Journal of Electronic Imaging

JElectronicImaging.org

Enhanced view random access ability for multiview video coding

Seif Allah Elmesloul Nasri
Khaled Khelil
Noureddine Doghmane

SPIE. IS&T imaging.org

Seif Allah Elmesloul Nasri, Khaled Khelil, Noureddine Doghmane, "Enhanced view random access ability for multiview video coding," *J. Electron. Imaging* **25**(2), 023027 (2016), doi: 10.1117/1.JEI.25.2.023027.

Enhanced view random access ability for multiview video coding

Seif Allah Elmesloul Nasri,[a,*] Khaled Khelil,[b] and Noureddine Doghmane[a]
[a]University Badji Mokhtar of Annaba, Faculty of Engineering Sciences, Department of Electronics, B.P.12, Annaba 23000, Algeria
[b]University Mohamed-Cherif Messaadia of Souk Ahras, Faculty of Sciences and Technology, LEER Lab, B.P.1553, Souk Ahras 41000, Algeria

Abstract. Apart from the efficient compression, reducing the complexity of the view random access is one of the most important requirements that should be considered in multiview video coding. In order to obtain an efficient compression, both temporal and inter-view correlations are exploited in the multiview video coding schemes, introducing higher complexity in the temporal and view random access. We propose an inter-view prediction structure that aims to lower the cost of randomly accessing any picture at any position and instant, with respect to the multiview reference model JMVM and other recent relevant works. The proposed scheme is mainly based on the use of two base views (I-views) in the structure with selected positions instead of a single reference view as in the standard structures. This will, therefore, provide a direct inter-view prediction for all the remaining views and will ensure a low-delay view random access ability while maintaining a very competitive bit-rate performance with a similar video quality measured in peak signal-to-noise ratio. In addition to a new evaluation method of the random access ability, the obtained results show a significant improvement in the view random accessibility with respect to other reported works. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.25.2.023027]

Keywords: multiview; video coding; random access; bit rate; peak signal-to-noise ratio.

Paper 15743 received Oct. 5, 2015; accepted for publication Mar. 28, 2016; published online Apr. 25, 2016.

1 Introduction

Multiview video is a three-dimensional (3-D) scene captured by at least two cameras located at different viewpoints. Multiview video systems are used in various applications such as 3-D cinemas, gaming, education, surveillance, immersive telepresence and videoconference, three-dimensional television, and free viewpoint television.[1] Practically, all the above-mentioned applications share similar components of the processing scheme, which is made of four basic parts: capturing, coding, transmission, and display. In comparison with a conventional two-dimensional video sequence, the 3-D scene representation usually requires a much larger amount of data. Consequently, efficient compression for data storage or transmission with less degradation and delay over limited bandwidth represents a challenging task.

The simplest solution for encoding this type of video is to encode each view independently with a video codec such as H.264/MPEG-4 AVC.[2] This method is called simulcast compression. However, this encoding scheme does not exploit the great similarity that exists between the different views of the same scene. This similarity means redundant information in the coding process, which could be eliminated by a combination of the motion estimation in the temporal level and the disparity estimation in the inter-view level. Based on the exploitation of both temporal and inter-view prediction, Merkle et al. introduced an approach that ensured a good tradeoff between the bit rate and the video quality.[3] It was adopted and implemented by the Joint Video Team of ISO/IEC, Moving Picture Experts Group and ITU-T, and Video Coding Experts Group in a reference model named the Joint Multiview Video Model (JMVM).[4]

As with any video coding standard, the main requirement for multiview video coding (MVC) is a high compression efficiency. Other requirements specific to MVC involve low-delay random temporal and view access, which are desirable features in video communication systems.[5] These requirements ensure the interactive capability to the coding scheme so that any image can be accessed, decoded, and displayed with a lower outlay of complexity.

Many research studies have proposed different MVC structures to meet the MVC requirements. In Ref. 6, an MVC algorithm based on distributed source coding is proposed to tackle the free viewpoint switching problem of compression efficiency. Even though it outperforms the solutions based on intra- or closed-loop predictive coding, it is less efficient compared to the H.264/AVC standard. Similarly, in Ref. 7, three approaches are proposed, including switching predicted/intra-coded frames frame coding, interleaved view coding, and secondary representation coding, which provide reduced delay view random access. However, the performance is inferior to MVC, the standard extension of H.264/AVC. Zhang et al.[8] proposed a method to adaptively select the best prediction mode among a set of given structures. This approach, which is based on a spatiotemporal correlation analysis using Lagrange cost, provides significant enhancement of the view random access, but with an additional encoding delay and a higher consumption of memory resources. In Ref. 9, Yang et al. suggest a prediction structure based on the enhancement of the encoding order of the B pictures and their reference frames as an extension of each independent view of the multiview video by applying a binary tree. This approach leads to a significant improvement in the bit-rate

*Address all correspondence to: Seif Allah Elmesloul Nasri, E-mail: seifallah.nasri@univ-annaba.org

performance but shows slow view random access due to the increased coding complexity.

In this paper, an inter-view prediction structure, aimed at improving the random access performance while maintaining high compression efficiency, is proposed. It consists of using two base views (I-view) with selected positions in a scheme of eight views, which inherently introduces a view random access delay and may decrease the compression efficiency of the coder. Nevertheless, this disadvantage is overcome through the use of four bi-predictive views (B-view), where two of them are successive. A generalization of the proposed scheme, for structures containing more than eight views, is then developed. In addition, a new evaluation method to fully assess the random access capability for the MVC is established.

The rest of this paper is organized as follows. Section 2 addresses the multiview video coding and summarizes the fundamental coding schemes including simulcast, intracoded pictures, predicted pictures, predicted pictures (IPP), intracoded pictures, bipredicted pictures, predicted pictures (IBP), and a recent related work in order to compare them later with our approach. The proposed inter-view prediction structure is described in Sec. 3. The random access enhancement and its evaluation are detailed in Sec. 4. Experimental results of the compression efficiency are presented and discussed in Sec. 5. Finally, some concluding remarks are given in Sec. 6.

2 Overview of the Multiview Video Coding Schemes

Typically, all video sequences are characterized by two types of correlation. The first one is the spatial redundancy, which exists in every picture as a natural space correlation. This redundancy is reduced using intra-picture prediction based on the intra-block matching. The second one is temporal redundancy, which means adjacent frames are highly correlated, called the interframe correlation. In addition to these two previous types of redundancies, MVC also employs the correlation between the neighboring views of the captured scene, referred to as inter-view correlation.

The simplest method for MVC is the simulcast scheme, which performs the compression of the multiview video by exploiting the spatiotemporal redundancies only and coding each view independently using a conventional video codec. By making use of H.264/AVC and the hierarchical B pictures, video compression has been efficiently improved in comparison to the traditional simulcast coding structures.[10] Figure 1 depicts the hierarchical B pictures structure where the number of frames in the group of pictures (GOP) is 8. The first picture is independently coded as an instantaneous decoder refresh (IDR) picture, and the so-called anchor or key pictures are coded in regular intervals. The B pictures, located between two I pictures and referred as nonanchor frames, are hierarchically predicted using the concept of hierarchical B pictures.

The simulcast compression for the multiview video is illustrated in Fig. 2. In this case, each group of groups of pictures (GGOP) is composed of eight views and eight pictures per GOP. S_n indicates the different views (cameras), while T_n represents the time location.

Simulcast compression is characterized by its fast view random access because all key pictures used within the prediction structure are independently intracoded. However, its coding efficiency is not optimal due to not exploiting the inter-view statistical dependencies. Usually, simulcast

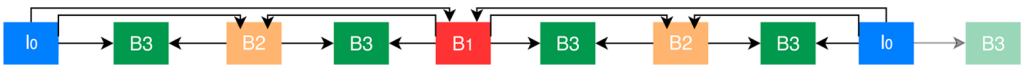

Fig. 1 Hierarchical B pictures structure.

