# Fast Encoder Design for Multi-view Video

**Fan Zhao[1], Kaiyang Liao[1], Erhu Zhang[1] and Fangying Qu[2]**
[1] Department of Information Science, Xi'an University of Technology, Xi'an 710048, China
[e-mail: vcu@xaut.edu.cn]
[2] Department of Mathematics, Northwest University, Xi'an 710127, China
[e-mail: fangyingqu@gmail.com]
*Corresponding author: Fan Zhao

*Received February 11, 2014; revised April 11, 2014; accepted May 10, 2014; published July 29, 2014*

## Abstract

Multi-view video coding is an international encoding standard that attains good performance by fully utilizing temporal and inter-view correlations. However, it suffers from high computational complexity. This paper presents a fast encoder design to reduce the level of complexity. First, when the temporal correlation of a group of pictures is sufficiently strong, macroblock-based inter-view prediction is not employed for the non-anchor pictures of B-views. Second, when the disparity between two adjacent views is above some threshold, frame-based inter-view prediction is disabled. Third, inter-view prediction is not performed on boundary macroblocks in the auxiliary views, because the references for these blocks may not exist in neighboring views. Fourth, finer partitions of inter-view prediction are cancelled for macroblocks in static image areas. Finally, when estimating the disparity of a macroblock, the search range is adjusted according to the mode size distribution of the neighboring view. Compared with reference software, these techniques produce an average time reduction of 83.65%, while the bit-rate increase and peak signal-to-noise ratio loss are less than 0.54% and 0.05dB, respectively.

***Keywords:*** multi-view video, compression coding, multi-view video coding, temporal correlation, inter-view correlation, prediction

This work is supported in part by the National Natural Science Foundation of China (NSFC) Project No. 11272253 and China Postdoctoral Science Foundation No. 20100481336.

**http://dx.doi.org/10.3837/tiis.2014.07.015**

## 1. Introduction

As a result of their greatly enhanced viewing experience and high interactivity, 3D video and free viewpoint television are attracting attention in various industries and research institutes [1-3]. Multiple synchronized cameras are usually used to capture the same scene from different viewpoints to form a 3D system, and the resulting multi-view video brings not only a whole new 3D impression, but also a large amount of data for storage and transmission. Encoding all the views efficiently is a crucial issue for future multi-view applications. Multi-view video coding (MVC) is a key technique for distributing multi-view video content through networks with limited bandwidth. MVC was developed to improve coding efficiency. In 2005, MPEG and ITU-T started the procedure of MVC standardization [4-5]. Although the standard activity for MVC (multi-view extension of H.264/AVC) is largely complete, there is a lot of room for improvement in reducing its complexity.

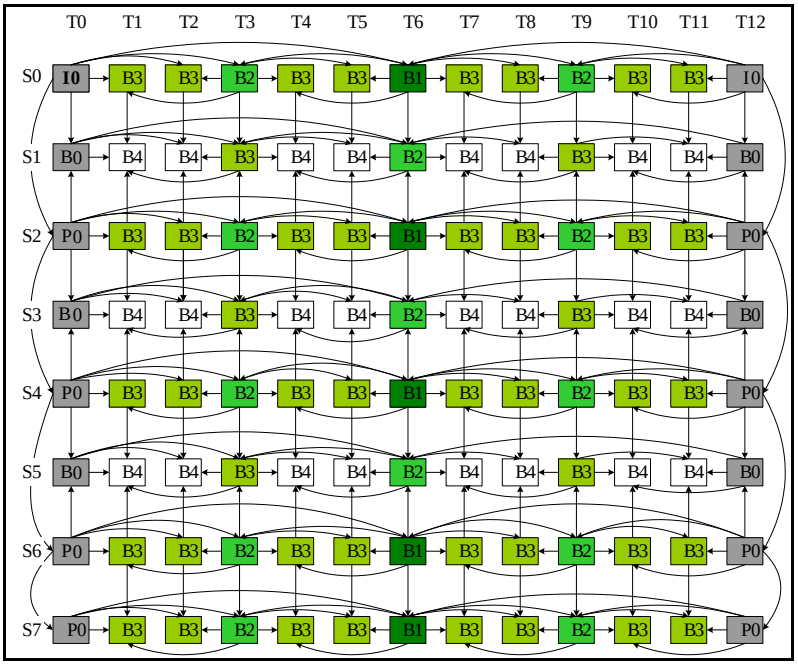


**Fig. 1.** JMVM reference prediction structure.

By fully utilizing the redundancy amongst temporally successive frames and adjacent views, many practical schemes for multi-view video compression have been proposed. Their common feature is the flexible use of prediction schemes. These typical prediction structures include the simulcast scheme, the CIP (Cross Island Parkway) scheme, and the HHI (Heinrich Hertz Institute) scheme [6-10]. The simulcast scheme encodes each single view independently using the hierarchical B-pictures, and is supported by H.264/AVC. Its performance is generally lower, for it does not employ inter-view correlations. The CIP scheme employs inter-view prediction only for the first picture of every GOP (Group of Pictures), whereas other pictures are encoded using the same prediction technique as in the simulcast scheme. Thus, its coding performance is only improved to a very limited extent. The HHI scheme, proposed by Fraunhofer HHI, exhibits the best compression performance by fully utilizing the

prediction both in the time axis and in the view direction [11]. Its prediction structure is shown in **Fig. 1** [6, 11, 15], where $S_n$ denotes the $n^{\text{th}}$ individual view sequence and $T_n$ denotes the $n^{\text{th}}$ time-instant frame. All the pictures in **Fig. 1** make up an MVC prediction unit, in which hierarchical B-pictures are used for each view, as in H.264, while in the view direction, only the I-, P-, and B-pictures are employed. In H.264, the prediction unit is applied to each GOP, whereas in MVC, the prediction unit operates on a GoGOP (group-of-GOPs). In an MVC prediction unit, views are named after the starting picture. For example, views starting with picture $I_0$ are called I-views, like $S_0$; views starting with picture $B_0$ are called B-views, such as $S_3$ and $S_5$; and views starting with picture $P_0$ are called P-views, such as $S_2$, $S_4$, $S_6$, and $S_7$. Although the HHI prediction structure exhibits the best coding performance, its complexity is too high. Thus, algorithms for practical applications must speed up the prediction process, especially in the view direction. This is because there are more reference frames in the MVC scheme, and predicting B-views is more complex than for other views.

