# Comparative Interactivity Analysis in Multiview Video Coding Schemes

You Yang, Qionghai Dai, Gangyi Jiang, and Yo-Sung Ho

**In a multiview video system, interactivity is important for users and should be considered in the design of multiview video coding (MVC). In this paper, we present an interactivity evaluation model for MVC schemes by using both weighted random graph and Markov approaches. The main factors that affect both the interactivity and rate-distortion (RD) performances of MVC schemes are analyzed and discussed in detail. By taking these factors into consideration, a new MVC scheme is proposed for high interactivity and RD gains. Experimental results show that the proposed scheme has a significant interactivity gain with little coding loss, compared to the state-of-the-art benchmark. As an extension to RD performance analysis, the interactivity evaluation model can be used as a design tool of alternative schemes for a future interactive multiview video system.**

**Keywords: Three-dimensional video, free viewpoint video, multiview video coding, interactivity evaluation, human-computer interaction.**

Manuscript received July 7, 2009; revised Mar. 5, 2010; accepted May 13, 2010.

This work was supported by the Major State Basic Research Development Program 973 of China (2009CB320905), the Natural Science Foundation of China (60872094, 60832003), the Project from Chinese Ministry of Education (200816460003), and the China Postdoctoral Natural Science Foundation (grant 20090460323).

You Yang (phone: +86 10 62788613, email: sayu_yangyou@163.com) and Qionghai Dai (email: qhdai@tsinghua.edu.cn) are with the Broadband Network and Multimedia Laboratory, Department of Automation, Tsinghua University, Beijing, China.

Gangyi Jiang (email: jianggangyi@126.com) is with Faculty of Information Science and Engineering, Ningbo University, Ningbo, China, and also works in State Key Laboratory for Novel Software Technology, Nanjing University, China.

Yo-Sung Ho (email: hoyo@gist.ac.kr) is with the School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Rep. of Korea.

doi:10.4218/etrij.10.0109.0391

## I. Introduction

The three-dimensional television (3DTV) system is an open, flexible, and modular immersive TV system. It is backwards-compatible to the conventional 2D digital television and able to support a wide range of different 2D and 3D displays [1], [2]. The interactivity and flexibility of the 3DTV system allows for the immersive viewing of large panoramic images and videos, and supports high resolution with brilliant quality of depth perception. In the 3DTV system, a scene is captured by a multiple baseline camera setup, from which the amount of video data is increased proportionally with the number of cameras.

Multiview video coding (MVC) has emerged as a new field to find solutions to encode multiview video signals by using the state-of-the-art video codec [3]. Various schemes combining temporal and interview predictions have been proposed where frames are not only predicted from the temporally neighboring frames but also from the corresponding frames in adjacent views. Several approaches are based on predictive coding from reference pictures and, therefore, are closely related to classic video coding, but take advantage of multiple views of the same scene, such as the compression techniques in [4]-[6]. In such a way, MVC allows a very flexible design of temporal and interview prediction dependencies. Different designs of this open architecture lead to considerably different coding performances. The Joint Video Team (JVT) has adopted the MVC-hierarchical B-picture (HBP) prediction structure presented in [4] as the non-normative structure for the joint multiview video model (JMVM) [7].

In the literature, comparison between different prediction structures only focused on rate-distortion (RD) performance. In

this work, we argue that using solely RD performance ignores important differences between structures when considering their implementations. Our motivation for this work is to address interactivity by providing a general framework for which an analysis can be developed in a systematic manner. The analysis in MVC schemes and real-time human-computer interaction are both important before developing any future interactive multiview video system (IMVS) [8], [9]. In [3], it was noted that both low latency and high coding performance are highly desirable for MVC schemes. In other words, the functionality of video cassette recording (VCR) enabling quick and user friendly browsing of multimedia contents is highly desirable in video applications. However, the functionality of free temporal browsing and interview skimming of multiview video contents, which is described as interactivity of MVS [1], is much higher and complex than the traditional VCR functionality in single-view video. Experiments have shown that more reasonable prediction dependencies in MVC schemes may result in higher coding performance but will increase the workload for real-time interaction [10].

Research on random accessibility is the first step for interactivity analysis. The user can access any single frame in temporal and view-wise directions when watching a multiview video program, and resource consumption for this accessing is considered as random accessibility of the IMVS [11]. In this case, frame random access is a mutually independent event and can be treated as a non-Markov procedure. Therefore, random accessibility was measured in the classical possibility space in previous works [10], [12]. On the other hand, using physical access time is another approach in measuring random accessibility [3]. Actually, interaction between a client and the IMVS is a continuous action that is formed by a series of frame accessing. Which frame will be accessed next is mainly dependent on the frame being accessed currently. Thus, interaction is essentially a Markov procedure, and it is not an appropriate way to measure the interactivity by single frame random accessibility in a non-Markov manner.

In this paper, we develop an interactivity evaluation model for MVC schemes that allows us to compare different prediction structures in terms of interactivity. This model provides new tools for the analysis and design of multiview prediction structures. We use this model to study the impact on interactivity of the main elements of the multiview prediction structure. From these existing schemes, we are then able to derive alternative prediction structures, with much higher interactivity and comparable RD performance. We present examples where two structures have similar RD performance, but the proposed schemes achieve 37.11% and 59.89% higher interactivity.

The rest of this paper is organized as follows. In section II, our interactivity evaluation model and some analytical results for MVC schemes are presented. In section III, we identify the effect of the interactivity evaluation model to the JMVM prediction structure. We also present our modifications of the JMVM structure to obtain new prediction structures with similar RD performance but significantly higher interactivity. In section IV, we present the experimental results to compare the JMVM prediction structures and our proposed schemes in terms of RD performance and interactivity. Finally, we conclude our work in section V.

## II. Interactivity Evaluation Model

### 1. Interactivity Evaluation Model

Interaction between a user and an IMVS is a Markov action in a period of time, and it is essentially different from the non-Markov random access. In research on random accessibility, it is assumed that the user can access any single frame on temporal and view-wise directions in the procedure of program watching, and the resource consumption for this action is considered as random accessibility of the IMVS [11]. Random accessibility was theoretically measured by the number of decoded frames in classical possibility space in previous works regardless of the frame position [10], [12]. Some works applied physical decoding time to measure the interaction latency [3], but the output evaluation results will be hardware dependent and are in an after-effect manner. Theoretical research is usually applied to predict the accessibility performance. Although there will be a gap with the physical after-effect result, the theoretical prediction is meaningful to IMVS designers. However, these non-Markov approaches are not suitable for interactivity analysis.