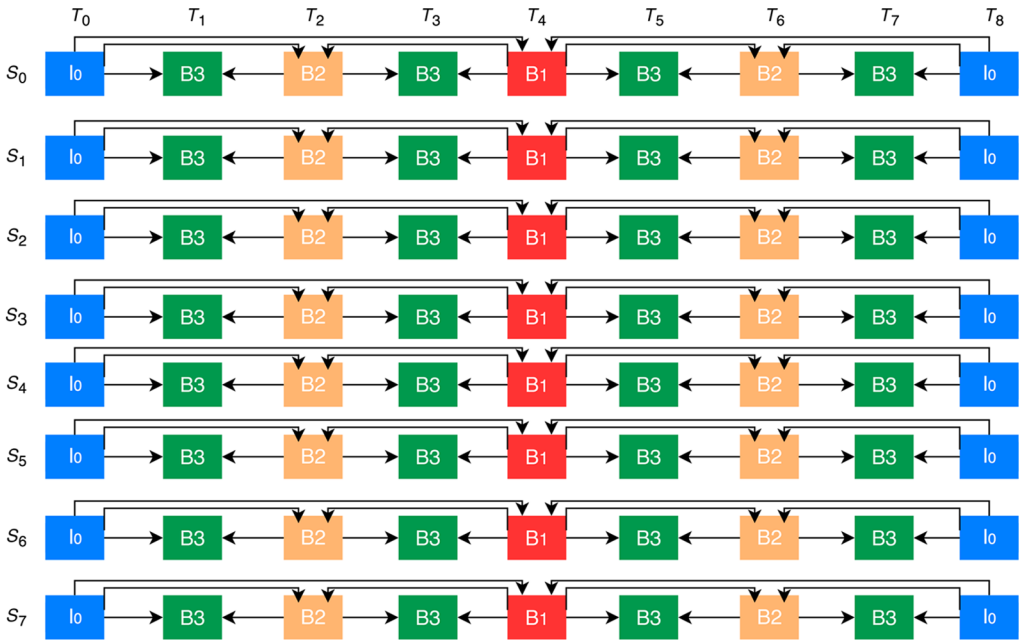

Fig. 2 Simulcast compression scheme.

coding is typically employed as a reference model for coding performance comparisons between different MVC schemes.

The analysis of temporal and inter-view prediction efficiency indicates that using simultaneously temporal and inter-view reference frames improves the coding efficiency.[6] This is due to the use of disparity or motion-compensated prediction to eliminate the redundant information between neighboring views, or successive frames respectively.

To satisfy the main MVC requirements, many inter-view prediction structures are proposed. These proposed schemes differ in several criteria, including the number of reference frames used either by anchor or nonanchor pictures, as well as the type of anchor pictures of different views, which can be I, P, or B pictures.

Figure 3 illustrates an IPP prediction structure[3] with a multiview sequence employing eight cameras and GOP length of 8. This IPP structure uses one intracoded picture per GGOP. The first view, S_0, called the base view, remains the same and always begins with an IDR picture. All the remaining views, S_1 to S_7, called P-views, begin with an anchor picture predicted from the previous anchor picture I/P. The nonanchor pictures of these views are predicted according to the temporal level and the inter-view level. In other words, each nonanchor frame is predicted using two pictures from the temporal level and one picture from the previous view. For example, the picture located at (S_1, T_4) is predicted from (S_1, T_0) and (S_1, T_8) in the temporal level and from (S_0, T_4) in the inter-view level. Generally, an IPP inter-view prediction structure achieves a significant gain in bit rate and video quality in comparison with a simulcast structure. On the other hand, IPP increases the coding complexity and slows down the view random access ability.

According to Ref. 11, the random access ability is measured by the maximum number of reference images $N_{\max}$ needed for decoding a given image. For the IPP structure, this number is defined as

$$N_{\max} = (H_{\max} + 1) + 5 \times [Nbr_{\text{view}} - 1], \quad (1)$$

where $H_{\max}$ is the highest level of B pictures in a hierarchical B picture coding structure ($H_{\max}$ is set to 3 for IPP), and Nbr_{view} represents the number of views in the structure.

The IBP prediction structure proposed in Ref. 6 is depicted in Fig. 4. This scheme, which exploits the inter-view correlation, was adopted as a default structure of JMVM. It uses three types of views: I-view (as one base view per GGOP), P-views (S_2, S_4, S_6, and S_7), and B-views (S_1, S_3, and S_5). Each B-view is located between an I/P-view and a P-view. The B view begins with a B key picture, which is bidirectionally predicted from I/P and P key pictures. The nonkey pictures of the B views are predicted using four pictures: two from the temporal level and the other two from the inter-view level. For example, the picture located in (S_1, T_4) is predicted from (S_1, T_0) and (S_1, T_8) in the temporal level, and from (S_0, T_4) and (S_2, T_4) in the inter-view level. Generally, the IBP structure ensures a good trade-off between bit rate and view random access. Compared to the simulcast scheme, an IBP coding structure allows a considerable gain in bit rate and video quality, because of the exploitation of the inter-view correlation. Also, the IBP structure allows improvement in view random access ability with respect to the IPP structure. The following equation gives the $N_{\max}$ of the IBP structure:

$$N_{\max} = 3 \times H_{\max} + 2 + 5 \times [Nbr_{\text{view}} - 1], \quad (2)$$

where $H_{\max}$ is set to 4 for the IBP structure.

In Ref. 12, an approach based on the use of two successive B-views between two views of types I and P is proposed. It has been shown that the use of successive B-views improves the gain in bit rate and considerably speeds up the view random access. The hierarchical level of the used B images is similar to that of the IBP structure. The gain in

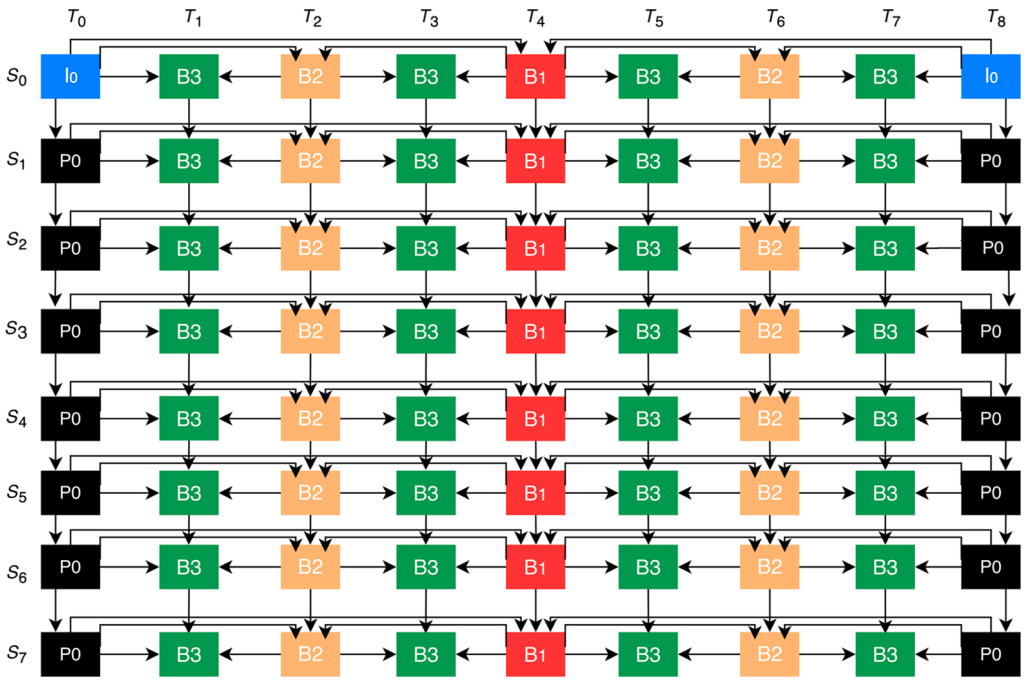

Fig. 3 IPP prediction structure.

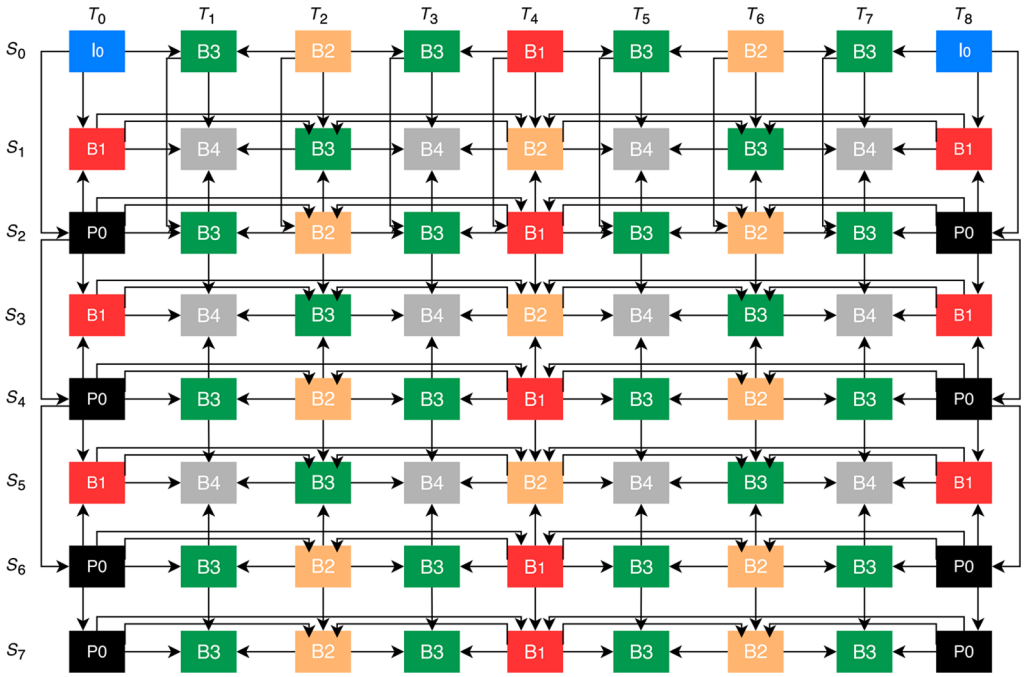

Fig. 4 IBP prediction structure.

bit rate obtained in this structure is due to the larger number of B-views, which provides a lesser bit rate than the I-view and P-views. The $N_{\max}$ used in Ref. 12 is given by

$$N_{\max} = 3 + H_{\max} + 2 \times [(Nbr_{\text{view}} - 2)/3]. \quad (3)$$

Currently, the improvements obtained for the random access ability need to be enhanced further to ensure more interactivity to the multiview video users while keeping a high video quality and a good bit rate saving.