MVC improves the coding efficiency by utilizing motion- and disparity-compensated prediction (M/DCP). However, the complexity of inter-frame prediction is very high, especially when rate-distortion optimization is used. Temporal prediction is applied first, and inter-view prediction will be skipped if the sum of absolute differences of the temporal prediction is small [12-13]. A fast strategy for deciding between MCP and DCP was designed in [14], where the GOP size of each view is set to four. This short GOP size hinders the full utilization of correlation in the temporal direction. At the same time, the motion vectors (MVs) must be stored, and the locations of static macroblocks (MBs) of the base view and anchor pictures are needed, so the search for co-located MBs takes a considerable amount of time. Reference [15] investigated the influence of removing inter-view prediction from the higher temporal decomposition levels, but failed to implement their algorithm properly. Fortunately, based on an analysis of the contribution of inter-view prediction to the coding gain at different temporal layers, a simplified prediction structure was proposed in [16], in which inter-view prediction was disabled if the temporal dependency dominates the current view. However, this is based on the assumption that the contribution of inter-view prediction is less in higher temporal layers. However, inter-view estimation works well at higher temporal decomposition levels, and the experimental results in [17] indicate that inter-view prediction is often more efficient than temporal prediction for a significant number of blocks under fast camera motion in the temporal direction.

Most previous complexity reduction techniques for fast MVC focus on the selection of mode size, prediction direction, reference frames, and search range. Experiments have shown that most MBs in regions with homogeneous motion or relatively static backgrounds would select the large mode size (16×16) for motion estimation (ME), while only MBs in the region with complex motion need disparity estimation (DE) and small mode size ME. To reduce the computational complexity of MVC, [18] utilized the spatial property of the motion field to limit the candidate prediction modes to a small subset. Zhang et al. noted that there is a high probability that MBs with smaller partition sizes will eventually select the same reference frame and prediction direction as those with larger partition sizes [19]. Thus, they proposed an algorithm for B-pictures in which MBs with smaller mode sizes follow the decisions of MBs with higher mode sizes to select the best prediction direction and reference frame. The approach in [20] increases the overall speed of MVC by reducing the size of the search range for both ME and DE in regions with homogeneous motion. Inter-view SKIP mode correlation has been exploited by Shen et al. [21]. Reference [22] reports a low-complexity DE technique that chooses the previous disparity vector (PDV) instead of the median prediction vector as the

search center for DE. By combining the MVC methods in [19-21] with PDV-DE, a novel complexity reduction technique can be derived. The correlation of rate distortion (RD) costs among inter-views in SKIP mode are studied and used to reduce the computational complexity of MVC in [23]. Khattak et al. have used the correlation between the RD costs of the SKIP mode in neighboring views and Bayesian decision theory to reduce the number of candidate coding modes for a given MB [24]. A hybrid optimal stopping model to solve the mode decision problem has been developed, whereby the predicted mode probability and estimated coding time are jointly investigated with inter-view correlations [25]. Yeh et al. proposed a fast mode decision algorithm to avoid the high computational complexity of MVC [26]. The minimum and maximum values of the RD cost in the previously encoded view are used to compute a threshold for each mode in the current view. Using these thresholds, the mode decision process becomes more efficient. A fast mode decision process is adopted to reduce the complexity of HEVC (the High Efficiency Video Coding) [30].

The frame at instant $T_6$ would select its reference from pictures at $T_0$ and $T_{12}$, with a GOP length span. The distribution of the encoding modes for MBs at $T_6$ reflects the temporal correlation in a given GOP, and our first contribution uses this fact to develop a method of MB-based inter-view prediction skipping. Inter-view correlations are high in a GOP (or a half-GOP). Pictures at instants $T_0$ and $T_{12}$ are always processed before those between them (i.e., $T_1$ to $T_{11}$) for any of the views according to the encoding order. Thus, our second contribution is to use the inter-view dependencies at instant $T_0$ or $T_{12}$ to judge whether frame-based inter-view prediction can be skipped at other instants. For the sake of the arrangement of multi-view cameras, boundary MBs may not find their reference in the neighboring views. Our third contribution is inter-view prediction skipping for these boundary MBs. To realize a low coding gain, finer partitions are skipped in the inter-view prediction for MBs in static image areas, such as those in the background. Because of the correlation in the mode size distribution of MBs in adjacent views, we propose to dynamically adjust the search range for the current MB.

The paper is organized as follows. The framework of our flexible prediction selection method is proposed in Section 2. Experimental results are presented in Section 3, and some conclusions are drawn in Section 4.

## 2. Flexible Prediction Selections in MVC

For convenience, we refer to a picture by both its view and temporal indices. $S_i/T_n$ $(0 \le i \le I, 0 \le n \le N)$ represents the picture of the $i^{\text{th}}$ view at instant $T_n$, where $I$ and $N$ represent the number of views and GOP length in an MVC unit, respectively. For simplicity, a prediction structure with three views is shown in **Fig. 2**. Pictures $S_i/T_0$ and $S_i/T_{12}$ are called the anchor pictures, and the others are non-anchor pictures. The flexible prediction selections are detailed as follows.