Given a specific MVC scheme (Fig. 1(a)), our interactivity evaluation model is based on generating a random graph $G$ with weights (Fig. 1(b)). In this weighted random graph, each node corresponds to a frame in the MVC scheme. A virtual start point and end point is placed before and behind the graph (that is, scheme) for the convenience of stochastic analysis, respectively. Directed edges in the graph show the possible interaction path from current to the other frame in the procedure of watching. The weight labeled on edge is in the simplified form of $(p_{\text{state } i}, f_c)$, where $p_{\text{state } i}$ ($i$=1, 2, 3) is the possibility of interaction state in set $S$={stay, left, right} indicating the user can stay in the current view or switch to the adjacent left or right neighbor view. The parameter $f_c$ is the frame cost that induced from the decoding time

$$D = \left( f_{\text{I}} w_{\text{I}} + f_{\text{P}} w_{\text{P}} + \sum\nolimits_{l} f_{\text{B}l} w_{\text{B}l} \right) C , \quad (1)$$

where $f_{\text{I}}$, $f_{\text{P}}$, and $f_{\text{B}l}$ is the number of I-, P-, and B$l$-type frames

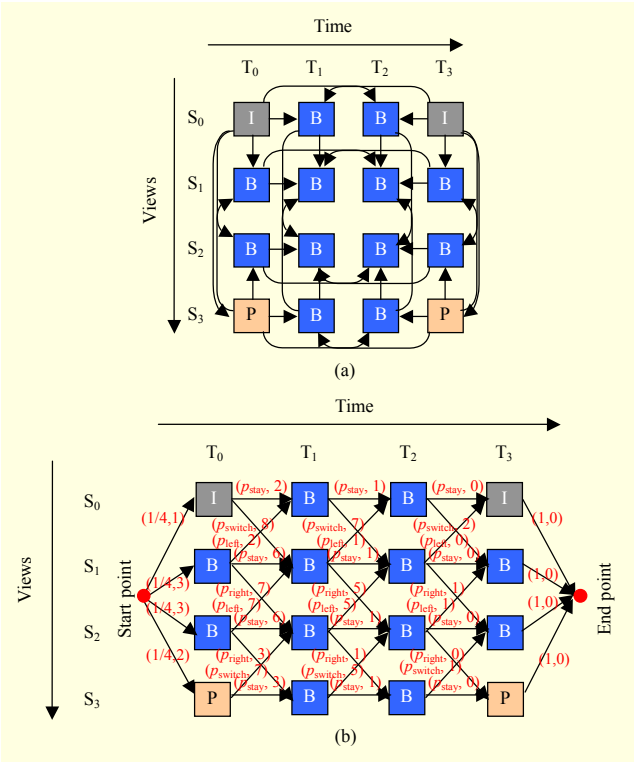



Fig. 1. (a) Multiview video coding sample scheme 1 and (b) weighted random graph model $G$ for sample scheme 1.

involved for switching. $w_I$, $w_P$, and $w_{Bl}$ is the weight for I-, P-, and B$l$-type frame, respectively. $l$ is the number of hierarchy levels for the B-type frame, and $C$ is the average decoding time for one frame when all required references are ready in memory. We arrange substantial experiments in the experiment section for decoding delay measurement, and the results will show that the latency for one frame is limited in a very small range of milliseconds regardless of its frame type when all required references are ready. Therefore, we set $w_I=w_P=w_{Bl}=1$ and let $f_c=f_I+f_P+f_{Bl}$, and simply label the weight as $(p_{\text{state }i}, f_c)$. The gap between physical and predicted decoding time will be very small in statistical manner. For example, the edge from S3T0 to S2T1 is labeled with $(p_{\text{left}}, 7)$ meaning the user may switch from S3T0 to S2T1 with possibility $p_{\text{left}}$, and 7 frames will be decoded for this switching action including frame S2T0, S2T1, S0T1, S3T1, S0T3, S2T3, and S3T3, thus decoding delay is predicted as $D=7C$. There are only three states provided in $G$ because these states will be empirically shown to be enough for user interaction through user behavior studies.

The interactivity of an IMVS can be evaluated by our weighted random graph model. Interaction from user in the period from T0 to T3 will form a random path $P$ from the start point to end point, while the possibility and frame-cost of $P$ are the product of all $p_{\text{state }i}$ and the sum of all $f_c$ in $P$, respectively. Therefore, the expected number of frames to be decoded for $P$ can be obtained in probability space as

$$E(P)=\mathbf{P}(P)W(P), \tag{2}$$

where $\mathbf{P}(P)$ and $W(P)$ are the possibility and frame-cost of random path $P$, respectively. $E(.)$ is the expectation operator for random variable in probability theory. The two parameters in (2) can be calculated as

$$\mathbf{P}(P)=\prod\nolimits_{(p_{\text{state }i})_j \in P} p_{\text{state }i}, \tag{3}$$

$$W(P)=\sum\nolimits_{(f)_j \in P}(f)_j, \tag{4}$$

where $(p_{\text{state }i})_j$ and $(f)_j$ ($j$=0,1,…,$N$) are distinctive weights on $P$ and $N$+1 is the length of the random path. For example, the value of $N$ is 4 in Fig. 1(b).

Given an MVC scheme and its corresponding weighted random graph model, $E(P)$ indicating the expected number of decoded frames for an interaction procedure from start point to end point can be calculated using (2) to (4). A higher value of $E(P)$ will result in longer latency for interaction. However, the value of $E(P)$ is calculated for one random path. The number of random paths will increase dramatically with the size of weighted random graph model. Therefore, it is not suitable to evaluate the interactivity of an IMVS by just one $E(P)$ value. To this end, the interactivity for $G$ that indicates the expected number of decoded frames for any random path in weighted random graph model is calculated by

$$E(G)=\sum E(P). \tag{5}$$

The average interactivity of an IMVS is obtained by (5), indicating the average decoding latency for any interaction in $G$. A smaller $E(G)$ value represents higher interactivity of IMVS, since less latency is needed.

## 2. Possibility for $p_{\text{state }i}$

Users can interact with an IMVS in the procedure of watching, and the possibility $p_{\text{state }i}$ for view switching is critical in our weighted random graph model. Figure 2 provides a series of interactivity performances for different MVC schemes when $p_{\text{state }i}$ ranging from 0 to 1. The names of the mentioned schemes will be described in the latter section. These results show that the gradient of interactivity is different but constant for MVC schemes when the value of $p_{\text{state }i}$ are changing. In this case, it is better to compare interactivity performance of schemes in a specific possibility interval or value.

To this end, an interactive multiview video player was developed and substantial experiments were arranged for user behaviors [13]. These empirical studies showed that users tend to interact with an IMVS in just a small range of views, although they are able to switch to far distance views. This is

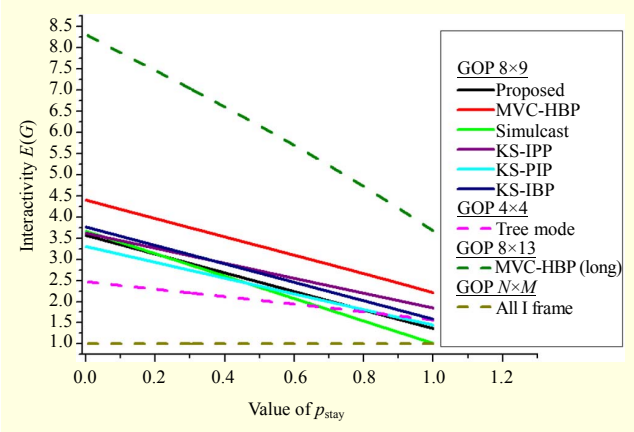



Fig. 2. Values of $E(G)$ for different MVC schemes when different values of $p_{state\,i}$ are selected.

