Improving View Random Access via Increasing Hierarchical Levels for Multi-view Video Coding

Amara Bekhouch, Imed Bouchrika and Noureddine Doghmane

Abstract — *Multi-view video has attracted considerable interest due to its wide use in a growing market including 3D television, free viewpoint video and intelligent surveillance. Besides the efficiency of compression methods, a vital requirement for multi-view video coding procedures is the view random access which is described as the capability to navigate quickly to any arbitrary view at any given time. In this paper, a new prediction structure based on the increase of the hierarchical level of B-views is proposed. This approach involves reducing the number of images needed for the prediction of B pictures for specific B-views. To show the effectiveness of the proposed prediction structure, a new metric for evaluating the view random access is described. In contrast to the metric proposed by the JVT group which is limited to consider only the image having the maximum number of frames needed for decoding, the key basis of the proposed metric is to consider evaluating all the images contained within a Group of Group Of Pictures. Experimental results have shown that compared with the IBP prediction structure of the reference model JMVM, the proposed algorithm improves the view random access by up to 33.5% with significant improvement in terms of bit rate.*[1]

Index Terms — View random access, inter-view prediction, compression efficiency. Multi-view video

I. Introduction

Multi-view video (MVV) has received recently considerable momentum from the research community due to the unprecedented advancement in 3D display technologies. The multi-view video can be defined as a set of cameras capturing synchronously the same scene from several angles giving the users an enhanced multimedia experience. Research in this area is fueled by the numerous applications of multi-view videos in a growing market including 3D television [1],[2], free viewpoint video (FVV) [3],[4] and intelligent surveillance. The free viewpoint video gives the ability to browse across multiple views of the same scene giving a realistic impression of the viewed scene. Because of the large amount of data being transmitted from the different viewpoints in addition to the concerns of limited storage and network bandwidth, it becomes critically important to employ an efficient compression algorithm to encode the multiple video streams that would assist in reducing the redundancy and speed up the transmission process. The Multi-view Video Coding (MVC) is created to meet the rapid development of visual communication services for multi-view video. The MVC has been standardized as an extension of the Advanced Video Coding (H.264/AVC) [5]. This standard is the result of the work by the Joint Video Team (JVT) standardization body. The JVT group was created jointly between ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group. A major feature which must be observed in the MVC is that all views of the MVV should use the same encoder such as H.264/AVC or even HEVC.

A rudimentary requirement for compression-based coding procedures of multi-view video is the view random access which is defined as the capability to quickly navigate to any view at any given time. Random access is needed either at the temporal or view level. For the temporal random access, pictures can be decoded via the intra-coded frames from the same viewpoint whilst view random access ensures that any frame from any selected arbitrary view can be decoded from one or more frames that are intra-coded. The spatial and temporal inter-view coding enforces certain dependencies between views during the encoding process which therefore deteriorates the random-access potency. Low-delay random access is necessary to give users the option to view an arbitrary view with an acceptable response time. Nevertheless, there are other requirements having similar importance in addition to random access such as error robustness and low delay.

Due to the dearth of approaches dealing with the requirement of faster random access for multi-view video, a new prediction structure is proposed to facilitate better random access for different views. The proposed approach named as PS-WOPB for *P*rediction *S*tructure *W*ith*O*ut inter-view *P*rediction for non-anchor *B* pictures, is based on the principle of increasing the hierarchical level of *B*-views. The PS-WOPB structure is devised with the purpose to attain better tradeoff between view random access and bit rate. To show the potency of the proposed work, extensive experiments are performed to assess the efficiency for view random access using different datasets. Another major contribution is proposing an evaluation criterion for view random access. For comparative reasons, this metric is used to assess the proposed structure as well as the one implemented for the Joint Multi-view Video Model (JMVM) [6]. Experimental results have shown the efficacy of the proposed method being applied to multi-view video coding. As a result,

Amara Bekhouch and Imed Bouchrika are with the Faculty of Science and Technology, University of Souk Ahras, Souk Ahras, 41000 Algeria (e-mail: a.bekhouch@univ-soukahras.dz , imed@imed.ws).

Noureddine Doghmane is with the Electrical Engineering Department, University of Annaba, Algeria, (e-mail: ndoghmane@univ-annaba.org).

Contributed Paper
Manuscript received 09/30/16
Current version published 01/13/17
Electronic version published 01/13/17

view random access is considerably improved and the bit rate is increased with the same quality.

II. View Random Access for MVV

The task of capturing the same scene simultaneously using multiple cameras from several viewpoints causes high inter-view redundancy. To effectively exploit the redundancy of data, the disparity estimation and compensation techniques must be utilized. These techniques are based upon the same principles as those used at the temporal level. The major difference is that the disparity estimation uses a higher number of blocks (or Macro-blocks) for the prediction of non-anchor pictures. This has been tied to the use of neighboring views of the current view for the prediction. It is worth mentioning that the anchor images are the first encoded images per Group Of Pictures (GOP). These images do not use the temporal prediction. Hence, only an inter-view or spatial prediction is applied to these images. The non-anchor frames are those contained between two anchor frames per GOP as shown in Fig. 1(a). One approach to provide a good bit rate tradeoff with video quality, is to use a mixed prediction (temporal and inter-view.) as the IBP structure [7]. As most of the structures used by the MVC, the IBP structure is based on the Hierarchical B Pictures (HBP) model [8]. Unlike the Simulcast method [7] where only the temporal prediction is used, the IBP structure uses also the inter-view redundancy leading to significant bit rate saving with stable video quality [7].

The view random access (VRA) is an important requirement for the MVC. The IBP structure has improved the aspect for random access via using the *P* and *B*-views in alternate fashion in addition to the base view which is the one that solely contains the Intra-coded picture *I*. The *B* and *P*-views are successively the views that start with *B* and *P* pictures as shown in Fig. 1(b). This choice is adopted to avoid slower VRA as the case of the IPP structure [7]. This is due to using only *P*-views. For this reason, the IBP structure is chosen as the default prediction structure of the reference JMVM model [6] which was created by the JVT group as a common platform for research on MVC. Several approaches for MVC which are based on the IBP structure are proposed in recent years [9]-[11]. Pei-Jun *et al*. [9] presented a Fast Mode Decision algorithm for MVC which uses a Mode Correlation (MC) to reduce the computational complexity. This method is based on reducing the "*coding mode candidates*". This is completed by the selection of the coding mode from inter-view or the intra–view predictions. The selection is achieved by using the MC between the neighboring views. The approach proposed by Pei-Jun *et al*. [9] reports promising results in terms of reduction for the computational complexity. However, random access suffers from the drawback of being slow due to the adoption of the IBP structure.