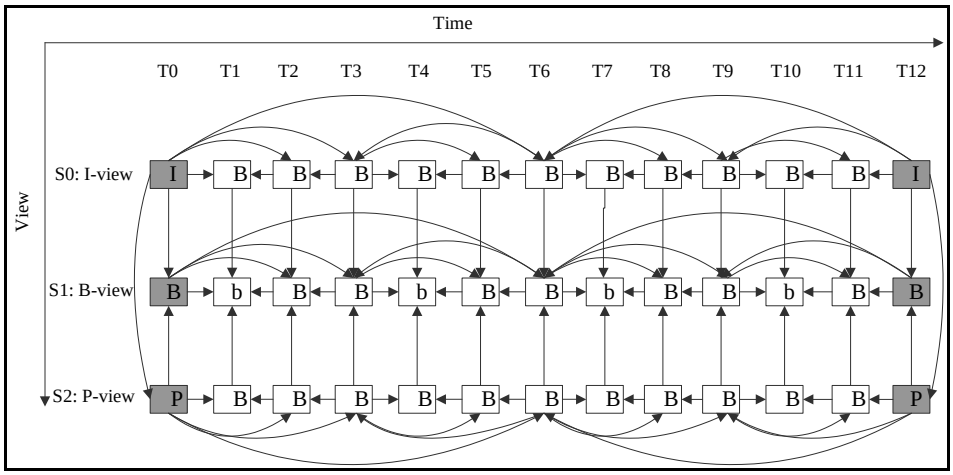


**Fig. 2.** JMVM reference prediction structure with three views.

## 2.1 MB-based inter-view prediction skipping for non-anchor pictures of B-views

*Input:* $Ratio\_Intra\_T_6$ *and* $Ratio\_For\_Bac$
*Output:* $flag\_skip(S_1/T_n)$, *where* $0<n<6, 6<n<12$
*if* $Ratio\_Intra\_T_6 \le thresh\_Intra$
  *if* $threh\_Lower \le Ratio\_For\_Bac \le threh\_Upper$
    $flag\_skip(S_1/T_n)=1,$ *for* $0<n<6, 6<n<12$
  *endif*
  *if* $Ratio\_For\_Bac > threh\_Upper$
    $flag\_skip(S_1/T_n)=1,$ *for* $0<n<6$
    $flag\_skip(S_1/T_n)=0,$ *for* $6<n<12$
  *endif*
  *if* $Ratio\_For\_Bac < threh\_Lower$
    $flag\_skip(S_1/T_n)=0,$ *for* $0<n<6$
    $flag\_skip(S_1/T_n)=1,$ *for* $6<n<12$
  *endif*
*else*
  $flag\_skip(S_1/T_n)=0,$ *for* $0<n<6, 6<n<12$
*endif*

**Fig. 3.** Pseudo-code for determining flag_skip.

As for the B-view pictures, the reference selection method for anchor pictures is the same as in JMVM. The MB-based prediction of non-anchor pictures of B-views proposed here differs from the method employed in JMVM. For MBs of B-views at all instants except $T_0$, $T_6$, and $T_{12}$, inter-view prediction is not employed if the temporal correlation of the encoding GOP unit is sufficiently strong. The temporal correlation is obtained by analyzing the prediction modes of the MBs at instant $T_6$, and this is derived before coding the current picture. First, we use the flag " *flag _ skip* " to indicate whether inter-view prediction should be used. A value of

0 means inter-view prediction is employed, and 1 denotes that its use will be reconsidered. Two parameters are introduced to determine the value of $flag\_skip$ after coding picture $S_1/T_6$. One is the ratio of the number of intra-coded MBs of the picture at instant $T_6$, which is labeled $Ratio\_Intra\_T_6$. The other is the ratio of the number of MBs of picture $S_1/T_6$ whose references are at $S_1/T_0$ and those whose references are at $S_1/T_{12}$, which is labeled $Ratio\_For\_Bac$. There exists a certain degree of temporal correlation at the current GOP if $Ratio\_Intra\_T_6$ is lower than the predefined threshold $thresh\_Intra$. In such cases, stronger temporal correlation exists in the half of this GOP that depends on $Ratio\_For\_Bac$. If $Ratio\_For\_Bac$ is more than a predefined $thresh\_Upper$, the period from $T_0$ to $T_6$ exhibits more temporal correlation than the period from $T_6$ to $T_{12}$. If $Ratio\_For\_Bac$ is lower than a predefined $thresh\_Lower$, the period from $T_6$ to $T_{12}$ is more temporally correlated than the period from $T_0$ to $T_6$. If $Ratio\_For\_Bac$ is between $thresh\_Lower$ and $thresh\_Upper$, the temporal correlation of the period from $T_0$ to $T_6$ is assumed to be very similar to that of $T_6$ to $T_{12}$. The $flag\_skip$ value is determined by combining these two parameters, as shown in **Fig. 3**. The coding performance of temporal prediction is better than that of inter-view prediction for static regions [12]. Hence, the MV of the corresponding MB is used to further determine whether inter-view prediction is employed when $flag\_skip$ = 1. In other words, when predicting the non-anchor pictures of B-views, the two temporal reference pictures are definitively selected, but the two inter-view reference pictures, say $S_0/T_n$ or $S_2/T_n$, are chosen as follows. First, a prediction is made using the forward reference picture in the temporal direction. If the MV is (0, 0), inter-view prediction using $S_0/T_n$ as a reference is skipped; otherwise, inter-view prediction is performed. Second, a prediction is made using the backward reference picture in the temporal direction. If the MV is (0, 0), inter-view prediction using $S_2/T_n$ as a reference is skipped; otherwise, inter-view prediction is performed.