Table 1. Experimental results for view switching possibilities.

| Left | | Right | |
|---|---|---|---|
| Switch step | Possibility | Switch step | Possibility |
| 1 | 0.0018 | 1 | 0.0018 |
| 2 | 0.0000 | 2 | 0.0000 |
| 3 | 0.0000 | 3 | 0.0000 |
| 4 | 0.0000 | 4 | 0.0000 |
| 5 | 0.0000 | 5 | 0.0000 |
| 6 | 0.0000 | 6 | 0.0000 |
| 7 | 0.0000 | 7 | 0.0000 |

because that far distance view switching will cause scene jittering and make the user very uncomfortable. According to statistical results, the value of $p_{stay}$ is 0.9964, and the possibilities for other states are listed in Table 1.

According to the results in Table 1, users tend to have just 3 interaction states in the procedure of watching, and may not switch to other views with more than 1 step. Therefore, the interaction state set $S$ with 3 states is enough for our random graph model. Furthermore, the possibilities for these states are

$$p_{stay}=0.9964,$$
$$p_{right}=0.0018,$$
$$p_{left}=0.0018.$$

On the other hand, $p_{switch}$ is 0.0036, and it is a special case in view 0 and $M$–1 since these views do not have adjacent left and right neighbor views, respectively.

## 3. Example of Weighted Random Graph Model

In this section we select two MVC schemes and their associated models as shown in Figs. 1 and 3. The value of $M$–1

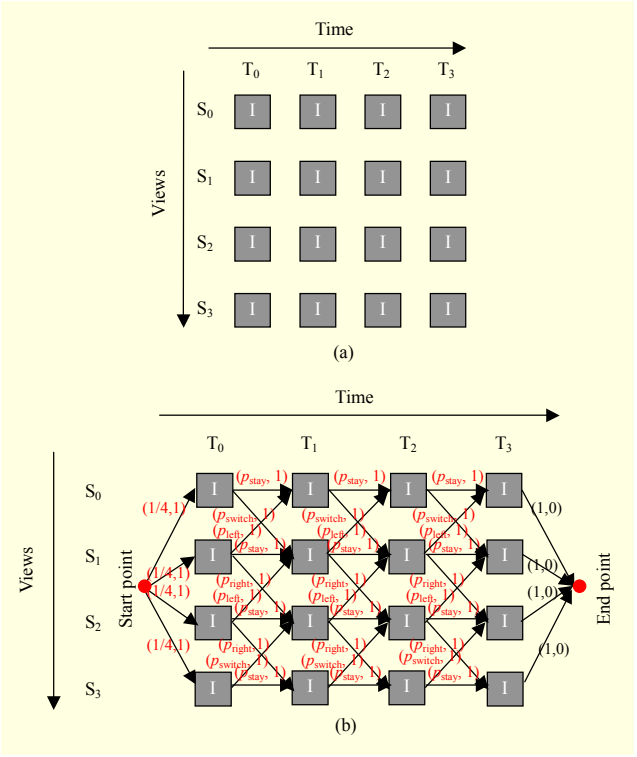



Fig. 3. (a) MVC sample scheme 2 and (b) weighted random graph model $G$ for sample scheme 2.

Table 2. Calculation for weighted random graph model in Fig. 3(b).

| No. | Random path $P$ | $\mathbf{P}(P)$ | $W(P)$ | $E(P)$ |
|---|---|---|---|---|
| 1 | Start, T0S0, T1S0, T2S0, T3S0, End | 0.2473 | 4 | 0.9892 |
| 2 | Start, T0S0, T1S1, T2S0, T3S0, End | 0.1614E-5 | 4 | 0.6457E-5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 68 | Start, T0S3, T1S3, T2S3, T3S3, End | 0.2473 | 4 | 0.9892 |
| $E(G)$ | | 4 | | |

and $N$ are both equal to 4 for these two schemes. In accordance with (2) to (5), the calculation for average interactivity of IMVS can be summarized as a 4-step algorithm.

1: Find out all the possible random paths $P$.
2: Obtain $\mathbf{P}(P)$ and $W(P)$ by (3) and (4), respectively.
3: Compute $E(P)$ by (2).
4: Determine $E(G)$ by (5).

Given the MVC schemes in Figs. 1(a) and 3(a), the calculation for these two schemes are listed in Tables 2 and 3, respectively. There are 68 random paths in total for the two models. For each of these paths, $\mathbf{P}(P)$, $W(P)$, $E(P)$, and $E(G)$ are obtained by (3), (4), (2), and (5), respectively.

Table 3. Calculation for weighted random graph model in Fig. 1(b).

| No. | Random path $P$ | $\mathbf{P}(P)$ | $W(P)$ | $E(P)$ |
|---|---|---|---|---|
| 1 | Start, T0S0, T1S0, T2S0, T3S0, End | 0.2473 | 4 | 0.9892 |
| 2 | Start, T0S0, T1S1, T2S0, T3S0, End | 0.1614E-5 | 16 | 0.2583E-4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 68 | Start, T0S3, T1S3, T2S3, T3S3, End | 0.2473 | 6 | 1.4839 |
| $E(G)$ | | 7.5211 | | |

Several phenomena can be found in the calculation of the two schemes. First, the two schemes both have 68 random paths in their correspondent graph model due to the same model size and structure. In other words, the number of $P$ is determined by the model size and structure. The model size, that is, the group of picture (GOP) size, is restricted to the number of views and time-stamps, and model structure is related to the cardinality of $S$. Therefore, different graph models will have the same number of $P$ if these schemes are with the same GOP size, and this is convenient for comparative study.

Second, the possibility $\mathbf{P}(P)$ is equivalent for correspondent random path in different graph models. For example, the values of $\mathbf{P}(P)$ for the first $P$ in Fig. 1(b) and Fig. 3(b) are both 0.2473. This is result from the same possibility and edge combination in $P$.

Third, the difference between two graph models is the value of $W(P)$, thus it will result in different $E(P)$ and $E(G)$. The value of $W(P)$ is result from the prediction structure of the MVC scheme. The correspondent frame in the same position of different prediction structures will have different frame-cost.

Finally, given different MVC schemes with the same GOP size, they will have different average interactivity. There are three factors that can affect the value of $E(G)$, including $f_c$, $p_{\text{state }i}$, and the cardinality of $S$. The latter two factors are determined by user behavior, which is settled and can be treated as a constant. Therefore, the average interactivity $E(G)$ is determined by the factor $f_c$, that is, the prediction structure of MVC scheme.