Hany *et al.* [10] proposed an algorithm which estimates blindly the Group Of Pictures of Prediction Structure (GOP-PS) combined with the use of stream order coding. Their approach is based on the blind source separation technique. This method relies on fully automated configuration and the adaptive detection of the optimal structure of the GOP. The fully automated configuration of GOP-PS allows an improvement in the compression efficiency. This configuration enhances the quality of the decoded video even when coding or transmission errors occur. The alternative use of *P* and *B*-views as well as the successive *P*-views can provide undesirable results in terms of VRA. This is regardless of the position used for the base view. Yang *et al.* [11] proposed a new structure based on improving the encoding order of *B* pictures in each view of the MVV. The encoding order is improved by using a binary tree. The same technique is then applied to all views of the MVV. This technique offers significant bit rate saving. Nevertheless, the major drawback of this structure is the increased encoding complexity leading to longer duration for random access.

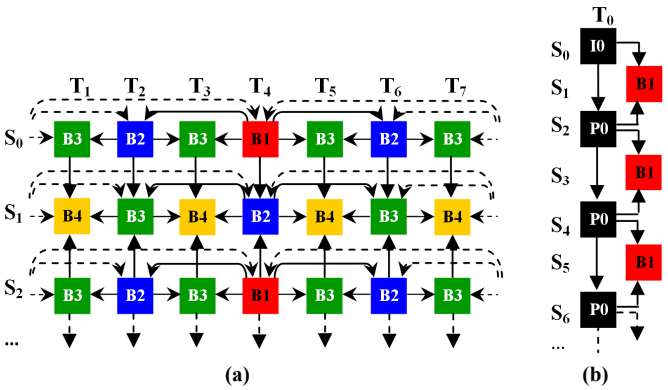

Fig. 1. Anchor and non-anchor frames of the IBP structure, (a) Non-anchor frames using only 3 views, (b) Anchor frames of 7 views.

As any existing video codec, the MVC must comply with a number of requirements [12] which can be classed either mandatory or desirable. Among the most important mandatory requirements is the high compression efficiency as well as the view random access. The compression efficiency is tied to a tradeoff between bit rate and video quality. This efficiency can be achieved by exploiting the inter-view redundancy in addition to the spatial and temporal redundancy. In the MVC, the techniques used for the elimination of spatial and temporal redundancy are the same methods employed in well-known standards such as H.264/AVC and HEVC. At the inter-view level, the disparity estimation and compensation is being performed through the Lagrange parameter [13] in the same way as the temporal case. Nevertheless, there are several differences between the temporal and inter-view level. A major difference is that the non-anchor *B* pictures in the inter-view level use four reference pictures for prediction. This level uses two reference images at the temporal level and two of the inter-view. Unlike the non-anchor *B* pictures, the anchor *B* pictures uses only two reference images at the inter-view level. It is obvious that *B* pictures provide the lowest bit rate compared to *I* and *P* pictures. The number of *B* pictures increases according to the size of the GOP which can be 8, 12,

15 pictures, etc. Fig. 2 and 3 show respectively an example of a GGOP of 3 views composed of 12 and 15 images by GOP. Moreover, increasing the number of *B* pictures by increasing *B*-views has a dual purpose. Firstly, it improves the total bit rate and VRA can be ameliorated in a considerable way. This is mainly due to the minimization of the *P*-views by Group of Group of Pictures (GGOP).

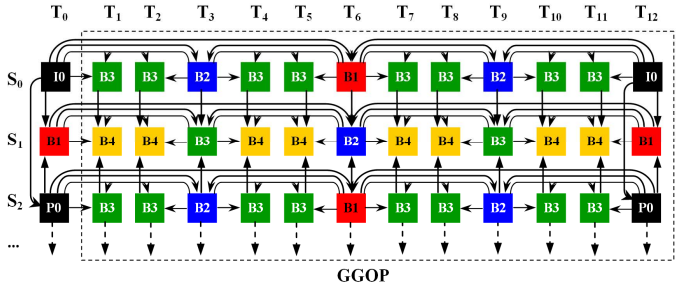

Fig. 2. An example of a GGOP of 3 views using 12 images per GOP.

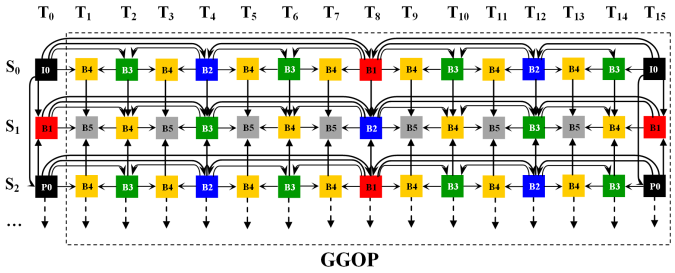

Fig. 3. An example of a GGOP of 3 views using 15 images per GOP.

View random access can be expressed by the number of reference pictures in a GGOP necessary for decoding a given image at time T_n. The number of reference images required each time must be minimized. This number can be calculated for all the *P* and *B* pictures of a GGOP. *I* pictures are spatially encoded. VRA depends on several parameters as the maximum hierarchical level of *B* images. Increasing the hierarchical level may increase the total number of pictures necessary for decoding a given picture. Another important factor is the number of successive *B*-views between two views of type *I/P*. The IBP structure uses a single *B*-view between two *I/P*-views. Another method which is reported to give better results in terms of random access is the structure proposed by Bekhouch *et al* [14],[15]. The method keeps the same maximum hierarchical level of the *B* pictures with the use of pairs of successive *B*-views. This structure is referred as PS-WPSB (*Prediction Structure With Pairs of Successive B-views*). The use of pairs of successive *B*-views without increasing the hierarchical level has led to significant improvement of VRA and bit rate. The improved bit rate is due to the high number of *B*-views with respect to the *P*-views by GGOP.

