17, no. 3, pp. 652-664, May 2007. Article (CrossRef Link).

[18] H. Wang, S. Kwong and C. W. Kok, “An efficient mode decision algorithm for H.264/AVC encoding optimization,” *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 882-888, May 2007. Article (CrossRef Link).

[19] Vetro, P. Pandit, H. Kimata and A. Smolic, “Joint multiview video model (JMVM) 7.0,” *Joint Video Team (JVT) of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, JVT-Z207, January 2008. Article (CrossRef Link).

[20] Liquan Shen et.al., “Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding,” *IEEE Transactions on Broadcasting*, vol. 55, no. 4, pp. 761-766, December 2009. Article (CrossRef Link).

[21] X. Li, D. Zhao, S. Ma and W. Gao, “Fast disparity and motion estimation based on correlations for multiview video coding,” *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 2037-2044, November 2008. Article (CrossRef Link).

[22] Z. Peng, G. Jiang, M. Yu and Q. Dai, “Fast macroblock mode selection algorithm for multiview video coding,” *EURASIP Journal on Image and Video Processing*, Article ID 393727, 2008. Article (CrossRef Link).

[23] G. Cernigliaro et al., "Fast mode decision for multiview video coding based on depth maps," in *Proc. of SPIE*, vol. 7257, no. 72570N, January 2009. Article (CrossRef Link).

[24] L. F. Ding et al., "Content-aware prediction algorithm with inter-view mode decision for multiview video coding," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1553-1564, December 2008. Article (CrossRef Link).

[25] Wei Zhu, Wei Jiang and Yaowu Chen, "A Fast Inter Mode Decision for Multiview Video Coding," in *Proc. of International Conference on Information Engineering and Computer Science*, pp. 1-4, December 2009. Article (CrossRef Link).

[26] M. Yu, Z. Peng and G. Jiang. "Statistical analysis of macroblock mode selection in JMVM," *JVT of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, JVT-Y026, October 2007. Article (CrossRef Link).

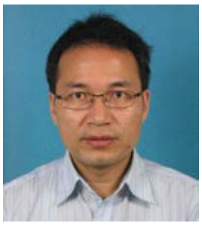
[27] Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff and T. Wiegand, "HHI test material for 3D video," *ISO/IEC JTC1/SC29/WG11*, *M15413*, April, 2008.

[28] L. Zitnick, S. B. Kang and M. Uyttendaele, "High-quality video view interpolation using a layered representation," *ACM SIGGRAPH and ACM Transactions on Graphics*, Los Angeles, CA, vol. 4, pp. 600-608, August 2004. Article (CrossRef Link).

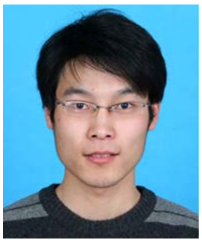
[29] M. Tanimoto, T. Fujii and N. Fukushima, "1D parallel test sequences for MPEG-FTV," *ISO/IEC JTC1/SC29/WG11*, *M15378*, April, 2008.

Mei Yu received her B.S. and M.S. degrees from Hangzhou Institute of Electronics Engineering, China, in 1990 and 1993, and Ph.D. degree from Ajou University, Korea, in 2000. She is now with the Faculty of Information Science and Engineering, Ningbo University, China. Her research interests include image/video coding and video perception.

Ping He received her M.S. degree from NingBo University in 2009. Her research interests include multi-view video coding and image processing.



Zongju Peng received his B.S. degree in computer science from Sichuan University, China, in 1998, and received his Ph.D. degree from Institute of Computing Technology, Chinese Academy of Science in 2010. He is now with the Faculty of Information Science and Engineering, Ningbo University, China. His research interests concentrate on image/video compression.



Yun Zhang received his B.S. and M.S. degrees from NingBo University in 2004 and 2007 respectively, and received his Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences in 2010. His research interests mainly include digital video compression and communications, SoC design and embedded system for consumer electronics.

Yuehou Si received her B.S. degree from Ningbo University, China, in 2008. She is currently a M.S. candidate of Ningbo University. Her research interests mainly include digital video compression and communications, multi-view video coding.

Gangyi Jiang received his M.S. degree from Hangzhou University，China, in 1992, and received his Ph.D. degree from Ajou University, Korea, in 2000. He is now a professor at Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include digital video compression and communications, multi-view video coding and image processing.