Light Field Image Coding with Flexible Viewpoint Scalability and Random Access

Ricardo J. S. Monteiro, Nuno M. M. Rodrigues, Sérgio M. M. Faria, and Paulo J. L. Nunes

Abstract

This paper proposes a novel light field image compression approach with viewpoint scalability and random access functionalities. Although current state-of-the-art image coding algorithms for light fields already achieve high compression ratios, there is a lack of support for such functionalities, which are important for ensuring compatibility with different displays/capturing devices, enhanced user interaction and low decoding delay. The proposed solution enables various encoding profiles with different flexible viewpoint scalability and random access capabilities, depending on the application scenario. When compared to other state-of-the-art methods, the proposed approach consistently presents higher bitrate savings (44% on average), namely when compared to pseudo-video sequence coding approach based on HEVC. Moreover, the proposed scalable codec also outperforms MuLE and WaSP verification models, achieving average bitrate saving gains of 37% and 47%, respectively. The various flexible encoding profiles proposed add fine control to the image prediction dependencies, which allow to exploit the tradeoff between coding efficiency and the viewpoint random access, consequently, decreasing the maximum random access penalties that range from 0.60 to 0.15, for lenslet and HDCA light fields.

***Keywords*— Light Field Image Coding, Viewpoint Scalability, Viewpoint Random Access, HEVC**

I. INTRODUCTION

The light field (LF) imaging technology allows to jointly capture the scene radiance and angular information using single-tier lenslet LF cameras (with narrow baseline) or high-density camera arrays (HDCA), with wider baselines. A LF camera is composed by the conventional main lens and image sensor, and an additional microlens array (MLA) [1]. The MLA allows the LF camera to capture both spatial and angular information about the light that converges to the sensor [2]. Depending on the LF capturing device, different levels of both spatial and angular resolution, i.e., number of pixels per viewpoint and number of viewpoints,

respectively, may be achieved [3].

The captured LF information has the ability to convey 3D information about the scene, by effectively capturing several points of view instead of representing just a single 2D perspective. Typically, a LF image is represented as a 4D signal [4], $LF(j,i,y,x)$, where dimensions j and i define the angular resolution, and dimensions y and x define the spatial resolution. This important feature enables several a posteriori image processing manipulations, such as changing the perspective and refocusing [1]. The richer multiview content of LF images also has applications in image recognition, medical imaging [5] and 3D television [6]. This imaging technology also allows for interactive media applications, such as interactive multiview video [7, 8], free viewpoint video streaming [9], and interactive streaming of light field images captured by HDCAs [10] or lenslet LF cameras [11].

The LF technology has recently attracted the interest of many research groups as well as standardization initiatives, such as JPEG Pleno [12] and MPEG-I [13], targeting to improve compression efficiency and to standardize representation formats for LF data, as well as for other types of content like point cloud, holographic and 360-degree video. The on-going research work aims to tackle the inefficiency of conventional image and video coding standards, such as JPEG and HEVC, to encode LF content. The limited performance of these encoders is justified by their inability to exploit the new forms of redundancy present in LF content, namely non-local redundancy, i.e., the redundancy between neighboring micro-images (MI).

Alternatively to the direct use of conventional image and video coding standards, there are three possible alternative approaches to improve the coding efficiency for LF images: 1) exploit the inter-view redundancy, by applying pre- and post-processing tools that convert the LF into a sequence of viewpoints, so called pseudo-video sequence (PVS), and encode it using a standard video codec; 2) exploit the non-local spatial redundancy, by adding novel predictions tools to an existing image codec; 3) designing novel coding approaches specifically for LF images.

Previous proposals in the three mentioned categories are able to achieve significant improvements over traditional imaging codecs in terms of LF coding efficiency [14], [15]. However, despite their coding efficiency, most of them lack efficient viewpoint scalability and random access functionalities.

Viewpoint scalability in a LF codec facilitates compatibility with legacy displays and capturing devices by enabling the use of different representations, e.g., 2D, 3D and multiview, as well as an incremental representation of LF viewpoints, i.e., incremental angular resolution. Viewpoint scalability also allows LF angular resolution to be adjusted dynamically, depending on specific requirements, such as storage space, network conditions in a transmission/streaming scenario or the final display capabilities in terms of angular resolution and processing power. Additionally, the ability to refocus the captured scene at a specific depth

can also be enhanced as the number of viewpoints increases [16]. Standard image codecs such as HEVC, have already scalable extensions like SHVC, but only for spatial, temporal, SNR and color gamut dimensions [17].