III. Proposed Approach

A. Initial architecture without improvement of VRA

In this research study, a faster view random access must be observed for all images of each GGOP. However, an arbitrary access to an image *I* in view *V* implies that the value of *N* which is the number of images to be decoded to access the image *I*, is the smallest possible value. This requires that the view *V* that contains the image *I* is as close as possible to the base view. In other words, a direct prediction of the view *V* from the base view without going through the other views, can significantly improve VRA. This property requires a good choice of the position of the base view. In the proposed approach, the position of the base view is selected in such a way that it allows a direct prediction of a maximum number of views. Fig. 4(a) and 4(b) show an example for the organization of the used views within the IBP and proposed structures respectively. In Fig. 4, only the anchor pictures of 8 views are used. The sequence used in the proposed structure as the base view is the *S4* view. This choice allows direct prediction from the base view to all views except two views which are *S1* and *S3* of type *B2* as shown in Fig. 4(b). For example, the view *S1* uses for its prediction the *S2* view as a reference in addition to the base view *S4*. However for the case of the IBP structure, only two views which are *S1* and *S2* enjoy the benefit of direct prediction from the base view. Moreover, the other views must use at least two reference views for the prediction of its images within the IBP structure.

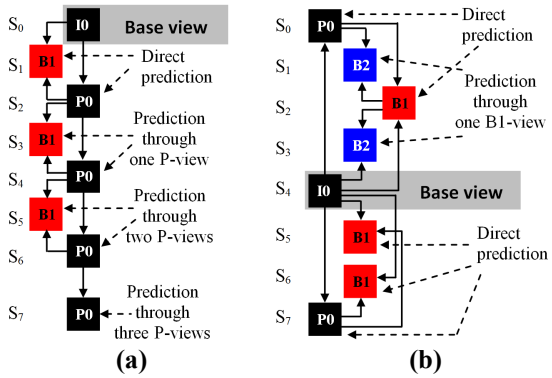

Fig. 4. Organization of the used views: (a) IBP. (b) Proposed structure.

The proposed approach is based on the use of three successive *B*-views between views having different types which must be either *I* or *P*. In Fig. 4, the three successive *B*-views are *S1*, *S2*, and *S3* of type *B2*, *B1*, and *B2* respectively. Unlike the approach PS-WPSB [15], the level of HBP is increased in the proposed method. The maximum hierarchical level of the approach PS-WPSB is the same as that of the IBP structure and is equal to four. The maximum level in the proposed approach is five. This increase is due to the use of two hierarchical levels for anchor pictures instead of a single level as the IBP and PS-WPSB structures. The second level is represented by the B2-views. It is noteworthy that in the IBP and the structure PS-WPSB, all views other than the views *I/P* are of type *B1*. The basic idea is to use maximum *B*-views which can automatically reduce the number of *P*-views in the GGOP leading to significant improvement for bit rate. This has resulted in creating an initial structure named as PS-WPB shown in Fig. 5 which stands for *P*rediction *S*tructure-*W*ithout

*P*rediction of *B* pictures. As the correlation increases between adjacent views, the maximum hierarchical level is increased proportionally allowing the use of *B1*-view as a reference in parallel with *I* and *P*-views. For instance, the view referenced by *S2* in Fig. 5 serves as a reference for two *B2*-views to be encoded which are *S1* and *S3*.

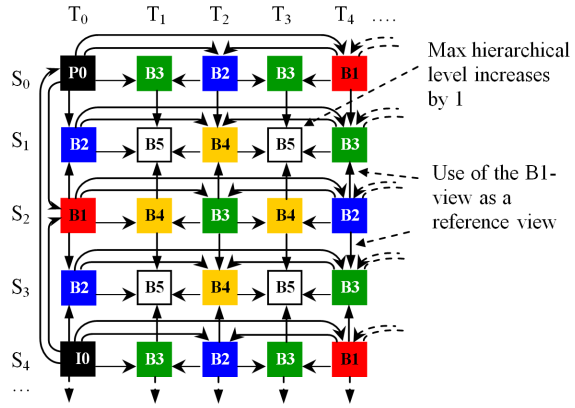

Fig. 5. PS-WPB: Initial proposed structure using five views without removing the inter-view prediction of B2-views.

B. Proposed approach with improved of VRA

The major disadvantage of increasing the level of hierarchical *B* pictures is the slow view random access to different images of GGOP. Specifically, VRA can be slower to images *B* of *B2*-views. This is due to the high number of reference pictures necessary for decoding a given image in a *B2*-view. To overcome this problem, the prediction of non-anchor pictures of *B2*-views is removed from *B1*-views from the initial structure PS-WPB to create the final proposed structure called PS-WOPB (*P*rediction *S*tructure *W*ith*O*ut inter-view *P*rediction for non-anchor *B* pictures) as shown in Fig. 6. Each non-anchor pictures of *B2*-view uses only three reference images instead of four images such that two at the temporal level and a single image at the inter-view level. The inter-view predictions of *B1*-views are kept to take advantage of direct prediction from the base view.

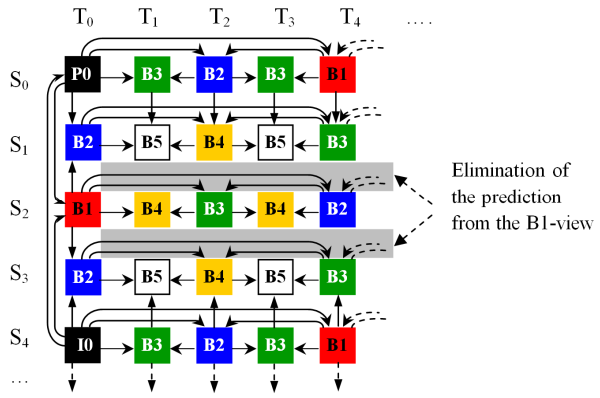

Fig. 6. PS-WOPB: Proposed approach using five views with removing the inter-view prediction of B2-views from B1-views.

A GGOP of eight views allows the use of a single group of three successive *B*-views. In this case, such group must be located before the base view. Indeed, eight views require the use of two successive *B*-views following the base view. In the proposed approach, the structuring and type of views after the base view can be devised depending on the number of used views. The structuring is conditioned upon several constraints that must be respected including, the maximum use of successive *B*-views. The desired number of successive *B*-views is 3. If the remaining views after the base view are not sufficient, two successive *B*-views as the case of eight views have to be used. A single *B*-view can be used as the case of seven or nine views. Fig. 7 shows an example of the layout for more than eight views starting with the base view *S4*. Further, unlike the IBP prediction structure, the case of successive *P*-views is undesirable. The only reason is to avoid view random access being slower [7].