**Table 1.** Statistical analysis of parameter distributions

| **Sequence** | **Ratio_Intra_T6** | | | **Ratio_For_Bac** | | |
|---|---|---|---|---|---|---|
| | **＜0.1 (%)** | **[0.1,0.2] (%)** | **＞0.2 (%)** | **＞0.8 (%)** | **[0.8,1.25] (%)** | **＜1.25(%)** |
| Ballroom | 40 | 60 | 0 | 16 | 67 | 17 |
| Exit | 100 | 0 | 0 | 0 | 85 | 15 |
| Rena | 35 | 65 | 0 | 30 | 70 | 0 |
| Race1 | 60 | 20 | 20 | 55 | 25 | 20 |
| **Average** | **58.75** | **36.25** | **5** | **25.25** | **61.75** | **13** |

We conducted statistical tests on four multi-view video (MVV) sequences, “Ballroom”, “Exit”, “Rena”, and “Race1”, using a JMVM encoder to examine $Ratio\_Intra\_T_6$ and $Ratio\_For\_Bac$. The statistical results for the two parameter distributions at instant $T_6$ are shown in **Table 1**. On average, the percentage of MBs is 58.75 when $Ratio\_Intra\_T_6$ was less than 0.1 The percentage of MBs is 36.25 when $Ratio\_Intra\_T_6$ is in [0.1, 0.2], The

percentage of MBs is 5 when $Ratio\_Intra\_T_6$ is greater than 0.2 . Thus, $Ratio\_Intra\_T_6$ at instant $T_6$ is less than 0.2 for most MBs about 95% percent. Hence, we set $thresh\_Intra$ to 0.2. As the percentage of MBs is greater than 50 when $Ratio\_For\_Bac$ is in [0.8, 1.25], we set $thresh\_Lower$ and $thresh\_Upper$ to 0.8 and 1.25, respectively.

The proposed scheme is realized in the MVC reference software JMVM 8.0 [27]. The features of six test datasets are listed in **Table 2**. The experimental results are presented in **Table 3**, where "DPSNR", "Dbitrate (%)", and "Dtime (%)" represent the peak signal-to-noise ratio (PSNR) change, percentage bitrate change, and the percentage change in total coding time, respectively, between our method and the reference software. From **Table 3**, we can see that the average time reduction is 25.41%. The average PSNR loss is 0.03 dB, and the increase in bitrate is about 0.29% on average. The proposed algorithm exhibits consistent speedup for all the sequences, with a minimum gain of 21.38% for the "Ballroom" sequence and a maximum gain of 29.04% for "Rena". These data show that the algorithm and parameter thresholds are feasible and effective for all video sequences.

**Table 2.** Features of the test datasets

| Test Data | Image Features | | | | Camera Properties | | |
|---|---|---|---|---|---|---|---|
| | Image Size | Frame Rate | GOP Number | GOP Size | Camera Number | Camera Densities | Camera Arrangement |
| AkkoKayo | 640×480 | 30fps | 19 | 15 | 15 | 5cm | 2D parallel |
| Ballroom | 640×480 | 25fps | 20 | 12 | 8 | 20cm | 1D parallel |
| Exit | 640×480 | 25fps | 20 | 12 | 8 | 20cm | 1D parallel |
| Flamenco2 | 640×480 | 30fps | 66 | 15 | 5 | 20cm | Cross array |
| Rena | 640×480 | 30fps | 19 | 15 | 16 | 5cm | 1D parallel |
| Race1 | 640×480 | 30fps | 35 | 15 | 8 | 20cm | 1D parallel |

**Table 3.** Experimental results: difference between our method and the reference software JMVM

| Sequences | DPSNR (dB) | DBitrate (%) | Dtime (%) |
|---|---|---|---|
| AkkoKayo | -0.03 | -0.41 | -26.06 |
| Ballroom | -0.03 | -0.52 | -21.38 |
| Exit | -0.02 | -0.15 | -25.78 |
| Flamenco2 | -0.04 | -0.09 | -25.32 |
| Race1 | -0.01 | -0.18 | -24.87 |
| Rena | -0.04 | -0.36 | -29.04 |
| **Average** | **-0.03** | **-0.29** | **-25.41** |

## 2.2 Frame-based inter-view prediction skipping when large differences exist between adjacent views

As can be seen from **Fig. 1**, the pictures at instant $T_0$ and $T_{12}$ are always processed before $T_1 - T_{11}$ for any of the views according to the encoding order. Hence, we can use the statistical

dependencies of inter-view pictures at instants $T_0$ and $T_{12}$ to estimate those at other instants. To simplify the description, we take the basic framework with $S_0, S_1$ and $S_2$ as an example, as shown in **Fig. 2**. Let $R_{Num\_S_0\_S_1}(T_0)$ denote the number of MBs with intra-mode at $S_1/T_0$ by forward DE, (that is, estimating $S_1$ with $S_0$ ), and let $R_{Num\_S_2\_S_1}(T_0)$ be the number of MBs with intra-mode by backward DE (that is, estimating $S_1$ with $S_2$ ). $R_{Num\_S_0\_S_1}(T_{12})$ and $R_{Num\_S_2\_S_1}(T_{12})$ denote similar quantities at $T_{12}$. When the length of a GOP is sufficiently large, the inter-view disparity cannot be manifested. Thus, we separate the GOP into two halves, and consider the inter-view similarities independently. Because the inter-view similarities do not change significantly if the time interval of each half is short, the following hypothesis is put forward:

$$R_{Num\_S_0\_S_1}(T_0) = R_{Num\_S_0\_S1}(T_1) = ... = R_{Num\_S_0\_S_1}(T_6) \quad (1)$$

$$R_{Num\_S_0\_S_1}(T_{12}) = R_{Num\_S_0\_S_1}(T_{11}) = ... = R_{Num\_S_0\_S_1}(T_7) \quad (2)$$

$$R_{Num\_S_2\_S_1}(T_0) = R_{Num\_S_2\_S_1}(T_1) = ... = R_{Num\_S_2\_S_1}(T_6) \quad (3)$$

$$R_{Num\_S_2\_S_1}(T_{12}) = R_{Num\_S_2\_S_1}(T_{11}) = ... = R_{Num\_S_2\_S_1}(T_7) \quad (4)$$