3 Random Access Enhancement

3.1 *Proposed Approach*

Generally, the view random access ability could be enhanced through a good selection of the reference view, which ensures a direct prediction of a maximum of the views in the scheme. The middle position of the I-view as shown in Ref. 13, in addition to a regular involvement of the B-views, generally gives better results than the standard structures, where the position of the base view is always set as the first view in the structure.

The proposed structure is chosen carefully to give a direct inter-view prediction for all the views of the structure. It is based on the use of two base views (I), which means that it uses two IDR pictures in one GGOP. These two views are independently coded by the use of the temporal prediction, which is based on the B hierarchy algorithm. In this scheme, S_2 and S_5 are selected as optimal positions for the reference views, allowing a direct inter-view prediction for all remaining views (S_0, S_1, S_3, S_4, S_6, and S_7). The structure contains two P-views, leading to a lower view random access complexity. As is shown in Fig. 5, these views are S_0 and S_7. These two views benefit from direct prediction through S_2 and S_5 as follows. The anchor pictures of the P-views are coded via one anchor picture from the I-views, whereas the nonanchor pictures of the P-views (S_0, S_7) are coded from three reference pictures, two from their temporal level and one from the inter-view level of the reference view. The scheme also includes four B-views (S_1, S_3, S_4, and S_6). The anchor pictures of these views are coded using two reference anchor pictures, while the nonanchor pictures of these views are coded with four reference pictures, divided evenly between the temporal and the inter-view levels. S_1 and S_6 benefit from a direct inter-view prediction. This is also applicable to S_3 and S_4. In addition, there are two successive B-views, which will reduce the maximum number of the decoded frame for accessing a given picture.

Since using B-views offers a better bit rate saving with respect to the use of P-views or I-views, the proposed scheme achieves a competitive bit rate saving with a good video quality even with the use of two reference views (I-view). Indeed, the use of a second I-view instead of P-view or B-view relatively reduces the encoding time duration, because the I-view uses only temporal prediction between its pictures, while B-view and P-view use an inter-view prediction process in addition to the temporal prediction. Also, improving the random access performance means reducing the number of the needed frames for coding or decoding an image in the multiview video scheme. This leads directly to a reduction of the time complexity.

Furthermore, the proposed prediction structure presents an additional backward compatibility due to the use of two base views that could be extracted. The resulting bit stream also complies with the H.264/AVC standard.

The proposed structure, referred to as PBI and depicted in Fig. 5 with eight views, has two main view coding orders that give the same results:

S_2-S_0-S_1-S_5-S_3-S_4-S_7-S_6 I-P-B-I-B-B-P-B

S_2-S_5-S_0-S_1-S_3-S_4-S_7-S_6 I-I-P-B-B-B-P-B

Figure 6 illustrates the difference between the proposed structure PBI and the previously studied structures in Sec. 2 by highlighting the anchor pictures of each structure.

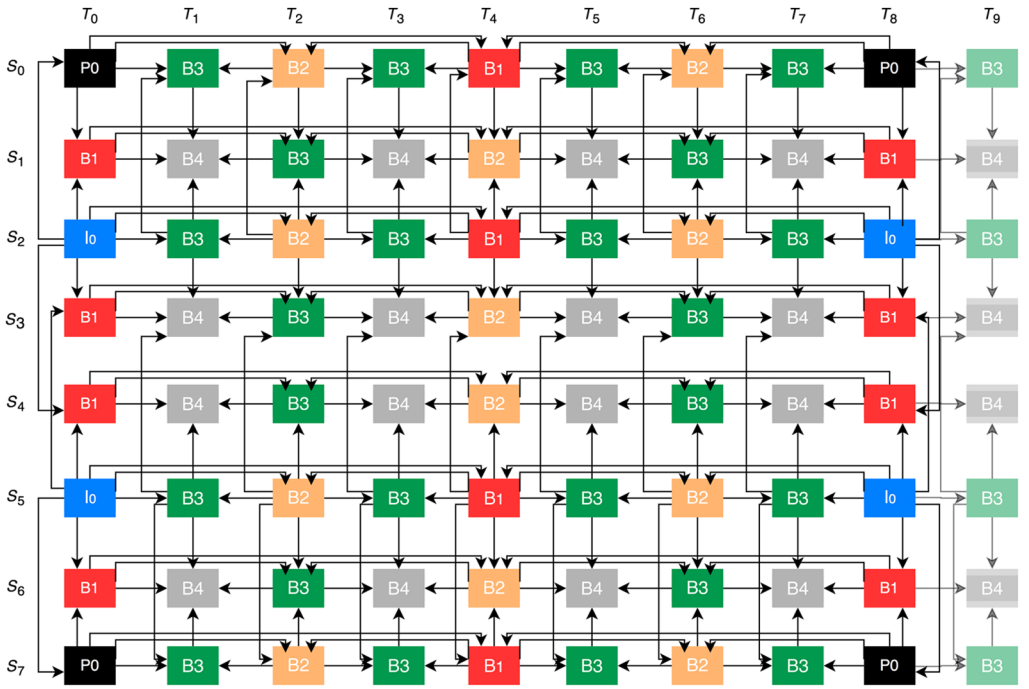

Fig. 5 Proposed prediction structure predicted pictures, bipredicted pictures, intracoded pictures (PBI).

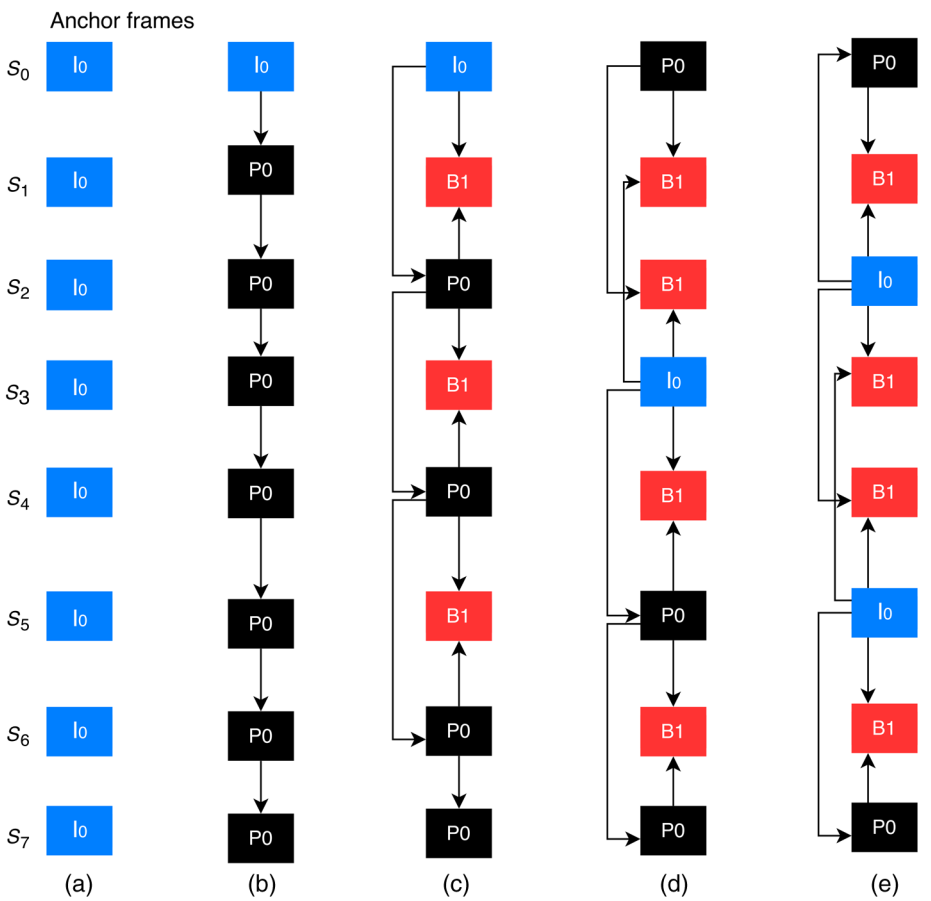

Fig. 6 Anchor picture comparison among (a) simulcast structure, (b) IPP structure, (c) IBP structure, (d) Ref. 12, and (e) proposed structure PBI.