Random access enables video encoders to access a specific frame or region without decoding the full video sequence [18], facilitating user interaction. For LF images, since they are represented by a single frame, temporal random access is not applicable. However, viewpoint random access can be extremely useful in scenarios such as LF streaming or in VR/AR applications. The viewpoint random access functionality allows to reduce the amount of decoded data required to render a target viewpoint within the LF image, reducing the decoding delay as well as the computation resources required on the decoder side of the transmission pipeline.

To address these problems, this paper proposes a new LF image coding framework with flexible viewpoint scalability and random access. The provided viewpoint scalability structure is defined using a configurable scalability layer mask, which allows for a large number of viewpoint scalability configurations, based on six different layers and the choice of an optimized reference picture selection. Additionally, flexible viewpoint random access functionality is also specifically addressed. By using different combinations of the associated control parameters, the proposed method is able to adapt to different tradeoffs between viewpoint random access capabilities and coding efficiency. The proposed scalability and random access tools are based upon the approach proposed in [14], which is focused on high coding efficiency and adaptable to any scanning order, outperforming codecs such as MuLE [19] and WaSP [20], developed as part of JPEG Pleno standard. The proposed functionalities in terms of viewpoint scalability and random access enable a large number of configurations, which can be specifically designed for certain application scenarios and requirements. To assess the performance of these encoding profiles, a comprehensive study is presented for, both, lenslet and HDCA LFs.

The remainder of this paper is organized as follows: Section II reviews the state-of-the-art on LF coding approaches, as well as the available solutions for LF viewpoint scalability and random access; Section III presents the proposed PVS-based LF image codec; Section IV presents the proposed viewpoint scalability solution; Section V presents the proposed control parameters that allow fine control viewpoint random access; Section VI presents the experimental results; and, finally, Section VII concludes the paper.

II. Related work

This section briefly reviews the state-of-the-art methods to encode LF images discussing, for the relevant cases, their viewpoint scalability and random access features.

A. LF image coding techniques

The available LF image coding techniques described in the literature rely, essentially, on three types of approaches [21] based on: 1) exploitation of the LF inter-view redundancy, by applying pre- and post-processing tools that convert the LF into a sequence of viewpoints, so called pseudo-video sequence (PVS), and encoding it using a conventional video codec; 2) exploitation of the LF non-local spatial redundancy by adding novel predictions tools to an existent image codec; 3) exploitation of the LF inter-view or non-local spatial redundancy by using novel coding approaches designed specifically for LF content.

1) PVS-based LF image coding

Conventional video encoders, e.g., H.264 and HEVC, use block matching algorithms to exploit the temporal redundancy of video data. These tools can also be successfully used to exploit the inter-view redundancy between viewpoints. To this purpose, some methods use specific scanning strategies [22]–[25] to convert a LF to a PVS that is then encoded as a common video sequence. The most popular scanning strategies include raster, serpentine and spiral scanning. Alternatively, more than one PVS can be generated, therefore interpreting the lenslet LF as a multiview signal. In [26], [27] the viewpoints are interpreted as a HDCA signal, which is then encoded with a modified MV-HEVC encoder using a two dimensional weighted prediction and rate allocation.

2) Exploiting the LF non-local spatial redundancy

Several methods to exploit the non-local spatial redundancy are proposed as additional coding tools for state-of-the-art video coding standards like HEVC. When encoding lenslet LF images, the non-local spatial redundancy is normally much more relevant than the traditional spatial redundancy, therefore, most methods rely on searching algorithms that try to exploit MI similarities. The searching algorithms can have different degrees of freedom and may use one or multiple references [15], [28]–[35]. In [28], a self-similarity (SS) compensated prediction is proposed, taking advantage of the flexible partition patterns used by HEVC. The authors in [15] extended this approach by developing a multi-hypothesis coding method that uses up to two hypotheses for prediction in spatial and time domain. The approaches based on SS can be considered low order prediction (LOP) methods because they are limited to two degrees of freedom (DoF). This limitation reduces the prediction ability to describe the changes in perspective between adjacent MIs. In [33] the authors proposed to evolve the HEVC coding architecture by integrating a high order prediction (HOP) method, which uses a geometric transformation (GT) with up to 8 DoF to compensate changes in perspective. More recently, the authors proposed to implicitly signal the HOP parameters using a training step, both at the encoder and decoder side [34]. Additionally, in [35], an alternative non-local spatial prediction method is investigated, relying on a prediction mode based on locally linear embedding (LLE) integrated in HEVC.