## III. JMVM Scheme Discussions

In this section, we will apply our weighted random graph model $G$ to the well known JMVM prediction structure [7], which has been shown to be efficient in terms of RD performance. The JMVM prediction scheme is a rather fixed prediction structure that makes use of hierarchical B frames on the temporal dimension and interview prediction on the spatial dimension. Schwarz and others [14] presented a detailed

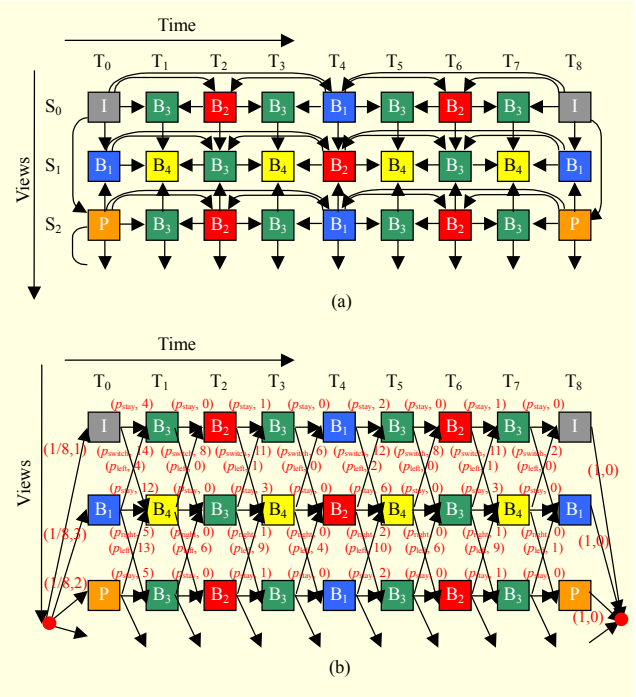



Fig. 4. (a) MVC-HBP scheme in JMVM prediction structures and (b) weighted random graph model $G$ for the JMVM prediction structure in (a).

description of the hierarchical B frame structure. As shown in Fig. 4(a), MVC-HBP is a typical hierarchical prediction structure with three stages of a dyadic hierarchy in temporal and interview predictions. Hierarchy levels of the prediction structure are denoted by the indices in the frames, and all frames are predicted using only frames of the same or a higher hierarchy level as references. This dyadic hierarchy structure is one of the schemes applied in the JMVM. The correspondent graph model for the MVC-HBP scheme is given in Fig. 4(b). Our graph model can complement RD performance values, providing a three-dimensional characterization for MVC schemes.

The results in Fig. 5(a) show the max-$E(P)$ in each view and the all $E(P)$ of the MVC-HBP graph model. Values of $E(P)$ within view interval $x$ indicate that they belong to those random paths started from S$x$T0. The max-$E(P)$ in each view is the maximal expectation number of decoded frames for $P$ started from the corresponding view at T0. There are 8 values for this category since the classical MVC-HBP scheme is with 8 views. The values of $E(P)$ shows the expected frame costs for all $P$ excluding the above 8 paths with the max-$E(P)$. As can be found in Fig. 5(a), there are 33,942 random paths, but most of the values of $E(P)$ are close to 0.

The results in Fig. 5(b) show the comparative study of max-$E(P)$ in each view and $W(P)$ for all 33,942 random paths. The random path with max-$E(P)$ will be with the minimal $W(P)$ in its correspondent view interval $x$. On the other hand, those

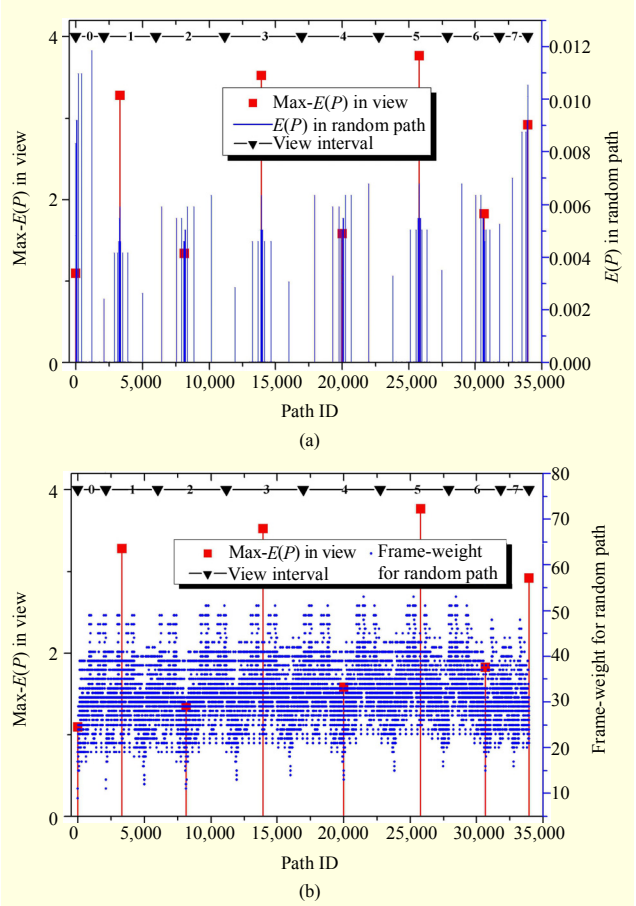



Fig. 5. (a) Comparative results of max-$E(P)$ in view and $E(P)$ in $P$ for MVC-HBP random graph model and (b) comparative results of max-$E(P)$ in view and $W(P)$ for $P$ for MVC-HBP random graph model.

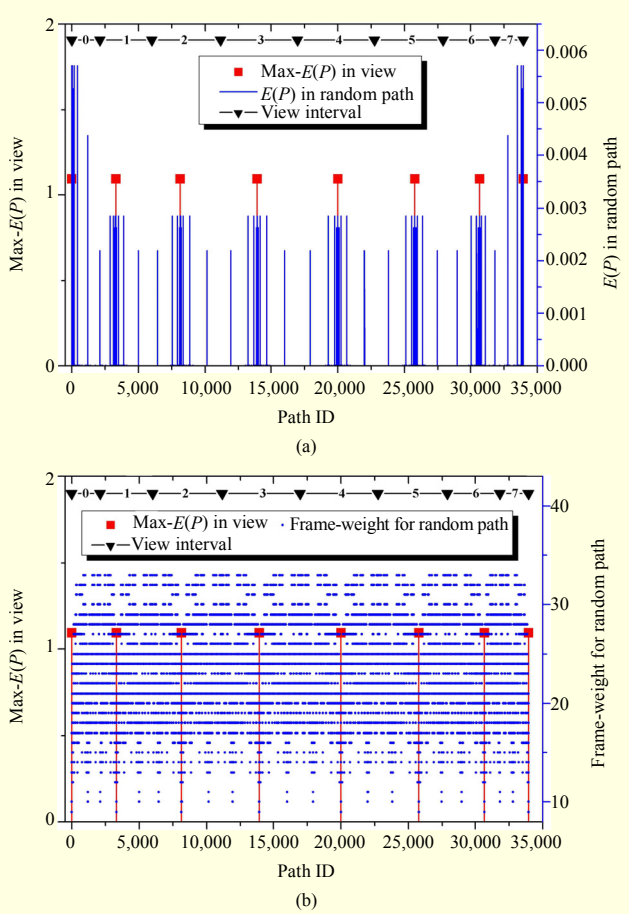



Fig. 6. (a) Comparative results of max-$E(P)$ in view and $E(P)$ in $P$ for Simulcast random graph model and (b) comparative results of max-$E(P)$ in view and $W(P)$ for $P$ for Simulcast random graph model.