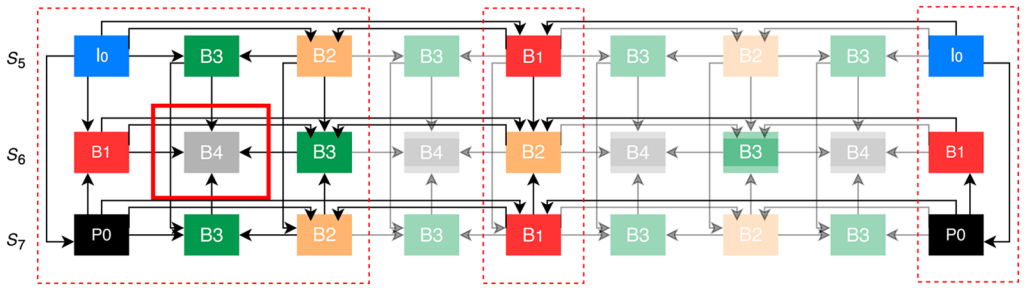

Fig. 7 Random access scheme for a given image.

The picture located in the B-view (S_6) in the position S_6/T_1 in Fig. 7 is noted as the B_4 picture. It has the maximum hierarchical level in the proposed structure PBI. Accessing the picture S_6/T_1 needs four reference pictures in the temporal level, which are $\{S_6/T_0, S_6/T_2, S_6/T_4,$ and $S_6/T_8\}$. It also needs five pictures per neighboring view: I-view (S_5) and P-view (S_7), which are as follows: $\{S_5/T_0, S_5/T_1, S_5/T_2, S_5/T_4, S_5/T_8\}$ and $\{S_7/T_0, S_7/T_1, S_7/T_2, S_7/T_4, S_7/T_8\}$.

The example of the picture S_6/T_1 is repeated four times in the same GOP of the B-view (S_6), as also happens for the rest of B-views (S_1, S_3, and S_4).

Therefore, the formula that describes the computation of the $N_{\max}$ for the proposed inter-view prediction can be deduced as

$$N_{\max} = 3 \times H_{\max} + 2, \tag{4}$$

where the maximum hierarchical level, $H_{\max}$, is equal to 4 in this proposed PBI structure. In order to show the simplicity of the calculation that characterizes our proposed structure PBI, the equations computing the $N_{\max}$ for all the studied structures are reported in Table 1.

To evaluate the random access ability, we have adopted the metric proposed in Ref. 12. It aims at calculating the number of necessary reference frames Nbr_{img} to be decoded for the access of a given picture. After analyzing the different types of pictures composing our proposed PBI structure, the results were as follows:

1. For anchor frames:
   - The number of pictures needed to be coded for accessing I picture is known to be equal to 0 in all structures.
   - For P anchor frames, which are two (S_0 and S_7), the number of anchor frames to be coded is equal to one and is the same for both views.
   - For the four B anchor frames (S_1, S_3, S_4, and S_6), each one needs two frames to be decoded.

Equation (5) summarizes the Nbr_{img} for the type of anchor frame, I, P, and B:

$$Nbr_{\text{img}} = \begin{cases} 0 \text{ for I anchor frame} \\ 1 \text{ for P anchor frame} \\ 2 \text{ for B anchor frame} \end{cases}. \tag{5}$$

2. For nonanchor frames:

The number of decoded pictures for accessing a given picture depends on its hierarchical level and the type of view (I, P, or B) this picture belongs to.

The different possibilities for calculating Nbr_{img} are given by

$$Nbr_{\text{img}} = \alpha \times \text{Hierarchy} + \beta, \tag{6}$$

$$\text{where } \begin{cases} \alpha = 1, \beta = 0 \text{ for I nonanchor frames} \\ \alpha = 2, \beta = 1 \text{ for P nonanchor frames} \\ \alpha = 3, \beta = 2 \text{ for B nonanchor frames} \end{cases}.$$

Table 2 summarizes the various possible cases of calculating Nbr_{img} for our PBI structure and the other two considered schemes. It highlights the simplicity of our equations in comparison to the other two structures.

Table 1 $N_{\max}$ equations comparison.

	$N_{\max}$ equations
IBP	$N_{\max} = (H_{\max} + 1) + 5 \times [Nbr_{\text{view}} - 1]$
IPP	$N_{\max} = 3 \times H_{\max} + 2 + 5 \times [Nbr_{\text{view}} - 1]$
Ref. 12	$N_{\max} = 3 + H_{\max} + 2 \times [(Nbr_{\text{view}} - 2)/3]$
PBI	$N_{\max} = 3 \times H_{\max} + 2$

3.2 *Extending the Proposed Structure*

In this part, we present an extension of the proposed PBI structure for cases where we have more than eight views. Note that this scheme maximizes the use of B-views to ensure a good bit rate gain and high random access ability. On the other hand, the scheme avoids the use of successive P-views that slow down the random access ability.

Figure 8 describes the order of views following the second base view S_5. Note that the order depends on the number of views, with the respect of the condition that guarantees the nonuse of successive P-views. For the proposed PBI structure, three sequencing orders are possible and are generated according to the following equation:

$$\text{Order} = \begin{cases} \text{I/P, B, P} & \text{if } (Nbr_{\text{view}} \bmod 3) = 2 \\ \text{I/P, B, P, B, P} & \text{if } (Nbr_{\text{view}} \bmod 3) = 1 \\ \text{I/P, B, B, P} & \text{if } (Nbr_{\text{view}} \bmod 3) = 0 \end{cases}, \tag{7}$$

where each letter I, P, and B corresponds to a view type, and the different orders represent the last views in every choice.

Table 2 Comparison between the equations computing Nbr_{img}.

Nbr_{img}	P anchor frames	B anchor frames	P nonanchor frames	B nonanchor frames
PBI	1	2	$2 \times \text{Hierarchy} + 1$	$3 \times \text{Hierarchy} + 2$
IBP	$\text{Num}_{\text{view}}/2$	$1 + (\text{Num}_{\text{view}}/2)$	$(\text{Hierarchy} + 1) + 2 \times [\text{Num}_{\text{view}}/2]$	$3 \times \text{Hierarchy} + 2 \times [\text{Num}_{\text{view}}/2]$
Ref. 12	$(\text{Num}_{\text{view}} - 2)/2$	$1 + [(\text{Num}_{\text{view}} - 3)/2]$	$(\text{Hierarchy} + 1) + 2 \times [(\text{Num}_{\text{view}} - 2)/3]$	$3 \times \text{Hierarchy} + 2 \times [(\text{Num}_{\text{view}} - 1)/2]$

Note: Num_{view} denotes the order of the numbers of view, which can be 1, 2…8, and Hierarchy is the hierarchical level of the picture to be consulted.

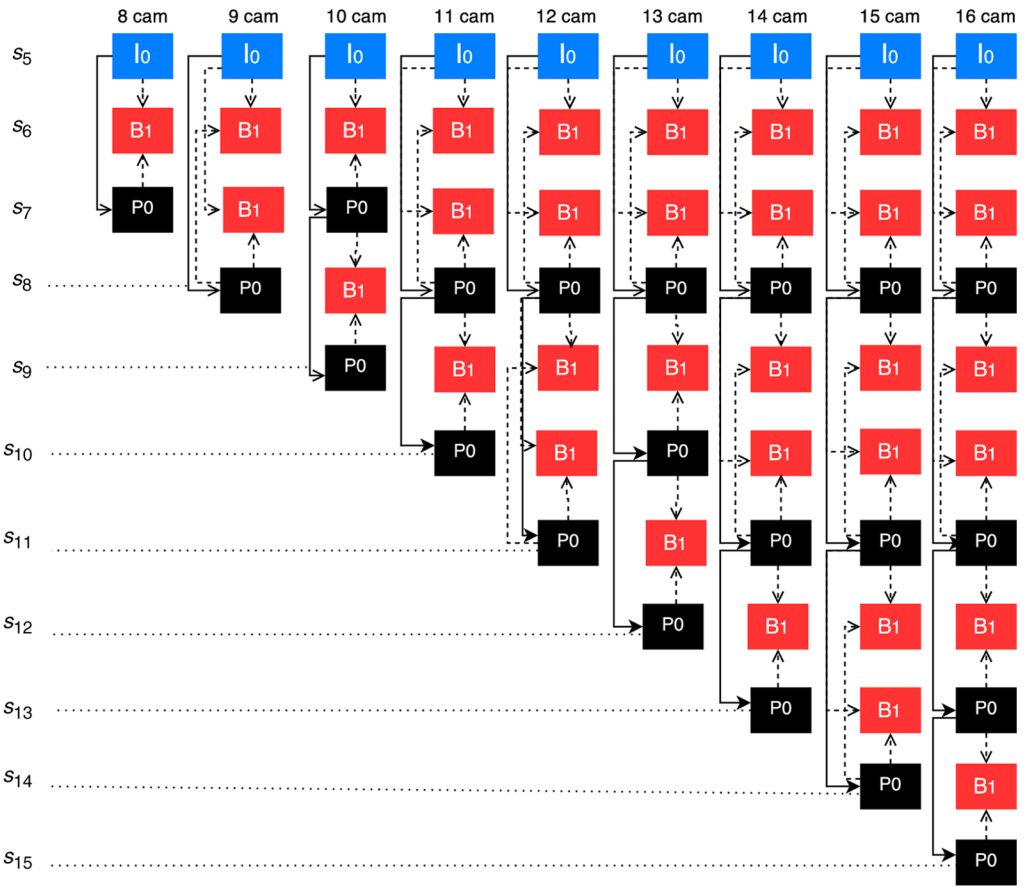

Fig. 8 Extended PBI structure.