Another important property implemented in the proposed approach is the use of the inter-view prediction of non-anchor pictures of the last *P*-views. The *B* pictures of the last *P*-views are also predicted from the penultimate *P*-view. The other *P*-views use only a temporal prediction. This property should not be used when the number of successive *B*-views between the last two *P*-views is 3. This is because the correlation decreases between very farther views as the case of nine views.

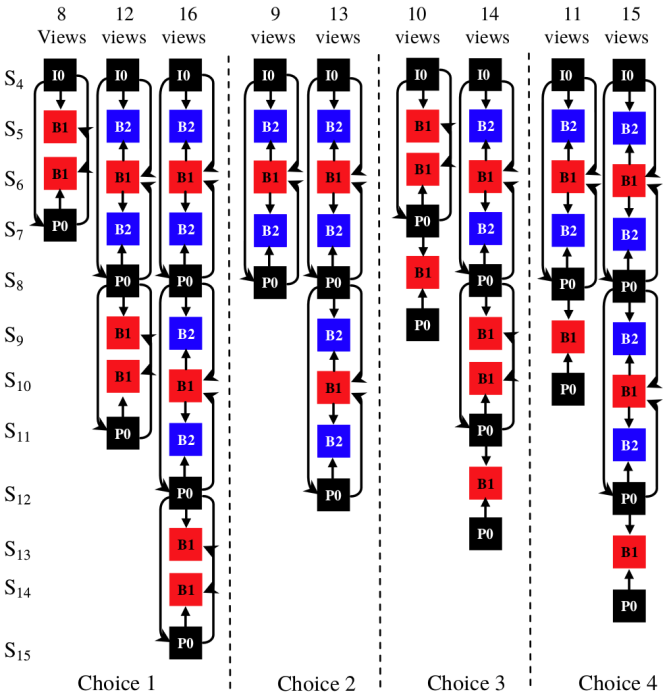

Fig. 7. The organization of the used views by total number of views.

Fig. 7 shows a specific case of using 16 views such that the arrangement or *Layout* of the last views can be deduced via the following set of equations regardless the number of the views such that Nbr_{views} is the total number of used views.

$$Layout = \begin{cases} "I/P, B_1, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 0 \\ "I/P, B_2, B_1, B_2, P" & if\ (Nbr_{views}\ MOD\ 4) = 1 \\ "I/P, B_1, B_1, P, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 2 \\ "I/P, B_2, B_1, B_2, P, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 3 \end{cases} \quad (1)$$

The number of the last views varies according to the total number of used views. Each choice must be always located after a sequence of type "*I/P, B2, B1, B2*" regardless of the number of used views. For example, the case of 14 views, the used choice is "*P, B1, B1, P, B1, P*" that must be preceded by the pattern "*I, B2, B1, B2*".

C. Proposed evaluation Metric for VRA

The second contribution in this study is the proposal of a new model for evaluating of the view random access due to its importance nature for MVC. The proposed evaluation model is based on two metrics that are N_{MAX} and *NBR-img*. The N_{MAX} is used by the IBP structure. This metric represents the maximum reference pictures to be decoded to access a given image S_n at moment T_n. This metric is given for the IBP structure using:

$$N_{MAX} = 3 * Hierarchy_{MAX} + 2 * \lfloor (Nbr_{views} - 1)/2 \rfloor \quad (2)$$

Where $Hierarchy_{MAX}$ is the maximum level of the HBP model. $Hierarchy_{MAX}$ is 4 in the case of IBP structure [7]. The N_{MAX} of the proposed structure PS-WOPB can be obtained in the same way using

$$\begin{aligned} &if\ Nbr_{view} <= 6 \\ &\quad N_{MAX} = 13 \\ &elseif\ (Nbr_{view} = (Nbr_{view} div\ 3 - 1) * 3 + Nbr_{view} div 3) \\ &\quad N_{MAX} = 3 * (Hierarchy_{MAX} - 1) + 2 * \left\lfloor \frac{Nbr_{view} - 2}{4} \right\rfloor - 1 \\ &else \\ &\quad N_{MAX} = 3 * (Hierarchy_{MAX} - 1) + 2 * \left\lfloor \frac{Nbr_{view} - 2}{4} \right\rfloor \end{aligned} \quad (3)$$

The $Hierarchy_{MAX}$ is equal to 5 for the proposed structures PS-WPB and PS-WOPB. This is explained by the increase in the inter-view hierarchical level. As shown in (3), the N_{MAX} of the PS-WOPB structure is conditioned by the different layout cases expressed in (1). Similarly the N_{MAX} proposed for the PS-WPB structure can be obtained by:

$$\begin{aligned} &if\ (Nbr_{view} - 5) < 4 \\ &\quad N_{MAX} = 4 * hierarchy_{MAX} - 1 \\ &else \\ &\quad N_{MAX} = 4 * (Hierarchy_{MAX}) - 1 + 2 * ((Nbr_{view} - 5) div 4 - \alpha) \\ &where \\ &\quad \alpha = \begin{cases} 1 & if\ (Nbr_{view} - 5) \bmod 4 == 0,2,3 \\ 2 & if\ (Nbr_{view} - 5) \bmod 4 == 1 \\ 1 & if\ ((Nbr_{view} - 5) \bmod 4 == 1) and ((Nbr_{view} - 5) div 4 = 1) \end{cases} \end{aligned} \quad (4)$$

Thus, in (2), (3), and (4), the N_{MAX} is given each time for a single image. For example, if the number of views is 8, the N_{MAX} of the three structures IBP, PS-WOPB, and PS-WPB is respectively equal to 18, 14, and 19. For the case of IBP structure the value 18 is given for images of type *B4* in the view *S5*. The values 14 and 19 are the N_{MAX} of pictures of type *B4* and *B5* in views *S2* and *S1* respectively for both structures WOPB-PS and PS-WPB.

```
for i = 1 to size-GOP
    for j = 1 to Num_Ord
        if P_view
            if GGOP(i,j) is Non_anchor frames
                Nbr_img = 1 * Hierarchy + 3
            else
                Nbr_img = 1
            endif
        endif
        if B1-view
            if GGOP(i,j) is Non_anchor frames
                Nbr_img = 3 * Hierarchy + 2
            else
                Nbr_img = 2
            endif
        endif
        if B2-view
            if GGOP(i,j) is Non_anchor frames of PS-WOPB
                Nbr_img = 2 * Hierarchy + 3
            else if GGOP(i,j) is Non_anchor frames of PS-WPB
                    Nbr_img = 4 * Hierarchy - 1
                else
                    Nbr_img = 3
                endif
            endif
        endif
end
```

Fig. 8. Algorithm to compute the metric NBR-img for all images before the base view S4 within the two structures: PS-WOPB & PS-WPB.