*If* $R_{Num\_S_0\_S_1}(T_0) \geq Cor\_Thr$
Forward frame based inter-view prediction is skipped for $S_1$ at instants from $T_1$ to $T_6$
*If* $R_{Num\_S_0\_S_1}(T_{12}) \geq Cor\_Thr$
Forward frame based inter-view prediction is skipped for $S_1$ at instants from $T_7$ to $T_{11}$
*If* $R_{Num\_S_2\_S_1}(T_0) \geq Cor\_Thr$
Backward frame based inter-view prediction is skipped for $S_1$ at instants from $T_1$ to $T_6$
*If* $R_{Num\_S_2\_S_1}(T_{12}) \geq Cor\_Thr$
Backward frame based inter-view prediction is skipped for $S_1$ at instants from $T_7$ to $T_{11}$

**Fig. 4.** Frame-based inter-view prediction skipping scheme.

On the assumption of high inter-view similarities within a half-GOP unit, we propose the frame-based inter-view prediction scheme in **Fig. 4**. That is, if the number of intra-coded MBs is beyond $Cor\_Thr$ at $T_0$ and $T_{12}$, a large difference is considered to exist between the two adjacent views, and frame-based inter-view prediction is disabled accordingly.

### 2.3 Inter-view prediction skipping for boundary MBs

MVVs are captured by various multiple camera arrangements, such as 1D and 2D arrays, 1D arcs, and so on. The arrangement of the multiple cameras means that some border MBs cannot find reference images in the adjacent views, regardless of whether the arrangement is regular. This is illustrated in **Fig. 5**, which shows the current view picture and its inter-view reference picture as well as the predicted picture. The white blocks at the border of the predicted picture show that matching MBs cannot be found in the inter-view reference picture. Omitting the inter-view prediction for these MBs would undoubtedly reduce the computational load.

(a)

(b)

(c)

**Fig. 5.** (a) Current view, (b) inter-view reference picture, and (c) predicted picture.

Similar to temporal prediction, DE is also conducted by minimizing the Lagrange cost function, which is the sum of the distortion $D$ and the rate $R$, weighted by the Lagrange factor $\lambda$. For each MB $B_i$ of the current view, the DE algorithm chooses a disparity vector $d_i=(d_{i\_x}, d_{i\_y})$ within a search range in the adjacent view so as to minimize $J$:

$$d_i = \arg\min J = \arg\min\{D(B_i, d_i) + \lambda \times R(B_i, d_i)\} \quad (5)$$

Here, the distortion $D$ is calculated as the sum of the squared errors between the current block

$B_i$ and the previously reconstructed reference block $B_i^{'}$ in the adjacent view:

$$D(B_i, d_i) = \sum_{(x,y)\in B_i} [B_i(x,y) - B_i^{'}(x - d_{i\_x}, y - d_{i\_y})]^2 \tag{6}$$

The intra-mode is normally chosen when its Lagrange cost is minimized, even though the current MB can find a reference from adjacent views. Therefore, the MBs with intra prediction modes can be classified into two categories: those whose elements could not determine their references because of large distance displacements in the cameras, and those whose elements have a cost function relatively smaller than any of the inter-modes. A texture descriptor $Text\_D$ can be introduced to derive MBs of the first type:

$$Text\_D = (\sum_{(x,y)\in B_i} [B_i(x,y)]^2 - [\sum_{(x,y)\in B_i} B_i(x,y)]^2)/256 \tag{7}$$

The inter-view prediction of the border MB is skipped if the prediction is in intra mode and the corresponding prediction distortion is large than the texture descriptor. The decision condition is added to the distortion function as follows:

$$\begin{aligned} &if\ \ Mode = "Intra"\ and\ \ D(B_i, d_i) \geq Text\_D \\ &\quad \text{Inter - view prediction is skipped;} \\ &else \\ &\quad \text{Inter - view prediction is adopted.} \end{aligned} \tag{8}$$

In comparison with the reference software JMVM, the experimental results given by the programs proposed in Sections 2.2 and 2.3 are listed in **Table 4**. Here, $Cor\_Thr$ is set to 0.25. The selection of the border MBs depends on the position relation of the cameras, which can be derived from the camera parameters. The border MBs are those within three MBs of the edge. The experimental results show an average runtime saving of 34.39% with only 0.04dB reduction in PSNR and 0.48% increase in bit rate. An even greater time reduction is obtained for sequences containing fast motion, such as Race1.

**Table 4.** Experimental results given by the proposed schemes compared with the reference software JMVM

| Sequences | DPSNR (dB) | DBitrate (%) | Dtime (%) |
|---|---|---|---|
| AkkoKayo | -0.02 | 0.49 | -28.23 |
| Ballroom | -0.02 | 0.41 | -29.75 |
| Exit | -0.03 | 0.56 | -38.61 |
| Flamenco2 | -0.02 | 0.32 | -34.08 |
| Race1 | -0.08 | 0.90 | -40.45 |
| Rena | -0.04 | 0.20 | -35.21 |
| **Average** | **-0.04** | **0.48** | **-34.39** |

### 2.4 Finer partitions skipping in inter-view prediction for MBs in static image areas

The supporting block sizes of motion compensation in H.264 vary from $16\times16$ to $4\times4$, and the luminance samples have many options between the two. In general, a larger partition size (such as $16\times16$, $16\times8$, or $8\times16$) is appropriate for homogeneous areas of the frame, and a smaller partition size (such as $8\times8$, $8\times4$, $4\times8$, or $4\times4$) is likely to be beneficial to areas with rich textures. The MB partitioning method also adapts to the disparity in motion compensation. Note that static image areas, such as background, cause a fundamental disadvantage to inter-view prediction. Hence, only the large partition size is expected in static image areas for inter-view prediction, and the unnecessary DE using finer partitions will be canceled. MBs in static image areas are picked out by the following method. First, we change the prediction order by ranking the intra-mode options at the top of the queue, followed by the inter-mode options. Second, the optimal intra-mode is recorded as $M_{intra\_}$ when all the intra-mode options are finished. Third, the current inter-view mode is labeled $M_{inter\_}$ after inter-view predictions with large partitions have been performed. Finally, whether the finer partitions are used for inter-view prediction depends on:

$$\begin{aligned} &if\ M_{\mathrm{int}ra\_}=16\times16\ and\ M_{\mathrm{int}er\_}=16\times16 \\ &\text{Finer partitions are skipped for inter - view prediction.} \end{aligned} \tag{9}$$

Thus, for scenes with large areas consisting of static background, as many finer partitions as possible are canceled.