random paths with much smaller $E(P)$ will have much higher $W(P)$. More view switching will lead to higher $W(P)$, but lower $\mathbf{P}(P)$ and thus lower $E(P)$.

The results in Fig. 5 show the possibility that $\mathbf{P}(P)$ is important in determining $E$(P) and that view dependencies are important in determining $W$(P).

First, we consider the importance of $\mathbf{P}(P)$ to determine $E$(P).The max-$E(P)$ in each view interval $x$ is obtained on the $P$ that without view switching, that is, the possibility for all directed edges on $P$ is $p_{stay}$. For example, the random path $P$ formed by start point, S0T0, S0T1, S0T2, S0T3, S0T4, S0T5, S0T6, S0T7, S0T8, and end point will be with the max $\mathbf{P}(P)$ 0.1214 in view interval 0.

The value of max-$E(P)$ is much larger than $E(P)$s in its correspondent view interval, since most of the $E(P)$ is close to 0. This phenomenon results from the possibility of switching. The user tends to stay in the current view with extra high possibility, as shown in Table 1 and discussed before. Therefore, the $\mathbf{P}(P)$ will be extra small for $P$ with more view switches. As can be found in Fig. 5(b), the $W(P)$ for $P$ with more view switches is much higher than the path without switching. However, more times of view switching will result in higher $W(P)$, but lower $\mathbf{P}(P)$ and $E(P)$. As depicted by Fig. 5(a), there are 33,942 random paths in MVC-HBP graph model, but only 120 of them can have $E(P)$ lager than 0.002.

Next, we examine the importance of view dependencies in determining $W(P)$. The view-independent random path will have the minimal $W(P)$ within its correspondent view interval $x$. For example, the view S0 is encoded temporally without any interview predictions, and the $P$ formed by start point, S0T0, S0T1, S0T2, S0T3, S0T4, S0T5, S0T6, S0T7, S0T8, and end point will with the minimal $W(P)$, that is, 9, in view interval 0.

The $W(P)$ for that one-view-dependent $P$ will be higher than that of view-independent one. Longer view dependencies will result in lager $W(P)$. As shown by Fig. 5(b), values of $W(P)$ distributed in view interval 4 are greater than those in view interval 2.

The $W(P)$ for dyadic-view-dependent random paths will be

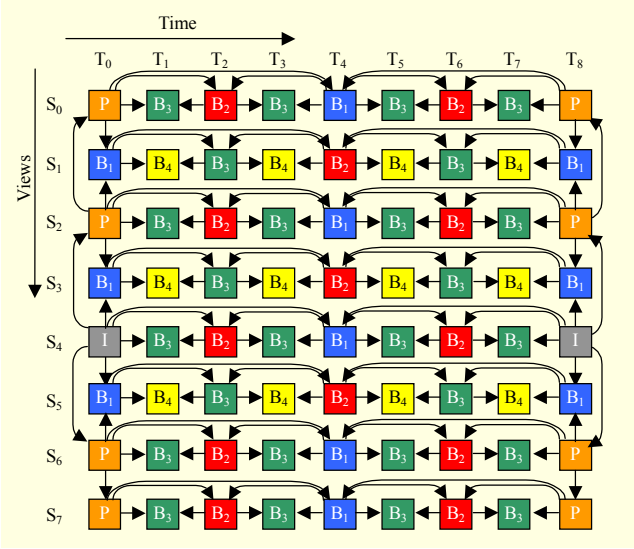



Fig. 7. Proposed MVC scheme based on JMVM.

much higher than that of view-independent and one-view-dependent paths. Longer view dependencies will have higher results. As shown by Fig. 5(b), values of $W(P)$ distributed in view interval 5 are greater than those in view interval 1.

More and longer view dependencies will result in lower interactivity. Figure 6 shows the results for Simulcast scheme as a comparative study to Fig. 5. Simulcast is another typical JMVM scheme that processes frames with just temporal predictions. As can be found in Fig. 6, the max-$E(P)$ is smaller than those correspondent in Fig. 5(a), and values of $W(P)$ is distributed regularly in view intervals. As discussed above, higher value of $E(P)$ will result in higher $E(G)$, and thus lower interactivity.

Based on the above analyses, we proposed an MVC scheme in Fig. 7 with shorter view dependencies. In this proposed scheme, the view S4 is encoded with the same method of S0 in the MVC-HBP scheme, while other views change their encoding method accordingly. We trim all the dyadic-view-dependencies for non-anchor frames so as to decrease the interview dependencies of frames.

## IV. Experimental Results

We provide interactivity and RD performance results for MVC schemes by comparing the MVC-HBP and the proposed schemes. In this section, the experiments will be performed on average decoding delay $C$, interactivity and RD performances of MVC schemes, respectively.

### 1. Average Decoding Delay

The average decoding delay is important in setting weights

Table 4. Selected MVC test sequences and experiment settings.

| Sequence | Resolution | Frame rate | Number of encoded frames |
|---|---|---|---|
| Ballroom | 640 × 480 | 25 | 241 |
| Breakdancer | 1024 × 768 | 15 | 97 |
| Exit | 640 × 480 | 25 | 241 |
| Race1 | 640 × 480 | 30 | 617 |

Table 5. Statistical results on decoding delay (ms) for different frame types, resolutions, QPs, and video contents.