The first choice, which is I/P, B, P, is obtained when the remainder of the division of the view number by 3 is always equal to 2, such as 8, 11, and 14. Akin to the second choice, I/P, B, P, B, P, this order avoids the use of successive P-views. It would appear that the sequence order I/P, B, B, P is the best possible choice as it allows the use of successive B-views and avoids the utilization of successive P-views. Additionally, it gives better results for both bit rate gain and random access ability. This order is selected when the remainder of the division of the view number by 3 is always equal to zero, like in the case of 9, 12, and 15 views. The three choices presented above are usually applied subsequently to the following order of views: I, B, B or P, B, B.

As the structure varies according to the number of views, the maximum number, $N_{\max}$, required to access a given picture varies accordingly. With respect to all the possible variations, the computation of $N_{\max}$ is given by

$$N_{\max} = 3 \times H_{\max} + 2 \times \left[\frac{(Nbr_{\text{view}} - 5) + \alpha}{3}\right], \tag{8}$$

where

$$\alpha = \begin{cases} 0 & \text{if } (Nbr_{\text{view}} \,\text{MOD}\, 3) = 2 \\ -1 & \text{if } (Nbr_{\text{view}} \,\text{MOD}\, 3) = 0 \\ 1 & \text{if } (Nbr_{\text{view}} \,\text{MOD}\, 3) = 1 \end{cases}$$

Nbr_{view} and $H_{\max}$ denote, respectively, the number of views and the highest level of the B pictures, usually set to four, in a hierarchical B picture coding structure. The application of the PBI prediction structure to a different number of views allows a fast view random access, especially when several pairs of successive B-views are introduced in the scheme. It also gives a similar video quality, measured in PSNR, compared to IBP, IPP, and Ref. 12.

To further show the effectiveness of the proposed PBI structure in terms of random access ability, the equations calculating the Nbr_{img} for all possible cases, regardless of the number of views considered, are developed:

- For the anchor picture:

$$Nbr_{\text{img}} = \beta + \left[\frac{(V_p - 5) + \alpha}{3}\right], \tag{9}$$

$$\text{where} \begin{cases} \beta = 0 \text{ for P anchor frames} \\ \beta = 1 \text{ for B anchor frames} \\ \text{and} \\ \alpha = \begin{cases} 1 & \text{if } (V_p \text{ MOD } 3) = 1 \\ 0 & \text{if } (V_p \text{ MOD } 3) = 2 \\ -1 & \text{if } (V_p \text{ MOD } 3) = 1 \end{cases} \end{cases}$$

In this case, V_p denotes the P-view number, also used to calculate the Nbr_{img} needed for accessing the anchor picture of the B-view. As an example, for the case of 12 views, to calculate the Nbr_{img} of the two anchors B pictures S_9 and S_{10}, V_p will be set to 12 and β to 1. Thus, Nbr_{img} is equal to 3 for both S_9 and S_{10}.

- For the nonanchor picture:

$$Nbr_{\text{img}} = \beta \times \text{Hierarchy} + 2 \times \left[\frac{(V_p - 5) + \alpha}{3}\right], \tag{10}$$

$$\text{where} \begin{cases} \beta = 1 \text{ for P nonanchor frames} \\ \beta = 3 \text{ for B nonanchor frames} \\ \text{and} \\ \alpha = \begin{cases} 1 & \text{if } (V_p \text{ MOD } 3) = 1 \text{ ,} \\ 0 & \text{if } (V_p \text{ MOD } 3) = 2 \\ -1 & \text{if } (V_p \text{ MOD } 3) = 0 \end{cases} \end{cases}$$

where Hierarchy is the hierarchical level of the picture, which can be set to 2, 3, or 4. For the nonanchor frames, V_p is used in the same way as in Eq. (9).

4 Evaluation of the Random Access Ability

4.1 *Global Random Access Evaluation*

In this part, we propose a new method to evaluate the random access ability, which allows a deeper look into the considered schemes. The evaluation, based on the calculation of the average cost for accessing each existing image, ensures a thorough estimation of each structure with respect to the others.

The evaluation process is composed of three phases. First, a global evaluation of the anchor pictures of each studied structure is carried out. Second, the nonanchor pictures with their different hierarchical levels are evaluated. Then an evaluation covering the entire structure, which includes anchor and nonanchor pictures of the studied structures, is performed.

The speed of access, G_{RA}, to the anchor pictures is estimated by measuring the average random access cost of the anchor pictures. It is equal to the sum of the random access cost of the frames divided by the number of views and is given by

$$G_{\text{RA}} = \frac{\sum_{i=1}^{V_n} Nbr_{\text{img}}(i)}{V_n}, \tag{11}$$

where Nbr_{img} is the number of encoded pictures to access an anchor picture (see Table 3) and V_n is the number of views in the structure.

Applying Eq. (11), we get the following G_{RA} values for the three considered schemes: $G_{\text{RA}} = 2.37$ for IBP, $G_{\text{RA}} = 1.62$ for Ref. 12, and $G_{\text{RA}} = 1.25$ for PBI.

The significant gain in G_{RA} is determined by

$$\Delta G_{\text{RA}} = \frac{G_{\text{RA}}(\text{compared}) - G_{\text{RA}}(\text{PBI})}{G_{\text{RA}}(\text{compared})} \times 100\%, \tag{12}$$

where $G_{\text{RA}}(\text{compared})$ takes either the value of IBP or that of Ref. 12. As illustrated in Fig. 9, the proposed PBI structure achieves gains of 47.25 and 22.83% compared to IBP and Ref. 12 structures, respectively.

The second random access ability evaluation consists of measuring the average cost of encoding the nonanchor pictures of the structure given by

$$G_{\text{RN}} = \frac{\sum_{i=1}^{Vn} \sum_{t=1}^{\text{GOP(size)}-1} [Nbr_{\text{img}}(i,t)]}{\text{GGOP(size)} - V_n}. \tag{13}$$

In other words, G_{RN}, representing the global random access speed for the nonanchor pictures, is defined by the sum of $Nbr_{\text{img}}(i,t)$ divided by the number of the crossed pictures,

Table 3 Nbr_{img} to access anchor pictures following the view order.

	Nbr_{img} 0	Nbr_{img} 1	Nbr_{img} 2	Nbr_{img} 3	Nbr_{img} 4	Nbr_{img} 5	Nbr_{img} 6	Nbr_{img} 7
IBP	0	2	1	3	2	4	3	4
Ref. 12	1	2	2	0	2	1	3	2
PBI	1	2	0	2	2	0	2	1

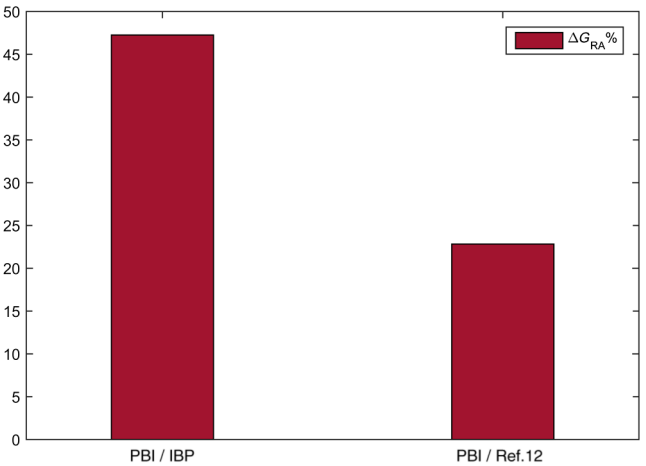

Fig. 9 Gains in G_{RA} compared to IBP and Ref. 12 structures.