### 2.5 Disparity search range adjustment based on mode distribution correlation

In the JMVM reference software, prediction can be done for either a 16×16 MB or its sub-block partitions (16×8, 8×16, 8×8, 8×4, 4×8, and 4×4). Large modes are suitable for predicting regions of homogeneous motion, whereas small modes are appropriate for areas of complex motion. Because the mode size can reflect motion activity to a certain extent, the search range for the current MB can be dynamically adjusted according to the mode size distributions of MBs in the previously coded neighbor view.

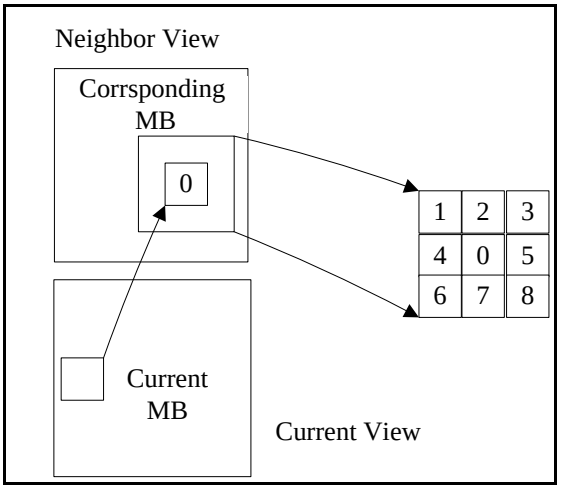


**Fig. 6.** Corresponding MB and neighboring MBs in previously coded view.

The current MB and the corresponding MB in the neighboring view are shown in **Fig. 6**. The modes of the corresponding MB and its eight neighbor MBs are used to estimate the motion activity of the current MB, thus determining its disparity search range. When all nine MBs have been predicted using either 16×16 or skip mode, the current MB is considered with the simple mode. When at least one of the nine MBs is predicted using a small mode (16×8 or 8×16), the current MB is considered with the medium mode. When at least one of the nine MBs is predicted using a smaller mode (8×8, 8×4, 8×8, or 4×4), it is considered with the complex mode.

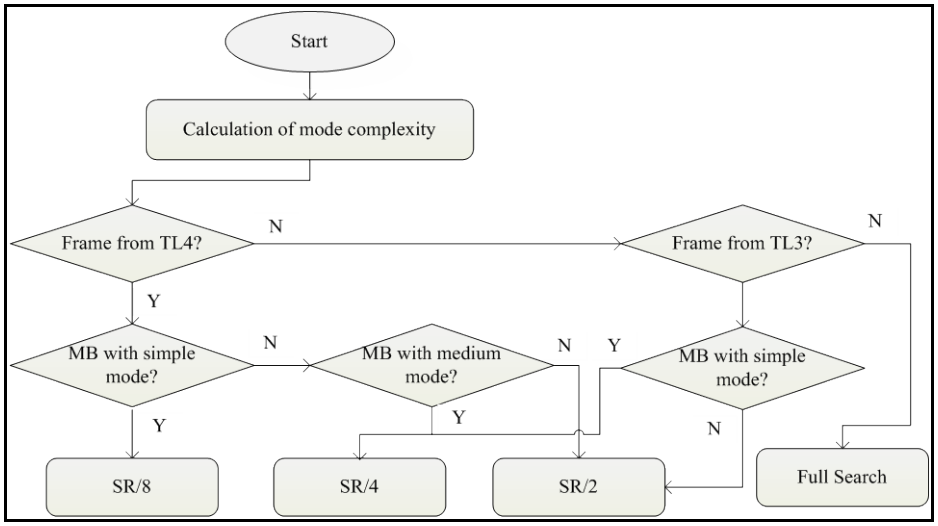


**Fig. 7.** Proposed disparity vectors search range adjustment algorithm. SR denotes the maximum search range, which is 64. As defined in [22], TL3 and TL4 are Temporal Level 3 and Temporal Level 4, respectively.

After the mode type of the current MB has been determined, the proposed disparity search strategy shown in **Fig. 7** is applied. Two conditions are checked: (i) Does the frame belong to TL4 or TL3? (ii) Does the MB belong to a simple, medium, or complex mode? Different search ranges are selected according to the answers.

## 3. Experimental Results and Analysis

In this section, we evaluate the results of the overall algorithm, which incorporates the proposed five techniques. The comparison results of the overall algorithm and JMVM [27] are given in Table 5, where BDPSNR and BDBR [28] are used to represent the average PSNR and bit rate differences with QP under 10, 20, 24, 28, 32, 36 and 40. The six test sequences [29] are evaluated in the experiment, which is run on a machine with Intel Core i7 CPU 860@2.8 GHz and 4 GB RAM.

As can be seen from **Table 5**, the overall algorithm reduces the encoding time by 83.65% on average and the coding efficiency loss is negligible, with 0.05 dB PSNR loss or 0.54% bit rate increase. The results show that the proposed approach achieves a consistent gain in coding speed, with the lowest gain of 70.32% for Race1 and the highest gain of 88.76% for AkkoKayo. The maximum PSNR loss is 0.13 dB and the maximum bitrate increment is 1.05%,

which are negligible compared with the saving in encoding time. The proposed algorithm exhibits considerable and consistent speedup for all sequences with large or small disparities, slow or fast motion. For example, for sequences containing still backgrounds and large disparities, such as Rena and Exit, the proposed algorithm can reduce the encoding time by more than 86%. For sequences containing complicated motion and less disparity, such as Race1 and Ballroom, the time saving is lower, but still significant, at about 70.32–82.87%.