| Seq. | Exit | | | | | |
|---|---|---|---|---|---|---|
| Res. | 640 × 480 | | | | | |
| Scm. | Proposed 8×8 | | | MVC-HBP 8×8 | | |
| QP | 22 | 27 | 32 | 22 | 27 | 32 |
| I | 63.13 | 63.63 | 63.75 | 62.00 | 59.75 | 58.50 |
| P | 74.56 | 71.81 | 64.66 | 75.16 | 66.56 | 59.69 |
| B1 | 78.59 | 76.07 | 70.32 | 77.80 | 69.93 | 63.02 |
| B2 | 76.48 | 73.81 | 68.67 | 76.79 | 67.83 | 62.64 |
| B3 | 72.14 | 70.39 | 65.36 | 73.64 | 64.20 | 60.06 |
| B4 | 69.06 | 66.98 | 63.67 | 69.79 | 61.83 | 59.02 |
| Seq. | Breakdancer | | | | | |
| Res. | 1024 × 768 | | | | | |
| Scm. | Proposed 8×8 | | | MVC-HBP 8×8 | | |
| QP | 22 | 27 | 32 | 22 | 27 | 32 |
| I | 155.88 | 147.00 | 144.63 | 173.63 | 157.38 | 143.00 |
| P | 167.50 | 150.41 | 140.28 | 172.66 | 150.25 | 139.66 |
| B1 | 179.61 | 161.93 | 152.36 | 181.80 | 162.89 | 151.86 |
| B2 | 181.00 | 165.25 | 157.98 | 182.27 | 168.29 | 158.65 |
| B3 | 175.94 | 160.66 | 152.61 | 177.43 | 162.00 | 153.89 |
| B4 | 175.04 | 160.31 | 151.94 | 176.10 | 159.81 | 152.83 |

(that is, $w_I$, $w_P$, and $w_{Bl}$) and the parameter $C$ in (1). In this part of the experiment, we will show the average decoding delay for different types of frames under various conditions, including QPs, multiview video contents, and coding schemes. The selected multiview test sequences and experiment settings are listed in Table 4, and the computer is equipped with 2.93G CPU and 4G memory. Each sequence is decoded 4 times on each QP and coding scheme to gather the statistical results. The results of sequences Exit and Breakdancer are listed in Table 5. In this table, I, P, B1, B2, B3, and B4 are the frame types, and Res. is the frame resolution. All the results are measured by milliseconds.

According to the results in Table 5, we find that the decoding delay for the frame is limited in a small range of milliseconds

for one sequence when all required references are ready in memory. For example, the decoding delay for the frame Exit sequence is around 67.81, and Breakdancer is around 161.24 regardless of the frame type, coding QP, and coding scheme. The statistical results for Ballroom and Race1 are also similar to the other sequences. Therefore, we can obtain $C$ easily and set $w_I=w_P=w_{Bl}=1$ for the convenience of theoretical analysis. There should be a gap between physical and theoretical decoding delay, but it will be small in statistical manner.

## 2. Interactivity evaluations for MVC schemes

As mentioned above, more and longer view dependencies in MVC scheme will result in lower interactivity. In this subsection of experiments, we select several typical MVC schemes in JMVM, including Simulcast, KS-IBP, KS-IPP, KS-PIP, MVC-HBP, and the proposed scheme, to find the effects of view dependency on interactivity. All of these schemes are with equivalent GOP size, that is, 8×8, but with a different number of view dependencies. The details of KS-IBP, KS-IPP, KS-PIP, Simulcast, and MVC-HBP are described in [4]. The results of $E(G)$ for correspondent schemes are listed in Table 6.

In these six MVC schemes, MVC-HBP is the most complex scheme that with more and longer view dependencies, resulting in the highest value of $E(G)$ and the lowest interactivity. On the other hand, the schemes KS-IPP, KS-IBP, and KS-PIP are with similar number but different length of view dependencies. The length of view dependency in KS-IPP is longer than that of KS-IBP, and thus the value of $E(G)$ for KS-IPP is greater than that of KS-IBP. The two schemes KS-IBP and KS-PIP are with similar dependency length, but KS-PIP is simpler, and the value of $E(G)$ changes accordingly. Furthermore, Simulcast is a special scheme without view dependency, and thus it has the lowest value of $E(G)$ and highest interactivity among these schemes. When comparing KS-PIP with the proposed scheme, KS-PIP has longer but simpler view dependencies, and thus the values of $E(G)$ for the two schemes are closed to each other.

Table 6 also lists the interactivity savings $\Delta E(G)$% for the proposed scheme when compared to other schemes. The value of $\Delta E(G)$% is obtained as

$$\Delta E(G) = (E_o - E_p)/E_o \times 100\ (\%), \quad (6)$$

where $E_p$ is the value of $E(G)$ for the proposed scheme, $E_o$ is the value of $E(G)$ for the KS-IPP, KS-PIP, KS-IBP, MVC-HBP, or Simulcast scheme. It can be found that the interactivity of the proposed scheme is better than MVC-HBP by 37.1130%.

We also extend the GOP size of the proposed scheme to 8×12 and compare the interactivity performance to the MVC-HBP (long) scheme. The results in Table 6 show that the proposed scheme will have 59.8860% savings on $E(G)$.

Table 6. Interactivity savings of the proposed scheme compared to other MVC schemes.

| Name of scheme | Interactivity $E(G)$ | $\Delta E(G)$ (%) |
|---|---|---|
| MVC-HBP | 19.5918 | 37.1130 |
| KS-IPP | 16.6821 | 26.1442 |
| KS-IBP | 14.3209 | 13.9670 |
| KS-PIP | 13.0576 | 5.6435 |
| Simulcast | 9.2094 | -33.7840 |
| Proposed | 12.3207 | - |
| MVC-HBP (long) | 40.8079 | 59.8860 |
| Proposed (long) | 16.3697 | - |

Table 7. Comparative results of interactivity on MVC schemes with different GOP size.

| Name of scheme | GOP size | $E(G)/N$ | Name of scheme | GOP size | $E(G)/N$ |
|---|---|---|---|---|---|
| MVC-HBP | 8×8 | 2.1769 | Simulcast | 8×8 | 1.0233 |
| KS-IPP | 8×8 | 1.8536 | MVC-HBP (long) | 8×12 | 3.6928 |
| Proposed | 8×8 | 1.3690 | Tree mode | 4×4 | 1.5661 |
| KS-IBP | 8×8 | 1.5912 | Transposed | 5×5 | 4.6000 |
| KS-PIP | 8×8 | 1.4508 | All I frame | $N$×$M$ | 1.0000 |

The evaluation function (5) is able to predict the interactivity performance for all MVC schemes. The value of $E(G)$ indicates the average number of decoded frames for watching $N$ frames. Therefore, for any scheme with GOP size $N$×$M$, the value of $E(G)/N$ is the average number of decoded frames for switching from current frame $S_iT_j$ to next frame $S_{i(+0,+1,-1)}T_{j+1}$. In this case, the value $E(G)/N$ is able to compare the interactivity among different schemes with different GOP sizes. In this subsection, we select the previous 6 schemes and the MVC-HBP (long) [4], Tree mode [6], and Transposed [15] schemes for interactivity comparison. Furthermore, we select All I frame scheme for the convenience to figure out the lower bound of interactivity. In this scheme, all frames are intra-coded without any temporal or interview predictions, and thus it can be with any GOP size. The interactivity results for these schemes are listed in Table 7.