The improvement of this approach in terms of VRA emerges clearly when using the *NBR-img* evaluation metric which represents the required number of reference images to be decoded to consult each image in all views of GGOP. Contrary to N_{MAX}, the metric *NBR-img* must be measured for all images of GGOP other than basic view. The VRA is the same for images of basic view regardless of the approach used. The *NBR-img* can be obtained differently depending on the type of the view that contains the image to be evaluated. Therefore, the relevant views are *B*-views and *P*-views. The type of image anchor or non-anchor gives also different results for *NBR-img*. It should be noted that the *NBR-img* for images of views preceding the base view in both PS-WPB and PS-WOPB structures are generated in different ways. This metric is proposed for both PS-WOPB and PS-WPB structures. The algorithm used for generating the *NBR-img* of

images preceding the basic view *S4* in both structures is described in Fig. 8. *NumOrd* represents the number of views before the base view which is 4 for both proposed structures. The *NBR-img* of all images in *B1*-views and *P*-views is generated by the same method for the structures PS-WOPB and PS-WPB. The difference between the two structures is presented only for the non-anchor image of *B2*-views. The *NBR-img* for images of the views located after the base view can be evaluated regardless of the number of views using the algorithm shown in Fig. 9. *NumOrdAft* refers to ordering number of the view *S5* which is located just after the base view *S4*.

for i = 1 to size-GOP
  for j = Num_Ord_aft **to** Nbr_view // 6 is the view S5
    if P-view
      if GGOP(i,j) is Non-anchor frames
        $Nbr_{img} = 1 * Hierarchy + 2 * \left\lfloor \frac{j-2}{4} \right\rfloor$
      else
        $Nbr_{img} = \left\lfloor \frac{j-2}{4} \right\rfloor$
      endif
    endif
    if B1-view
      if GGOP(i,j) is Non-anchor frames
        $Nbr_{img} = 3 * Hierarchy + 2 * \left\lfloor \frac{j-1}{4} \right\rfloor$
      else
        $Nbr_{img} = 1 + \left\lfloor \frac{j-1}{4} \right\rfloor$
      endif
    endif
    if B2-view
      if GGOP(i,j) is Non-anchor frames of PS-WOPB
        $Nbr_{img} = 2 * Hierarchy + 1 + 2 * \left\lfloor \frac{j-1}{4} \right\rfloor$
      else if GGOP(i,j) is Non-anchor frames of PS-WPB
        $Nbr_{img} = 4 * Hierarchy - 3 + 2 * \left\lfloor \frac{j-1}{4} \right\rfloor$
      else
        $Nbr_{img} = 2 + \left\lfloor \frac{j-1}{4} \right\rfloor$
      endif
    endif
  endif
end

Fig. 9. Algorithm for computing *NBR-img* for all images after the base view *S4* within the two prediction structures: PS-WOPB & PS-WPB.

IV. Experimental Results

The performance evaluation of the proposed approach is achieved under the common test conditions described in [16]. The encoding parameters and the properties of some MVV [16] are the same used for testing the IBP and PS-WPSB structures. Only videos with a linear and arc arrangement type of cameras are considered for the evaluation. Firstly, the compression efficiency is assessed via the bit rate and video quality measured by the Peak Signal To Noise Ratio (PSNR). Subsequently, the evaluation of VRA using both metrics N_{MAX} and *NBR-img* is performed.

A. Bit Rate Analysis

To evaluate the proposed approach, various video sequences are taken from multiple publicly available datasets including: Ballroom, Exit, Vassar, Race 1, and Rena [16]. The chosen values of the quantization parameter QP used for evaluation are 22, 27, 32, 37, and 40. The choice of five QP values is made for the reason of ease comparison. Thus, these values are used for the encoding of all the prediction structures. For all presented results in this section, the bit rate saving is obtained by:

$$\Delta_{bit\_rate} = \frac{(bit\_rate_{utmost} - bit\_rate_{smallest})}{(bit\_rate_{utmost})} \times 100\,[\%] \quad (5)$$

Where $bit\_rate_{utmoust}$ is the greatest bit rate of the two compare structures and $bit\_rate_{smallest}$ is the smallest.

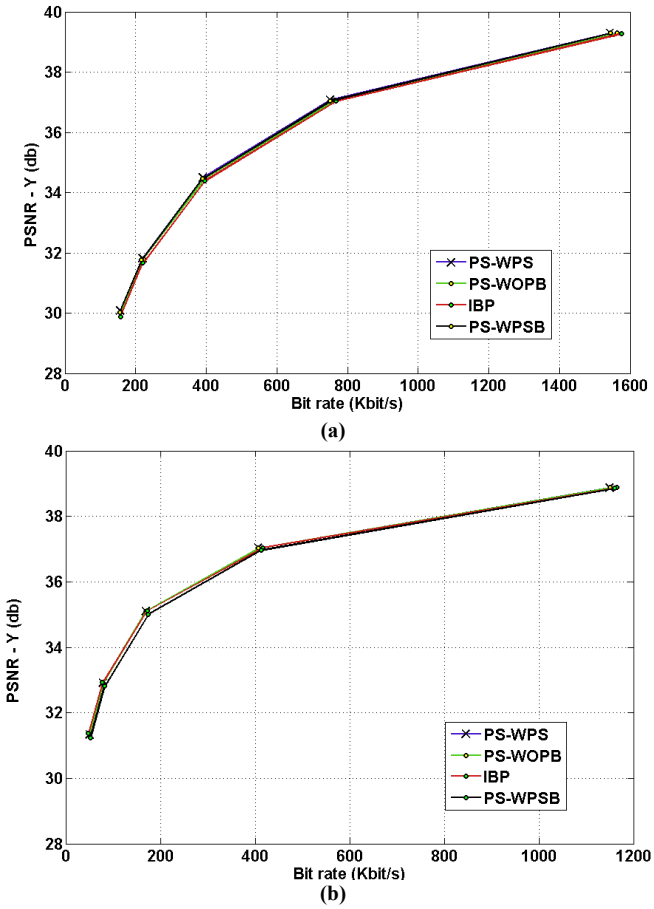

Fig. 10. PSNR results with respect to Bit rate: (a) Ballroom. (b) Vassar.