**Table 5.** Experimental results compared with the reference software JMVM .

| Sequences | BDBR (%) | BDPSNR (dB) | Dtime (%) |
|---|---|---|---|
| AkkoKayo | 0.62 | -0.04 | -88.76 |
| Ballroom | 1.05 | -0.03 | -82.87 |
| Exit | 0.52 | -0.03 | -86.55 |
| Flamenco2 | 0.30 | -0.04 | -85.97 |
| Race1 | 0.42 | -0.13 | -70.32 |
| Rena | 0.31 | -0.04 | -87.43 |
| **Average** | **0.54** | **-0.05** | **-83.65** |

**Table 6.** Experimental results of proposed overall algorithm and the methods in [19] and [20].

| Sequence | Method | DTime (%) | | | | | | BDRR (%) | BD PSNR (dB) |
|---|---|---|---|---|---|---|---|---|---|
| | | Basis QP | | | | | Average | | |
| | | 20 | 24 | 28 | 32 | 36 | | | |
| Ballroom | [19] | -62.10 | -66.70 | -69.38 | -71.30 | -72.63 | -68.42 | 0.68 | -0.10 |
| | [20] | -74.90 | -76.40 | -78.10 | -78.90 | -81.20 | -77.90 | 0.44 | -0.09 |
| | **Ours** | **-79.17** | **-80.55** | **-83.27** | **-86.05** | **-87.83** | **-84.28** | **1.09** | **-0.02** |
| Exit | [19] | -61.98 | -69.66 | -72.57 | -74.28 | -74.84 | -70.67 | 0.23 | -0.08 |
| | [20] | -77.10 | -81.80 | -85.70 | -87.20 | -89.00 | -84.16 | 0.12 | -0.04 |
| | **Ours** | **-83.72** | **-86.27** | **-88.38** | **-90.60** | **91.32** | **88.06** | **0.30** | **-0.03** |

To the best of our knowledge, the state-of-the-art results are those reported in [22], which combines state-of-the-art methods ([19], [20], and [21]) with its PDV-DE technique. Compared with the methods in [19], [20], and [21], PDV-DE increases the time saving by only 1.5%. To evaluate the performance of each algorithm alone, we further compare the MVC methods in [19] and [20] with our overall MVC encoding technique proposed in Section 2. Three views from the standard test sequences Ballroom and Exit are used. For each sequence, results are obtained for QP values of 20, 24, 28, 32, and 36.

In addition to JMVM, we also compare our algorithm to MVC methods in [19] and [20] in **Table 6**. Compared to MVC methods in [19] and [20], the overall algorithm can obtain better results on average in the evaluated performances. The proposed algorithm reduces 86.17% coding time for these two sequences while [19] and [20] only reduce 69.6% and 81.03% coding time respectively. Meanwhile, the proposed algorithm can achieve a better RD performance with 0.7% bit rate increase. Note that our method is particularly useful for sequences with large disparities. For example, the time saving for "Exit" was over 4% greater than that for "Ballroom", which has less disparity

## 4. Conclusion

A fast encoder design for MVC has been proposed. The primary contributions are our frame- and MB-based inter-view prediction skipping techniques, which are based on the statistical properties of previous intra-coded MBs. Based on the mode size distribution of MBs in the previously coded neighbor view, we also developed a dynamical adjustment method for the search range of the current MB. Experimental results showed that the proposed algorithm can significantly reduce the computational complexity of MVC and maintain almost the same coding efficiency, demonstrating its applicability to various types of videos.

## References

[1] ISO/IEC 14 496-10, "Information Technology-Coding of Audio-Visual Objects-Part 10: Advanced Video Coding," *Final Draft International Standard*, December, 2003. Article (CrossRef Link)

[2] ISO/IEC JTC/SC29/WG11, "Report on 3DTV Exploration," *N5878*, July, 2003. Article (CrossRef Link)

[3] Masayuki Tanimoto, "Overview of Free Viewpoint Television," *Signal Processing: Image Communication*, vol.21, no.6, pp.454-461, July, 2006. Article (CrossRef Link)

[4] Y.-S. Ho and K.-J. Oh, "Overview of multi-view video coding," in *Proc. of IEEE Int. Workshop on Signal, System, and Image Processing*, pp. 5-12, June, 2007. Article (CrossRef Link)

[5] Joint Video Team of ISO/IEC & MPEG & ITU-T VCEG, "Joint Draft 4.0 on Multi-view Video Coding," *Doc. JVT-X209*, June, 2007. Article (CrossRef Link)

[6] Huaiyi Pan and Feng Pan, "Development of Multi-view Video Coding Using Hierarchical B Pictures," *in Congress on Image and Signal processing*, vol.1, pp.497-503, May, 2008. Article (CrossRef Link)

[7] Philipp Merkle, Aljoscha Smolic, Karsten Müller, and Thomas Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Transactions on circuits and systems for video technology*, vol.17, no.11, pp. 1461-1473, November, 2007. Article (CrossRef Link)
ISO/IEC JTC1/SC29/WG11, "Multi-view Coding using AVC," *JVT Doc. m12945*, Jane, 2006. Article (CrossRef Link)

[8] Athanasions Leontaris and Pamela C. Cosman, "Compression Efficiency and Delay Tradeoffs for Hierarchical B-Pictures and Pulsed-Quality Frames," *IEEE Transactions on image processing*, vol.16, no.7, pp. 1726-1740, July, 2007. Article (CrossRef Link)

[9] Heiko Schwarz, Detlev Marpe and Thomas Wiegand, "Hierarchical B Pictures," *ITU-T and ISO/IEC Joint Video Team, Doc. P014*, July, 2005. Article (CrossRef Link)

[10] ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MVC," *MPEG2006/W7798*, Jane, 2006. Article (CrossRef Link)