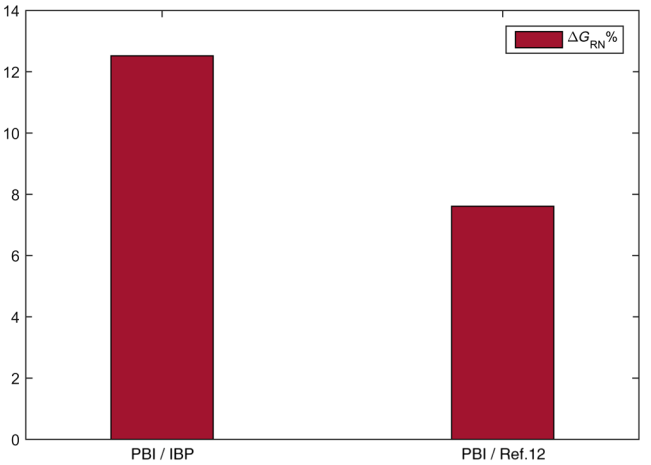

Fig. 10 Gains in G_{RN} compared to IBP and Ref. 12 structures.

which is equal to GGOP(size) $- V_n$. $Nbr_{\text{img}}(i,t)$ is the encoded picture number to access a nonanchor picture at view level position i and instant position t. V_n is the number of used views, and GGOP(size) is equal to V_n multiplied by the size of the GOP.

To demonstrate the efficiency of our proposed PBI structure in terms of random access cost, regardless of the hierarchical level of the encoding pictures, Eq. (13) is applied to the three considered structures that are composed of eight successive views and GOP size equal to 8. The obtained G_{RN} values for the structures are as follows: G_{RN} for IBP $= 10.41$, G_{RN} for Ref. 12 $=9.85$, and G_{RN} for PBI $= 9.10$. By adapting Eq. (12) for G_{RN}, and as illustrated in Fig. 10, our proposed scheme improves the random access by $\sim$12.52 and 7.61% compared to the IBP and Ref. 12 structures, respectively.

The global random access evaluation, including both anchor and nonanchor pictures, is given by

$$G_R = \frac{\sum_{i=1}^{V_n} \sum_{t=1}^{\text{GOP(size)}} [Nbr_{\text{img}}(i,t)]}{\text{GGOP(size)}}. \tag{14}$$

The value of the parameter G_R allows a full assessment of the considered structure, taking into account the random access cost of each existing picture in the MVC structure, regardless of its type and hierarchical level. Moreover, it may be considered as the most appropriate evaluation of the random access speed for a given MVC structure, since it reflects the average value of the random access cost. Therefore, by adapting Eqs. (14) and (12) for G_R, it can be observed in Fig. 11 that the proposed PBI scheme provides significant G_R gains that amount to $\sim$13.5 and 8% compared to the IBP and Ref. 12 structures, respectively.

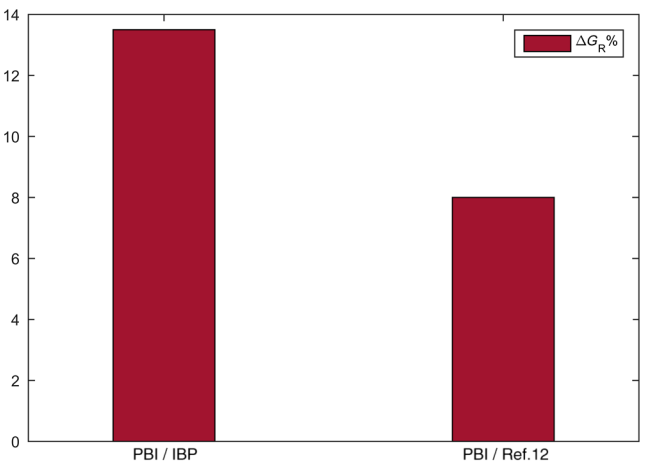

Fig. 11 Global random access gain according to IBP and Ref. 12.

4.2 *Evaluation of View Random Access Using $N_{\max}$*

The maximum number of reference frames $N_{\max}$ representing the random access ability depends on several parameters, such as the number of used views, the number of reference views (I-view) with their positions in the structure, the picture hierarchical level within the GOP, and the GOP size. The chosen positions of the two reference views result in a significant reduction of $N_{\max}$. The relative gain in $N_{\max}$ with respect to the default IBP structure is given by

$$\Delta N_{\max} = \frac{N_{\max}(\text{IBP}) - N_{\max}(\text{compared})}{N_{\max}(\text{IBP})} \times 100\%, \tag{15}$$

where $N_{\max}$ (compared) represents the $N_{\max}$ for either PBI or Ref. 12 structure. Note that the proposed PBI structure usually provides good results regardless of the GOP size. Hence, for the sake of simplification, and without loss of generality, the GOP size is set to 8 for all the considered structures. Since the basic number of views in multiview video coding is 8, a comparison, in terms of the gain $\Delta N_{\max}$, is carried out with a number of basic$_{\text{viewnumber}} * 2 + 1 = 17$ in order to show the periodicity of gain in the proposed PBI structure.

The proposed scheme maintains good random access ability despite the increasing number of views. Note that the best results are achieved when more successive B-views are utilized. This is ensured by the second choice mode presented in Sec. 3.2, namely, I/P, B, B, P, where the entire B-views, after the first base view (I), are successive B-views. In this case, the largest obtained gain $\Delta N_{\max}$ slightly exceeds 30% for a scheme of 15 views, where $N_{\max}$ of the proposed PBI scheme is equal to 18 and that of the IBP structure is equal to 26.

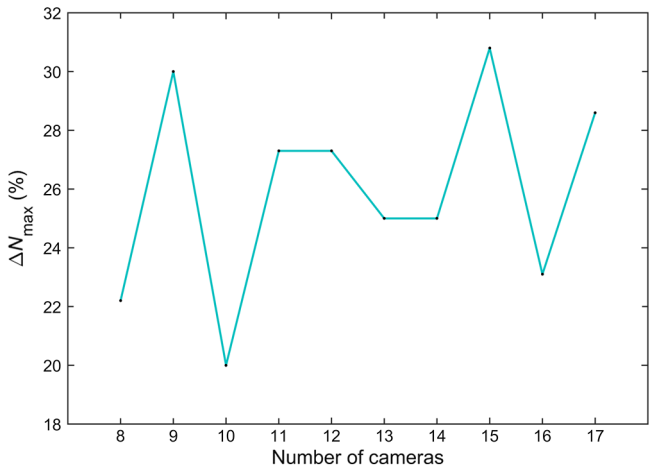

Fig. 12 Random access gain of the PBI structure over the IBP structure.

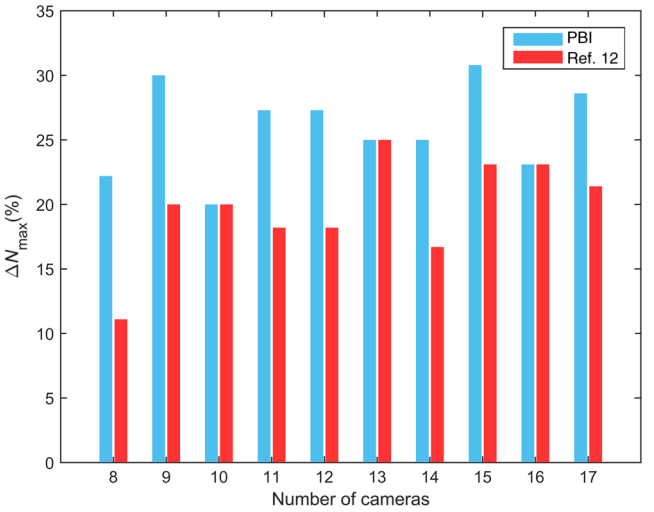

Fig. 13 Random access gain with respect to the IBP structure: comparison between the proposed PBI scheme and Ref. 12 structure.

Figure 12 shows the enhancement in random access ability achieved by the proposed PBI structure compared to the IBP structure, where the minimum gain ΔN_{max} is ~20% and is obtained for the case of 10 views. Figure 13 illustrates a comparison, in terms of the random access gain, between the proposed and Ref. 12 schemes. It is clear that the proposed scheme is more efficient, with an average gain of ~11%, except for the cases of 10, 13, and 16 views, where the two schemes show the same performances. However, using the third choice of the views mode combination, presented in Sec. 3.2, I/P, B, P, B, P, both PBI and Ref. 12 structures provide the same performances. The obtained values of N_{max} for the three considered structures as well as the relative gain ΔN_{max} of (PBI/IBP), (PBI/Ref. 12), and (Ref. 12/IBP) are reported in Table 4. The results demonstrate that the proposed PBI structure significantly reduces the maximum number N_{max} of the reference images needed for decoding a given image, which in turn leads to an improved random access ability of the proposed multiview video coding structure.

5 Experimental Results

Experimental results obtained with the proposed coding scheme are provided and discussed in this section. The compression efficiency of the proposed PBI structure is measured in terms of bit rate (Kbit/s), while the video quality is evaluated through the peak signal-to-noise ratio (PSNR), which is widely used as an objective measure and is given by

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right). \tag{16}$$

MSE represents the mean square error between the compressed and the original video signal. Furthermore, the multiview video compression efficiency is evaluated in terms of the PSNR (dB) of the luminance signal versus the bit rate.