The structure PS-WPB uses a larger number of reference images that PS-WOPB structure for prediction of *B* images in *B2*-views. This property allows a significant bit rate saving of the PS-WPB structure with respect to the PS-WOPB. Fig. 10 shows the similarity of the compression efficiency in terms of bit-rate between the examined structures (PS-WPB, PS-WOPB, IBP and PS-WPSB) with an observed improvement for the PS-WPB compared to other structures. The results presented in Fig. 10 are obtained using the "Ballroom" and "Vassar" videos. Fig. 11 shows the bit rate gain of the PS-WPB structure with respect to the PS-WOPB. The results presented in Fig. 11 are obtained by using the video of "Rena". The bit rate saving is maximum when the number of used views is 13. This is due to the use of a maximum number of groups of three successive *B*-views. Indeed, the increase in the number of groups of three successive *B*-views, automatically augment the number of *B2*-views. This improves the bit rate of the non-anchor pictures of *B2*-views in the case of PS-WOPB structure. The reason for this property is that each *B* picture in *B2*-views uses three reference images for prediction not four as in the case of PS-WPB.

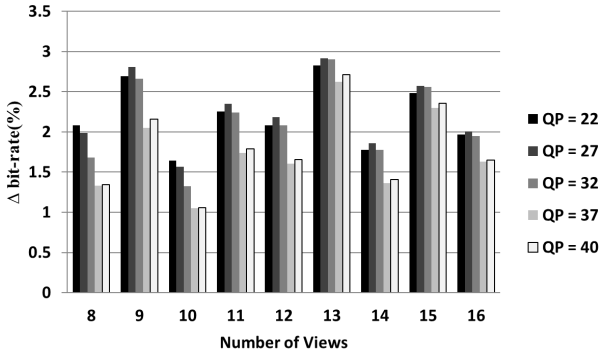

Fig. 11. The bit rate saving for the PS-WPB against PS-WOPB.

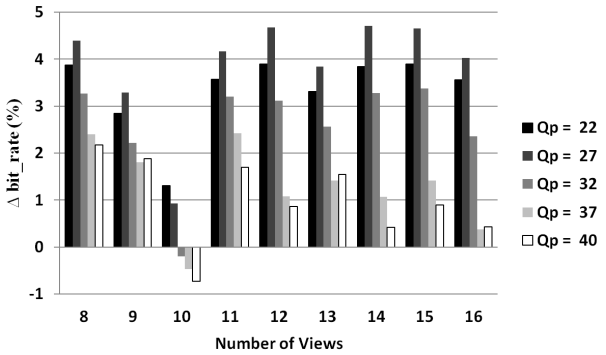

Fig. 12. The bit rate saving for the PS-WOPS against IBP.

The main objective in this work is to provide a tradeoff between bit rate and view random access. This compromise can be fully provided by the PS-WOPB structure. Regardless of the use of three reference pictures for predicting the non-anchor pictures of *B2*-views, the PS-WOPB structure enables a significant bit rate gain compared to the IBP structure. Fig. 12 shows the obtained gain of the PS-WOPB structure in comparison with IBP structure. In fact, several constraints control the variation of bit rate performance such as the number of group of three successive *B*-views used each time. Another specific case is the alternative use of successive *P*-views in the IBP structure according to the number of used views. The bit rate saving is gradually augmented with the increase of the number of used views.

B. Improving VRA using N_{MAX} and NBR-img

The evaluation is carried out using both N_{MAX} and *NBR-img* metrics presented in the previous section. The obtained results in terms of N_{MAX} and *NBR-img* are completely independent of the used video [6]. Instead, these two metrics are dependent on the used prediction structure PS-WPB, PS-WOPB, IBP and PS-WPSB. Comparison with the PS-WPB structure is performed to show the benefits of removing the prediction of *B* image in *B2*-views. The N_{MAX} is the same for both cases where the GOP size is 8 or 12 as the maximum hierarchical level is the same for both sizes. The maximum hierarchical level increases by 1 for the case of GOP of size 15. This property leads to the increase of N_{MAX} proportionally for all studied prediction structures. Fig. 13 shows the N_{MAX} gain for GOP of size 8 or 12 meanwhile the case of GOP of size 15 is depicted in Fig. 14. The N_{MAX} gain is obtained by:

$$\Delta N_{MAX} = \frac{(structure\_N_{MAX} - proposed\_N_{MAX})}{structure\_N_{MAX}} \times 100\,[\%] \qquad (6)$$

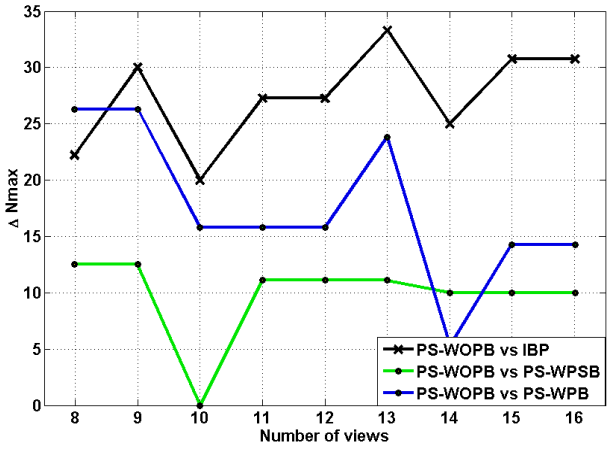

Fig. 13. The obtained N_{MAX} gain using 8 or 12 images per GOP.