[11] LiFu Ding, ShaoYi Chien, YuWen Huang, YuLin Chang and LiangGee Chen, "Stereo video coding system with hybrid coding based on joint prediction scheme," *IEEE International Symposium on Circuits and Systems*, vol. 6 , pp. 6082-6085, May, 2005. Article (CrossRef Link)

[12] L.-F. Ding, S.-Y. Chien, and L.-G. Chen, “Joint prediction algorithm and architecture for stereo video hybrid coding systems,” *IEEE Trans. Circuits and Systems for Video Technology*, vol.16, no.11, pp.1324-1337, November, 2006. Article (CrossRef Link)
[13] Jheng-Ping Lin, Tang A.C.-W. “A fast direction predictor of inter frame prediction for multi-view video coding,” *IEEE International Symposium on Circuits and Systems*, pp.2589-2592, 2009. Article (CrossRef Link)
[14] ISO/IEC JTC1/SC29/WG11, “Core Experiments on view-temporal prediction structures,” *MPEG2006/M13196*, April, 2006. Article (CrossRef Link)
[15] Junyan Huo and Yilin Chang, “Study on improving the coding efficiency of multiview video coding,” *Xi'an University of Electronic Science and Technology, Ph. D. Thesis*, April, 2008, Article (CrossRef Link)
[16] Zhao Fan, Liu Guizhong, Ren Feifei and Na Zhang, “Flexible predictions selection for multi-view video coding,” in *Proc. of 2009 Data Compression (DCC)*, pp: 471-471, March, 2009. Article (CrossRef Link)
[17] Liquan Shen, Zhi Liu, Suxing Liu, Zhaoyang Zhang, and Ping An, "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)
[18] Y. Zhang, S. Kwong, G. Jiang, and H. Wang, "Efficient multi-reference frame selection algorithm for hierarchical B pictures in multi-view video coding," *IEEE Transaction on Broadcasting*, vol. 57, no.1, pp.15-23, 2011. Article (CrossRef Link)
[19] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "View-adaptive motion estimation and disparity estimation for low-complexity multiview video coding," *IEEE Transactions on Circuits Systems for video technology*, vol. 20, no.6, pp.925-930, 2010. Article (CrossRef Link)
[20] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "Early SKIP mode decision for MVC using inter-view correlation," *Signal Processing: Image Communication*, vol. 25, pp. 88-93, February, 2010. Article (CrossRef Link)
[21] S. Khattak, R. Hamzaoui, S. Ahmad, and P. Frossard, "Low-complexity multi-view video coding," in *Proceedings of Picture Coding Symposium (PCS'2012)*, pp.97-100, May, 2012. Article (CrossRef Link)
[22] Liquan Shen, Zhi Liu, Ping An, Ran Ma, and Zhaoyang Zhang. "Low-Complexity Mode Decision for MVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no.6, pp.837-843, June, 2011. Article (CrossRef Link)
[23] Shadan Khattak, Raouf Hamzaoui, Thomas Maugey, Shakeel Ahmad, and Pascal Frossard, “Bayesian Early Mode Decision Technique for View Synthesis Prediction-Enhanced Multiview Video Coding,” *IEEE Signal Processing Letters* , vol. 20, no.11, pp.1126-1129, November, 2013. Article (CrossRef Link)
[24] Tiesong Zhao; Kwong, S.; Hanli Wang; Zhou Wang; Zhaoqing Pan; Kuo, C.-C.J., "Multiview Coding Mode Decision With Hybrid Optimal Stopping Model,", *IEEE Transactions on Image Processing*,vol. 22, no.4, pp.1598-1609, April, 2013. Article (CrossRef Link)
[25] Chia-Hung Yeh, Ming-Feng Li, Mei-Juan Chen, Ming-Chieh Chi, Xin-Xian Huang, and Hao-Wen Chi, “Fast Mode Decision Algorithm Through Inter-View Rate-Distortion Prediction for Multiview Video Coding System,” *IEEE Transactions on Industrial Informatics*, vol.10, no.1, pp.594,603, February, 2014. Article (CrossRef Link)
[26] ISO/IEC Standard ITU-T VCEG, “Joint Draft 8.0 on Multiview Video Coding,” *JVT-AB204*, 2008. Article (CrossRef Link)
[27] G. Bjontegaard, “Calculation of average PSNR difference between RD curves,” *ITU-T Q.6/SG16 VCEG 13th Meeting, Doc. VCEG-M33*, April, 2001. Article (CrossRef Link)
[28] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, “Common test conditions for multiview video coding,” *Doc. JVT-T207*, July, 2006. Article (CrossRef Link)
[30] Han Huang, Yao Zhao, Chunyu Lin and Huihui Bai, “Fast Intraframe Coding for High Efficiency Video Coding,” *KSII Transactions on Internet and Information Systems*, vol.8, no.3, pp.1093-1104, March, 2014. Article (CrossRef Link)

**Fan Zhao** received a Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, Xi'an, China, in 2009. She is an associate professor in the Department of Information Science at Xi'an University of Technology, Xi'an, China, and is now working as a postdoctoral fellow in the Department of Computer Science and Engineering, Xi'an Jiaotong University. Her research interests include image processing, video compression, and pattern recognition. (E-mail: vcu@xaut.edu.cn).

**Kaiyang Liao** received a Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, Xi'an, China, in 2013. He is currently a lecturer with the School of Printing and Packaging Engineering, Xi'an University of Technology, Xi'an, China. His research interests include data mining, pattern recognition, video analysis and retrieval.
(E-mail: liaokaiyang@xaut.edu.cn).

**Erhu Zhang** received a Ph.D. in Biomedical Engineering from Xi'an Jiaotong University in 2008. He is currently a professor in the Department of Information Science, Xi'an University of Technology, Xi'an, China. His research interests include digital image processing, pattern recognition, and intelligence information processing.
(E-mail: eh-zhang@xaut.edu.cn)

**Fangying Qu** is currently studying for a Bachelor's degree in Applied Mathematics at the Northwestern University, Xi'an, China.
(E-mail: fangyingqu@gmail.com)