The experimental results in Table 7 show that the interactivity performance for MVC schemes can be compared even though they are with different GOP sizes. The All I frame scheme has the lowest value of $E(G)/N$ since it has no temporal or interview prediction. Any switching from $S_iT_j$ to $S_{i(+0,+1,-1)}T_{j+1}$ will result in one frame decoded in this scheme. Therefore, the value of $E(G)/N$ for All I frame scheme is the lower bound for

all MVC schemes.

The results in Table 7 for different schemes also show the effect of view dependency on interactivity. Experimental results indicate that more and longer dependencies will result in lower interactivity, which shows the agreement to previous analyses.

### 3. Experimental results on RD performance

We use JMVM software version 7.0 [7] to test the RD performances of MVC-HBP and the proposed schemes. The MVC test sequences selected for RD performance and their settings are listed in Table 4. All RD tests were performed using the following fixed QP values: 22, 27, 32, and 37, and all DeltaLayerQuants are set as 0. The MVC-HBP and the proposed scheme with the same 8×8 and 8×12 GOP size are compared. These comparisons will be impartial on the interactivity evaluation model.

The RD performances for different test sequences and four selected schemes are shown in Fig. 8. For schemes with the same GOP size, they have the same temporal prediction structure within one view, but adopt different prediction dependencies among views. Therefore, the comparison settings among schemes are with the same GOP size and the same temporal prediction structure, but with different view prediction dependency. As we have mentioned above, the interactivity performance will be affected by the number and length of view dependency in one scheme. In this subsection, we will further show the effect on RD performance.

According to the results in Fig. 8, the MVC-HBP or MVC-HBP (long) scheme with more complex and longer view dependencies have better RD performance in almost all of the test sequences. However, the coding gain is quite small when compared to the proposed scheme. We use BDPSNR and BDBR algorithms in [16] to figure out the RD performance of the proposed scheme when comparing it to the MVC-HBP. The results in Tables 8 and 9 show that the proposed scheme has similar RD performance with the MVC-HBP scheme on four test sequences when different GOP sizes are selected. On the other hand, the results in Table 6 show that the proposed and the proposed (long) scheme can save 37.1130% and 59.8860% interactivity when comparing it to the MVC-HBP and MVC-HBP (long) scheme, respectively. Therefore, an IMVS can have more interactivity gains with small RD loss when it chooses the proposed scheme at the encoder rather than the MVC-HBP. The proposed scheme will be an alternative selection with high interactivity and appropriate RD performance for IMVS.

Table 8. Parameters of BDBR and BDPSNR for the proposed scheme comparing to MVC-HBP in JMVC with 8×8 GOP size.

| Sequence | BR range (kbps) | BDPSNR |
|---|---|---|
| Ballroom | [327.2107, 3091.0945] | 0.0151 |
| Breakdancer | [267.2797, 3847.6909] | 0.0088 |
| Exit | [171.9791, 2169.9734] | –0.1791 |
| Race1 | [496.8723, 3987.3069] | –0.0411 |
| Sequence | PSNR range (dB) | BDBR |
| Ballroom | [32.7436, 41.0606] | 0.2958 |
| Breakdancer | [35.8248, 40.7059] | 0.2903 |
| Exit | [35.2829, 41.5753] | –7.6353 |
| Race1 | [33.3885, 42.2170] | –1.0977 |

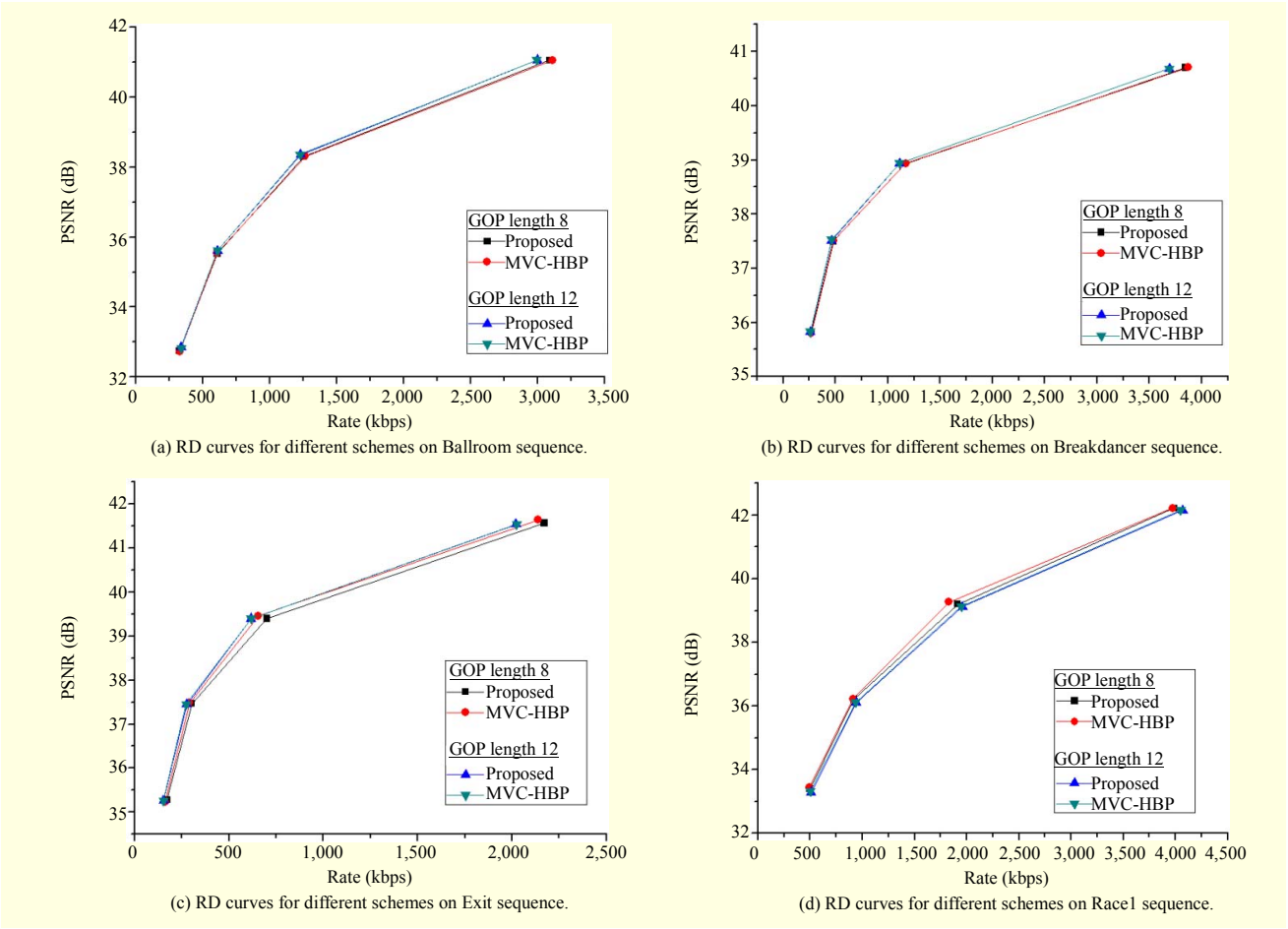
Table 9. Parameters of BDBR and BDPSNR for the proposed scheme comparing to MVC-HBP in JMVC with 8×12 GOP size.