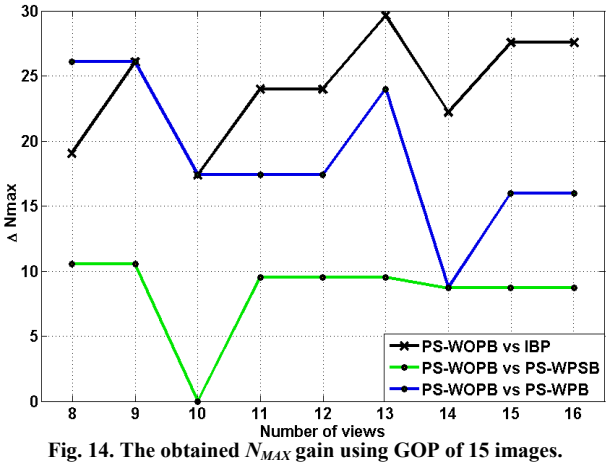

Fig. 14. The obtained N_{MAX} gain using GOP of 15 images.

The gain is observed to be relatively the same for both sizes of GOP whilst the gain value reaches its highest when using 13 cameras. The N_{MAX} gain exceeds 29% using GOP of 15 images and up to 12% compared to the structure PS-WPSB using GOP of 8 and 12 images. For evaluating the time needed to randomly access a certain view based on the N_{MAX} using the different discussed structures, three set of videos are considered for the assessment as shown in Table I. It is observed that the structure PS_WOPB enjoys the benefit of faster access time for view random access. The access time differs from video to video under the same structure cause of the different properties related to each video.

TABLE I
ACCESS TIME FOR VIEW RANDOM ACCESS IN MILLISECONDS

	IBP	PS_WPSB	PS_WPB	PS_WOPB
Ballroom	720	640	760	560
Race1	594	528	627	462
Rena	858	660	693	594

The evaluation using the second metric *NBR-img* is considered more important than the evaluation with N_{MAX}. This is because the *NBR-img* is calculated for all images of GGOP. Meaning, the newly introduced metric is estimated for all images in all views except the base view. The *NBR-img* metric depends on several criteria. For example, the *NBR-img* for images *B* of *B*-views must be larger than those of the *P*-frames. The evaluation using the *NBR-img* gain value is estimated as shown in (7). Table II shows the results for the *NBR-img* of the proposed PS_WOPB structures compared against other structures. The maximum gain is obtained for the case of the proposed structure compared to the IBP method using 16 views whilst it exceeds 20% when using 15 pictures by GOP and reaches almost 24% for GOP of size 8. For the other two structures, considerable improvement is observed with respect to the PS-WPSB.

$$\Delta Nbrimg = \frac{\sum structure\_Nbrimg - \sum proposed\_NBRimg}{\sum structure\_Nbrimg} \times 100[\%] \quad (7)$$

TABLE II
THE NBR-IMG GAIN OF PS_WOPB AGAINST OTHER STRUCTURES

		SIZE OF GOP		
		8	12	15
PS_WOPB vs IBP	8 views	7.16	06.19	04.30
	16 views	23.86	22.82	20.75
PS_WOPB vs PS_WPSB	8 views	7.25	07.54	08.13
	16 views	10.86	10.82	10.75
PS_WOPB vs PS_WPB	8 views	11.35	11.84	12.82
	16 views	13.75	14.13	15.84

The Size-GGOP varies depending on the size of used GOP which can be 8, 12 or 15 images. This variable depends on the chosen number of views. An analysis of the reliability of the PS-WOPB method compared to the other structures in terms of *NBR-img* is shown in Fig. 15. This analysis is performed using 16 views and the three chosen size of GOP which are 8, 12, and 15. The experimental results which are calculated each time using 8 images per GOP are further presented in Table III. For the case of the IBP structure, there are only 19 images that require less than 5% of frames for decoding from the total number of images within a GGOP. Meanwhile for the proposed prediction structure, there are 30 cases that require less than 5% of images from the overall number of frames. For the requirement of more than 15% of frames, there are 27, 8 and 8 cases for the IPB, PS-WPSB and PS-WPB respectively. Surprisingly, there are no cases for the proposed structure which explains the merit of the structure as random access requires generally less than 15% of images. Similarly, Tables IV and V show respectively the different results obtained for 12 and 15 images per GOP.

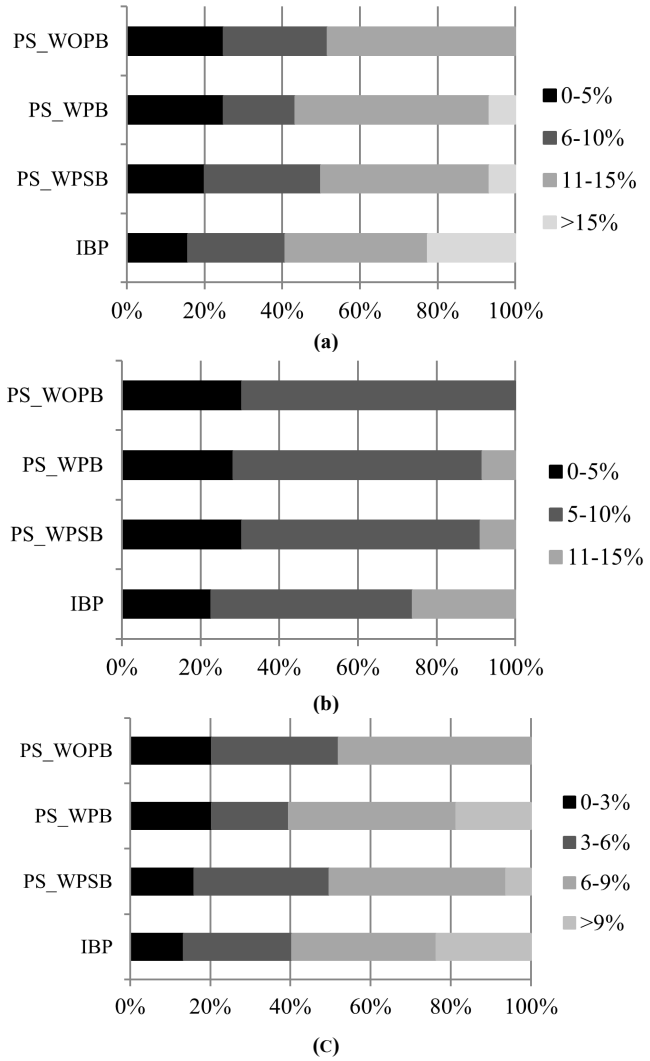

Fig. 15. View random access based on *NBR-img* using 16 views for different sizes of GOP: (a) 8. (b) 12. (c) 15.