This allows the number of references to be blockwise adjusted between one (similar to unidirectional searches) and up to eight reference signals.

3) Coding approaches specifically designed for LF

As mentioned in Section I, MuLE [19] and WaSP [20], as parts of JPEG Pleno standard, have been specifically designed for LF image coding. MuLE was developed to exploit the full 4D redundancy of the LF image by partitioning the LF image into 4D blocks and then applying a 4D-DCT to each block. The transform coefficients of the 4D-DCT are grouped using hexadeca-trees to generate a stream, which is encoded using an adaptive arithmetic coding. WaSP uses a hierarchical approach to LF image coding that is based on warping, merging and sparse prediction. The reference viewpoints are warped to the location of the current viewpoint; the warped reference viewpoints are merged using one optimal least-squares merger; finally, the overall image, merged to the original viewpoint, is adjusted using a sparse predictor.

Alternatively, one common approach to LF coding is to only encode and transmit part of the viewpoints, normally referred to as structural key views (SKVs), and then using additional side information to generate the remaining non-SKVs at the decoder. Usually, this type of approaches only differ in the type of additional information which is transmitted [36]–[41]. In [36], the non-SKVs are generated using a convolutional neural network (CNN) based on an angular super-resolution algorithm. In [37], coefficients are generated via linear approximation, that are used to generate the non-SKVs as a weighted sum of the SKVs. In [38], non-SKVs are generated using approximated disparity maps that are transmitted to the decoder. In [39], the non-SKVs are generated using depth-image-based rendering (DIBR). In [40], a graph-based transform derived from a coherent super-pixel over-segmentation of the several views is used to encode non-SKVs. In [41], the non-SKVs are encoded using a graph learning approach, which estimates the disparity among the views composing the LF.

B. Viewpoint scalability

In [42], the authors started exploring scalability functionalities for LF by proposing a two-layer LF coding approach for the focused LF camera model. It uses a LF representation that consists of a sparse set of MIs and associated disparity maps. Based on the sparse set of MIs and the associated disparity maps (first layer), a reference prediction LF image is obtained through a reconstruction method that relies on disparity-based interpolation and inpainting. This reconstructed LF image is then used to encode the original LF image (second layer), by encoding the prediction residue. This approach was extended [43] with a third layer of scalability and the use of lossy encoded disparity maps, which improve coding efficiency when compared to the lossless transmission of the disparity maps used in the previous approach.

In order to increase compatibility with legacy displays, the authors in [44] propose a three-layer approach. A certain number of viewpoints was assigned to each layer, i.e., the first layer encodes the central view, the second layer encodes stereo or multiview and the third layer encodes the full LF image. In order to increase the coding efficiency inter-layer prediction was used, to exploit the redundancy between layers. This work was recently extended [11] to a higher number of scalability layers, allowing to improve the coding efficiency by using an exemplar-based algorithm for texture synthesis.

Also, some techniques described in Section II.A.3, such as those described in [36]–[41], may be considered as scalable approaches. Several authors propose to encode only a few viewpoints and use additional information to generate the remaining ones. This type of approaches allow for viewpoint scalability, because the LF image is in fact encoded using two scalable layers. The base layer comprises the SKVs and the enhancement layers include the non-SKVs.

Although some of the previously mentioned coding solutions address viewpoint scalability, in general, they restrict the scalability to a reduced number of layers, i.e., normally less or equal to three. A higher number of scalability layers will benefit the deployment of LF content applications, allowing to have smoother variations in terms of angular resolution and compatibility with a larger variety of LF displays that support different angular resolutions. Additionally, in the existing coding solutions the flexibility in terms of the selected scalability structure is also generally reduced. In the proposed approach, the increased flexibility will allow to select among different configurations, according to the envisaged use case.

*C. **Viewpoint random access***

In order to provide viewpoint random access, the coding algorithms typically constrain the coding dependencies between viewpoints. The more constrained these dependencies are, the higher the viewpoint random access capabilities, however, the coding efficiency tends to be penalized. In [45], the authors propose to eliminate prediction at the encoder, therefore eliminating viewpoint dependency, by using Wyner-Ziv coding for compressing LF images. This work was extended in [46] by using SP-frame predictive encoding. More recently, in [47], the authors decompose 15×15 viewpoints into 25 groups of viewpoints and allocate 4 different dependency levels to each group. In [48], it was proposed a MV-HEVC based coding solution, that allows diagonal viewpoint prediction instead of exclusively allowing horizontal and vertical viewpoint prediction. Experimental results show that allowing diagonal viewpoint prediction provides a good compromise between coding efficiency and viewpoint random access when compared to algorithms that are exclusively based on horizontal and vertical viewpoint prediction. Finally, in [49], the authors propose to split the LF image into 4 different PVS, which then allow for viewpoint scalability over 4 layers using MV-HEVC. Viewpoint random access in [49] is achieved through minimization of the number of dependencies

per viewpoint. To this effect, two reference picture selection (RPS) variants are proposed that increase random access capabilities at the cost of some coding efficiency due to this reduction in viewpoint dependencies.