5.1 *Test Conditions*

In order to get comparable results of the proposed scheme against the previously reported structures, common initial conditions and specific test data are required. The used multiview video sequences vary and depend on different parameters, such as the frame rate or frames per second, image resolutions, number of cameras, camera arrangements, and distance between cameras. Table 5 summarizes the used test video sequences and their parameters. This initial configuration is included in the bit stream, which will be sent to the video decoder side.

To provide a fair comparison, the same encoding configuration is employed for both the proposed PBI scheme and the other studied structures (simulcast, IPP, IBP, and Ref. 12). The main encoding parameters are reported in Table 6.

The symbol mode specifies the used entropy coding mode, which is context-adaptive binary arithmetic coding (CABAC), which usually enhances the coding efficiency. Quantization parameter (QP) is the most essential parameter to control the bit rate of a bit stream and the video quality.

The GOP size is directly related to the video frame rate. As reported in Table 6, the GOP lengths of 12 and 15 are assigned for frame rates equal to 25 Hz (Vassar Ballroom and Exit) and 30 Hz (Race 1 and Rena), respectively. The fast motion search algorithm, with a search range of 64, is employed since it significantly reduces the encoding time.

5.2 *Results and Discussion*

In this section, the effectiveness of our proposed PBI scheme is evaluated from a compression efficiency perspective. The compression efficiency is expressed in terms of PSNR (dB) versus bit rate (Kbps). From the results obtained in Fig. 14, it can be shown that the two reference views S_2 and

Table 4 N_{max} and ΔN_{max} gain of the proposed PBI structure in comparison to IBP and Ref. 12 structures.

View number	8	9	10	11	12	13	14	15	16	17
N_{maxIBP}	18	20	20	22	22	24	24	26	26	28
N_{maxPBI}	14	14	16	16	16	18	18	18	20	20
$N_{max\,Ref.\,12}$	16	16	16	18	18	18	20	20	20	22
$\Delta N_{max(PBI/IBP)(\%)}$	22.2	30.0	20.0	27.3	27.3	25.0	25.0	30.8	23.1	28.6
$\Delta N_{max(PBI/Ref.\,12)(\%)}$	12.5	12.5	0.0	11.1	11.1	0.0	10.0	10.0	0.0	9.1
$\Delta N_{max(Ref.\,12/IBP)(\%)}$	11.1	20.0	20.0	18.2	18.2	25.0	16.7	23.1	23.1	21.4

Table 5 Used multiview video sequences for analysis and evaluation.

Database	Sequences	Frame rate	Image resolution	Camera parameters
KDDI	Race 1	30 fps	640 × 480	8 cameras, 20 cm spacing, 1-D parallel
Tanimoto Lab	Rena	30 fps	640 × 480	100 cameras, 5 cm spacing, 1-D parallel
	Vassar	25 fps	640 × 480	8 cameras, 20 cm spacing, 1-D parallel
MERL	Ballroom	25 fps	640 × 480	8 cameras, 20 cm spacing, 1-D parallel
	Exit	25 fps	640 × 480	8 cameras, 20 cm spacing, 1-D parallel

Table 6 Encoding configuration.

Parameter	Setting
Symbol mode	CABAC
QP	22, 27, 32, 37, 40
GOP size	12,15
Search mode	Fast search
Search range	64

S_5 (I-view), which exploit only the temporal compensation, take relatively larger bit rate values while providing a good video quality. The P-views (S_0 and S_7), using the temporal compensation in addition to the disparity compensation from only one view, have lower bit rate values in comparison to the I-views. The rest of the views, S_1, S_3, S_4, and S_6 (B-views), which make use of the temporal compensation besides the disparity compensation from two views, provide the lowest bit rate values while maintaining acceptable video quality.

Note that all the evaluation tests were carried out using the five QP values mentioned in Table 6.

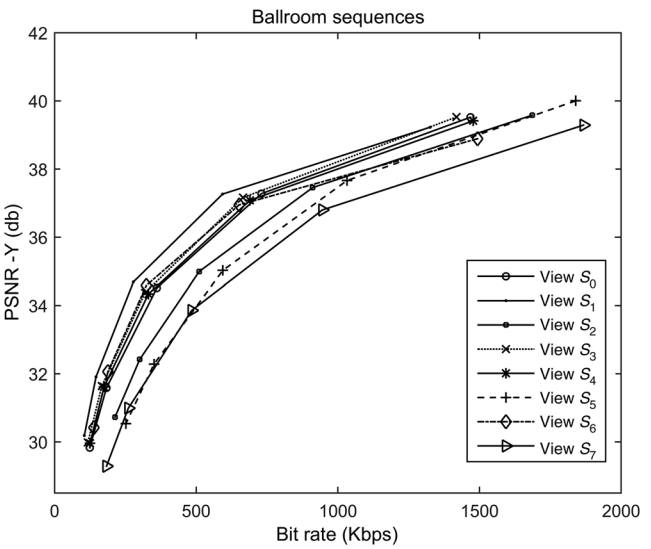

Fig. 14 Distribution of the views of the proposed PBI scheme.

The QP controls the compression efficiency of the video. The lower the value of the quantization parameter, the higher the values of bit rate and video quality.

Table 7 illustrates an example of the quantization parameter effects on the compression efficiency of the proposed structure PBI.

As depicted in Fig. 15, the compression efficiency in terms of PSNR (dB) versus bit rate (Kbps) of the proposed PBI structure is evaluated and compared against that of the four prediction structures (simulcast, IBP, IPP, and Ref. 12), using the five test sequences mentioned in Table 5.

From the results of Fig. 15, it can be easily observed that our PBI proposed structure outperforms the simulcast structure and provides nearly the same performances as the three other schemes in terms of bit rate saving and PSNR gain, which are, respectively, given by

$$\Delta_{\text{bitrate}} = \frac{\text{bitrate}_{\text{compared}} - \text{bitrate}_{\text{proposed}}}{\text{bitrate}_{\text{compared}}} \times 100\%, \tag{17}$$

$$\Delta\text{PSNR}_Y = \frac{\text{PSNR}_Y(\text{proposed}) - \text{PSNR}_Y(\text{compared})}{\text{PSNR}_Y(\text{compared})} \times 100\%, \tag{18}$$

where $\text{bitrate}_{\text{compared}}$ and $\text{PSNR}_{\text{compared}}$ take the values of the considered structures to be compared against our proposed scheme.

For the majority of the tested sequences, the IPP structure produces a rather better bit rate saving despite its high complexity and poor random access performance. The video quality is practically similar in all the considered structures, including the proposed structure.

Generally, the performances of a given video encoding structure should be demonstrated at an appropriate quantization parameter value that guarantees a high video quality offering the viewer a clear and comfortable watching experience. Thus, the results obtained for QP = 22, which ensures the best video quality among the five QP values

Table 7 Quantization parameters effects (Exit sequences).

QP	$Q = 22$	$Q = 27$	$Q = 32$	$Q = 37$	$Q = 40$
Bit rate (Kbps)	1278.602	533.489	266.182	152.5808	109.564
PSNR (db)	40.378	38.778	36.847	34.569	32.949