The evaluation presented in this section is independent of used video sequences and its encoding properties. In other words, the proposed approach gives the same results in terms of view random access regardless of the used videos or encoding standards. The use of HEVC or H.264/AVC shall give the same results in terms of view random access. This makes the proposed algorithm more convenient and generalizable. GOP of sizes 8, 12, and 15 with a number of

views between 8 and 16 are used during the experiment. The method offers promising results for different sizes of GOPs.

TABLE III
THE NBR-IMG ACCORDING TO A GIVEN INTERVAL, USING 8 IMAGES PER GOP AND 16 VIEWS

	<= 5% OF GGOP	6-10% OF GGOP	11-15% OF GGOP	>15% OF GGOP
IBP	19	30	44	27
PS_WPSB	24	36	52	8
PS_WPB	30	22	60	8
PS_WOPB	30	32	58	0

TABLE IV
THE NBR-IMG USING 12 IMAGES PER GOP AND 16 VIEWS

	<= 5% OF GGOP	6-10% OF GGOP	11-15% OF GGOP	>15% OF GGOP
IBP	41	92	47	0
PS_WPSB	55	109	16	0
PS_WPB	51	114	15	0
PS_WOPB	55	125	0	0

TABLE V
THE NBR-IMG USING 15 IMAGES PER GOP AND 16 VIEWS

	0-3% OF GGOP	3-6% OF GGOP	6-9% OF GGOP	>9% OF GGOP
IBP	30	61	81	53
PS_WPSB	36	76	99	14
PS_WPB	46	43	94	42
PS_WOPB	46	71	108	0

V. CONCLUSION

In this paper, a new inter-view prediction structure is proposed for faster view random access of different views. The proposed approach named as PS-WOPB for *P*rediction *S*tructure *W*ith*O*ut inter-view *P*rediction for non-anchor *B* pictures, is based on increasing the hierarchical level of *B*-views. The PS-WOPB structure is created with the aim to achieve better tradeoff between view random access and bit rate. For evaluation purposes, a new metric is described to examine the efficiency of prediction structures for view random access. Experiments conducted on different datasets of videos, have confirmed the efficacy of the proposed method being applied to multi-view videos taken from different datasets. As a result, view random access is considerably improved and the bit rate is increased with the same quality. Compared to the IBP structure, considerable bit rate saving is obtained with the proposed prediction structure. This is due to increasing the number of *B*-views with respect to the *P*-views in the developed structure.

REFERENCES

[1] J. Park, J. Choi, I. Ryu, and J. Park, "Universal view synthesis unit for glassless 3DTV," *IEEE Trans. Consumer Electron,* vol. 58, no. 2, pp. 706-711, 2012.

[2] J. Konrad and M. Halle, "3-D displays and signal processing – an answer to 3-D ills?," *IEEE Signal Processing Magazine*, vol. 24, no. 6, Nov. 2007.

[3] Y. Morvan, D. Farin, and H.N. Peter, "System architecture for freeviewpoint video and 3D-TV," *IEEE Trans. Consumer Electron*, vol. 54, no. 2, pp. 925–932, 2008.

[4] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies" *Proceedings of the IEEE*, vol. 93, no. 1, pp. 98-110, Jan. 2005.

[5] ITU-T, "Advanced video coding for generic audiovisual services," Recommendation ITU-T H.264.Feb, 2014.

[6] 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, 30th Meeting, Geneva, 2009.

[7] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol,* vol. 17, no. 11, pp. 1461–1473, 2007.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," *in Proc. IEEE International Conference on Multimedia and Expo*, Toronto, Canada, pp. 1929-1932, Jul. 2006.

[9] Pei-Jun Lee, Ho-Ju Lin, and Kuei-Ting Kuo, "Faster mode determination algorithm using mode correlation for multi-view video coding," *IET Signal Process,* vol. 8, no. 5, pp. 565–578, 2014.

[10] S. H. Hany, M. El-Khamy and M. El-Sharkawy, "Blind configuration of multi-view video coder prediction structure", *IEEE Trans. Consumer Electron,* vol. 59, no. 1, pp. 191–199, 2013.

[11] Y. Yang, Q. Dai, J. Jiang, and Y-S. Ho, "Coding order decision of B frames for rate-distortion performance improvement in single-view video and multiview video coding," *IEEE Trans. Image Processing,* vol. 19, no. 8, Aug. 2010.

[12] A. Vetro, T. Wiegand and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE,* vol. 99, no. 4, Apr. 2011.

[13] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol,* vol. 13, no. 7, pp. 688–703, Jul. 2003.

[14] A. Bekhouch,and N. Doghmane, "Proposing a new evaluation metric to improved view random access for multi-view video coding". *in Proc. IEEE 4th International Conference on Image Processing Theory, Tools and Applications,* Paris, France, pp. 167-172, Oct. 2014.

[15] A. Bekhouch and N. Doghmane, "Multiview video coding with an improved prediction structure for faster random access," *Journal of Electronic Imaging*, vol. 22, no.4, 043010, Oct. 2013.

[16] A. Vetro, Y. Su and A. Smolic, "Common test conditions for multiview video coding," JVT-T207, Klagenfurt, Austria, Jul. 2006.

BIOGRAPHIES

Amara Bekhouch received the engineering and magister degrees in computer science from the University of Annaba, Algeria in 2005 and 2009 respectively. He defended successfully his PhD from the same university in 2014. He is now working as computer science lecturer at the University of Souk Ahras, Algeria. His current research interests include areas of image and video compression and multi-view video coding.

Imed Bouchrika received a Ph.D. degree in electronics and computer science from the University of Southampton, Southampton, U.K. in 2008. He has been a Research Fellow with the Information: Signals, Images, Systems Research Group, University of Southampton. He is now a lecturer of computer science at the University of Souk Ahras. His research areas are image processing and visual surveillance.

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 University of Annaba, Algeria. He is also head of the research team Multimedia and Digital Communications in the laboratory of automatic and signal processing of Annaba. His main research interests include signal processing, digital communications and image/video coding.