| Sequence | BR range (kbps) | BDPSNR |
|---|---|---|
| Ballroom | [338.6926, 2999.0107] | –0.0075 |
| Breakdancer | [259.6464, 3690.7524] | –0.0109 |
| Exit | [152.7541, 2021.4348] | –0.0068 |
| Race1 | [509.5121, 4072.0627] | –0.0434 |
| Sequence | PSNR range (dB) | BDBR |
| Ballroom | [32.8599, 41.0712] | –0.2412 |
| Breakdancer | [35.8214, 40.6859] | –0.8432 |
| Exit | [35.2615, 41.5419] | –0.3459 |
| Race1 | [33.2495, 42.1587] | –1.1479 |

## V. Conclusion

We presented a new interactivity evaluation model for MVC schemes, and found that some of the considerations have to be taken into account for a high interactivity design for the scheme. This new interactivity evaluation model allows us to analyze the interactivity efficiency of MVC prediction structure, and provides a new tool in MVC scheme designing.

We also discussed the prediction structure of JMVM considering both RD and interactivity performance. The factors that can affect the interactivity performance were figured out, and we found these factors are important to both interactivity and RD performances. Furthermore, we modified the JMVM prediction structure and obtained the proposed scheme. The experimental results show that the new scheme has a similar RD performance compared to the MVC-HBP scheme but considerably gains on interactivity performance. The proposed




(a) RD curves for different schemes on Ballroom sequence.

(b) RD curves for different schemes on Breakdancer sequence.

(c) RD curves for different schemes on Exit sequence.

(d) RD curves for different schemes on Race1 sequence.

Fig. 8. RD performances for different schemes with same GOP size.

scheme can be an alternative selection for multiview video encoder in future interactive multiview video system with high interactivity and compression requirements. Finally, the results show that the interactivity issue is significant for future multiview applications when selecting an MVC scheme.

## References


[1] C. Fehn and P. Kauff, "Interactive Virtual View Video (IVVV) - The Bridge between Immersive TV and 3D-TV," *Proc. SPIE Three-Dimensional TV, Video and Display I*, Boston, MA, USA, Aug. 2002, pp. 14-25.

[2] A. Smolic et al., "3D Video and Free Viewpoint Video-Technologies, Applications and MPEG Standards," *IEEE Int. Conf. Multimedia Expo (ICME)*, Toronto, ON, Canada, Jul. 2006.

[3] ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MVC," MPEG2006/W8019, Montreux, Switzerland, Apr. 2006.

[4] P. Merkle et al., "Efficient Prediction Structures for Multiview Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, Nov. 2007, pp. 1461-1473.

[5] K.J. Oh and Y.S. Ho, "Multiview Video Coding Based on the Lattice-Like Pyramid GOP Structure," *Picture Coding Symp.*, Beijing, China, Apr. 2006.

[6] Y. Yang et al., "Hyper-Space Based Multiview Video Coding Scheme for Free Viewpoint Television," *Picture Coding Symp.*, Beijing, China, Apr. 2006.

[7] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, "Joint Multiview Video Model (JMVM) 7.0," JVT-Z207, Antalya, Turkey, Jan. 2008.

[8] H. Kalva et al., "Challenges and Opportunities in Video Coding for 3D TV," *IEEE Int. Conf. Multimedia Expo*, 2006, pp. 1689-1692

[9] J.G. Lou, H. Cai, and J. Li, "Interactive Multiview Video Delivery Based on IP Multicast," *Advances in Multimedia*, Jan. 2007, doi:10.1155/2007/97535.

[10] Y. Zhang, M.I. Yu, and G. Jiang, "Evaluation of Typical Prediction Structures for Multi-View Video Coding," *ISAST Trans. Electron. Signal Proces.*, vol. 2, no. 1, 2008, pp. 7-15.

[11] ISO/IEC JTC1/SC29/WG11, "Requirements on Multi-View Video Coding v.6," N8064, Montreux, Switzerland, Apr. 2006.

[12] Y. Liu et al., "Multi-View Video Coding with Flexible View-

Temporal Prediction Structure for Fast Random Access," *Pacific Rim Conf. Multimedia, Lecture Notes Computer Science (LNCS)* 4261, 2006, pp. 564-571.

[13] Y. Yang et al., "User Interaction and Random Accessibility Analysis for Multiview Video System," *Int. Conf. Consumer Electronics*, Las Vegas, Jan. 2008, pp. 4-22.

[14] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," *Proc. IEEE Int. Conf. Multimedia Expo*, July 2006, pp. 1929-1932.

[15] ISO/IEC JTC1/SC29/WG11, "Transposed Picture Ordering for Dynamic Light Field Coding," M10929, Redmond, USA, Jul. 2004.

[16] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T SG16 Doc. VCEG-M33, Mar. 2001.

**You Yang** received the BS and MS in applied mathematics from the Chinese University of Mining and Technology in 2000 and 2003, respectively, and the PhD in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences, in 2009. He is now a post-doctoral fellow in the Department of Automation, Tshinghua University. His research interests include digital image and video coding, three-dimensional image modeling and representation, multiview and free viewpoint video, 3D television, and image quality assessment in the above fields.

**Qionghai Dai** received the BS in mathematics from Shanxi Normal University, China, in 1987, and the ME and PhD in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the faculty of Tsinghua University, Beijing, China, where he is currently a professor and the Director of the Broadband and Networks and Digital Media Laboratory. His research areas include signal processing, multimedia networks, and computer vision and graphics.

**Gangyi Jiang** received his MS from Hangzhou University in 1992, and his PhD from Ajou University, Korea, in 2000. He is now a professor in the Faculty of Information Science and Engineering, Ningbo University, China, and also works in State Key Laboratory for Novel Software Technology, Nanjing University. His research interests include digital video compression and communications, multiview video coding, and image processing.

**Yo-Sung Ho** received the BS and MS in electronic engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the PhD in electrical and computer engineering from the University of California, Santa Barbara, in 1990. He joined ETRI, Daejon, Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, New York. In 1993, he rejoined the technical staff of ETRI. Since 1995, he has been with Gwangju Institute of Science and Technology (GIST) in Korea, where he is currently a professor of the School of Information and Communications. Since 2003, he has been the Director of the Realistic Broadcasting Research Center (RBRC) at GIST. He is currently an associate editor of the *IEEE Trans. on Circuits and Systems for Video Technology (CSVT)*. His research interests include digital image and video coding, advanced source coding techniques, three-dimensional image modeling and representation, multiview and free viewpoint video, 3D television, and realistic broadcasting systems.