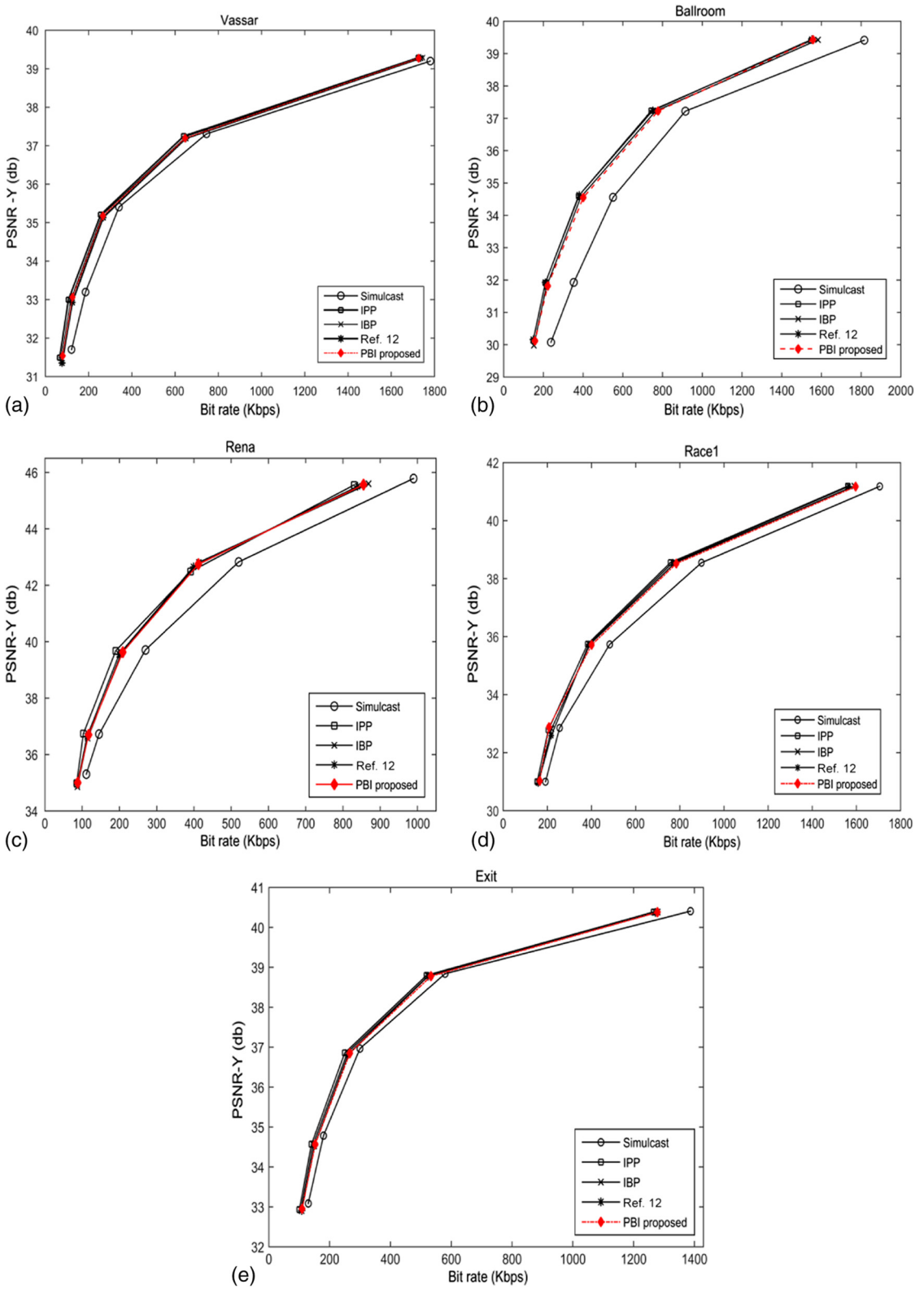

Fig. 15 (a) through (e) Compression efficiency comparison through different test video sequences.

used, are highlighted and reported in Table 8. Note that these results only cover the comparison of the proposed structure against IBP and Ref. 12 structures, because of the low bit rate saving and the poor random access performance achieved by simulcast and IPP schemes, respectively.

From the results of Table 8, our proposed structure shows practically similar PSNR values compared to the two other structures. In addition, the proposed PBI scheme provides better results with respect to IBP structure, with an average bit rate saving of ~1.3%. However, our proposed scheme seems to be

Table 8 Compression efficiency evaluation of the proposed PBI structure.

		IBP		Ref. 12	
	QP	ΔPSNR %	Δ bit rate %	ΔPSNR %	Δ bit rate %
Ballroom	22	0.013	1.59	−0.02	−0.35
Race 1	22	−0.043	−0.41	−0.016	0.0018
Vassar	22	−0.0341	1.0426	0.0108	0.5156
Exit	22	0.13	0.17	0.007	−0.53
Rena 8_{views}	22	−0.07	1.558	0.131	−1.9
Rena 9_{views}	22	−0.015	2.94	0.116	−1.72
Rena 10_{views}	22	−0.061	0.557	0.087	−0.189
Rena 11_{views}	22	−0.029	0.676	0.17	−3.177
Rena 12_{views}	22	0.32	2.86	0.131	−3.63
Rena 13_{views}	22	−0.002	1.60	0.17	−3.74
Rena 14_{views}	22	−0.021	3.18	0.131	−3.15
Rena 15_{views}	22	−0.012	2.11	0.109	−1.61
Rena 16_{views}	22	0.0023	−0.78	0.121	−2.56

slightly less efficient than the Ref. 12 structure with an average bit rate loss of ∼1.7%, which can be considered to be insignificant with regard to improvement in the random access performance provided by our approach.

The main advantage of choosing the proposed PBI structure is its significant improvement of the random access performance over the reported structures (IBP and Ref. 12). This aspect has been enhanced on each of the key and non-key pictures in the MVC structure. Moreover, the proposed structure has an inter-view coding scheme with less complexity, which means a lower time for coding or decoding any picture within the structure. In addition, the PBI structure satisfies other main requirements, such as the compression efficiency and the video quality. These requirements have been illustrated in Fig. 15. However, in some cases, the Ref. 12 structure provides better bit rate saving than the proposed PBI, as is reported in Table 8. Overall, the video quality, which is measured in PSNR, appears to be quite similar across the reported structures in Table 8.

The proposed structure would be the better choice for a more smooth view switching and interactivity feature.

6 Conclusion

In this paper, a PBI inter-view prediction scheme, incurring better random access performance for the multiview video encoder, is presented. The proposed PBI structure, which uses B-views to provide a good bit rate saving, is conceptually based on using two base views (S_2, S_5) per GGOP with positions allowing a direct interview prediction of the rest of the views. Depending on the number of views and taking into consideration three possible choices following the second base view (S_5), a generalization of the PBI scheme is performed, leading to an improvement in the view random access ability. The proposed coding approach is evaluated and compared against two relevant works, namely IBP and Ref. 12 coding structures, in terms of random access capability evaluated using the new proposed metric G_R and the gain ΔN_{max} in maximum number of pictures to be decoded. Despite the insignificant loss in bit rate with respect to the Ref. 12 structure, the proposed scheme performs better and achieves significant coding gains in G_R and ΔN_{max} that amount to ∼8 and 11%, respectively. Additionally, compared to the IBP coding structure, the obtained results show that the proposed scheme attains average performance gains in G_R and ΔN_{max} of ∼13.5 and 20%, respectively, with an average bit rate saving of ∼1.3%.

References

1. A. Smolic et al., "3D video and free viewpoint video—technologies, applications and MPEG standards," presented at *IEEE Int. Conf. Multimedia and Expo*, pp. 2161–2164, IEEE, Toronto, Ontario, Canada (2006).
2. ISO/IEC 14496-10 and ITU-T Rec. H.264, Advanced video coding (2003).
3. P. Merkle et al., "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.* **17**(11), 1461–1473 (2007).
4. Y. Chen, P. Pandit, and S. Yea, "WD 4 reference software for MVC," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207 (2009).
5. "Requirements on multi-view video coding v.4," ISO/IEC JTC1/SC29/WG11, Doc. N7282, Poznan, Poland (2005).
6. N. Cheung and A. Ortega, "Distributed source coding application to low-delay freeviewpoint switching in multi-view video compression," in *Picture Coding Symp. 2007*, Portugal (2007).
7. Y. Liu et al., "Low-delay view random access for multi-view video coding," in *IEEE Int. Symp. on Circuits and Systems 2007*, pp. 997–1000 (2007).
8. Y. Zhang et al., "Adaptive multiview video coding scheme based on spatio-temporal correlation analyses," *ETRI J.* **31**(2), 151–161 (2009).
9. Y. Yang et al., "Coding order decision of B frames for rate-distortion performance improvement in single-view video and multiview video coding," *IEEE Trans. Image Process.* **19**(8), 2029–2041 (2010).
10. L. Ma and F. Pan, "Efficient compression of multi-view video using hierarchical B pictures," in *Int. Conf. on Multimedia and Ubiquitous Engineering*, pp. 118–121 (2008).
11. U. Fecker and A. Kaup, "Complexity evaluation of random access to coder multi-view video data," ISO/IEC JTC1/SC29/ WG11, N8019 (2006).
12. A. Bekhouch and N. Doghmane, "Multiview video coding with an improved prediction structure for faster random access," *J. Electron. Imaging* **22**(4), 043010 (2013).
13. X. Lv, L. Ma, and J. Guo, "Multiview video coding scheme based upon enhanced random access capacity," *Int. J. Comput. Sci. Issues* **10**(1), 285 (2013).

Seif Allah Elmesloul Nasri received his BS and MS degrees from the University of Badji Mokhtar, Annaba, Algeria, in 2009 and 2011, respectively, both in communications and digital processing. Currently, he is a PhD student in multimedia and digital communications and a member of LASA laboratory. His main research interests include multiview video coding and image/video processing.

Khaled Khelil received his master's and PhD degrees from Ferhat Abbes University, Setif, Algeria, in 1996 and 2006, respectively. Currently, he is a professor at the University of Souk Ahras and is a scientist director of the Power and Renewable Energies Laboratory. His current research interests include multimedia signal processing, multiple description coding, wavelets, and renewable energies.

Noureddine Doghmane received his engineering degree in electronics from Annaba University, Algeria, in 1984 and his PhD from Lyon University Claude Bernard, France, in 1988. He is now a professor in the Engineering Sciences Faculty, Annaba University Badji Mokhtar, Algeria. He is also director of the research laboratory Automatic and Signal Processing of Annaba (LASA). His main research interests include signal processing, digital communications, and image/video coding.