These solutions, although capable of achieving some balance between random access capabilities and coding efficiency, do not provide fine control over this tradeoff, which is of the utmost importance to support different application scenario constraints. In general, the control over this tradeoff is only managed by varying the size of the reference picture list or by manually changing the RPS, which is very limited.

III. PVS-BASED LIGHT FIELD CODING USING OPTIMIZED RPS

The scalability method presented in this paper is built upon a light field PVS encoder, which is discussed below. A PVS is comprised of a sequence of viewpoints, whose inter-view redundancy may be efficiently exploited by the HEVC inter-prediction tools. The scanning order used to convert the viewpoints into the PVS plays a very important role on exploiting the viewpoint redundancy because: 1) the existence of redundancy between the selected viewpoints is crucial for the performance of the prediction techniques; 2) the scanning order influences the reference pictures used for each viewpoint. Since HEVC is unaware of the scanning order that was used to generate the PVS, the encoder is not able to fully exploit the RPS. Therefore, in this paper the RPS optimization method developed in [14] is used to increase the coding efficiency, which is achieved by implicitly signaling the scanning order to the decoder.

The proposed PVS-based LF image coding approach may be used with any scanning order, which may be signaled to the decoder. Fig. 1 shows the spiral and serpentine scans being applied to an $N \times N$ matrix of viewpoints, where $N = 7$, and $j = [0, 1, \dots, N-1]$ and $i = [0, 1, \dots, N-1]$ are the vertical and horizontal axis spatial positions, respectively, for each viewpoint in the matrix. Given a specific scan order, the decoder can then determine unambiguously the spatial position of each decoded viewpoint.

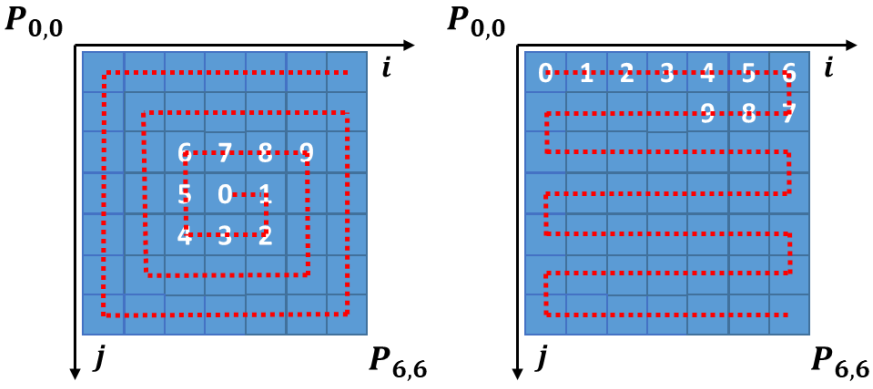


Fig. 1. Spiral (left) and serpentine (right) PVS scanning order applied to a $N \times N$ matrix of viewpoints, with $N = 7$.

The importance of optimizing the RPS is illustrated in the example of Fig. 2, assuming the application of a PVS-based LF coding default configuration, such as the HEVC "Low Delay" configuration [22]. This configuration is a variation of the classic low delay configuration with a QP offset where the first reference picture is the last frame that was encoded and the remaining reference pictures are the last $N - 1$ frames, (N is the total number of reference pictures that have a QP offset of 1 or 0). For example, for $N = 4$, as shown in Fig. 2, frame 22 will have frame 21 as the first reference picture, and the remaining three reference frames are the last frames that used a QP offset of 1 or 0, i.e., frames 20, 16 and 12. This means that for the PVS scenario illustrated in Fig. 2, for both the spiral and serpentine scans, the inherent 2D spatial locations of each viewpoint are not considered and the used reference pictures are not selected based on their expected correlation with the current viewpoint, as shown for frames 14 and 22. As can be seen in the spiral scanning order example of Fig. 2 (left), when encoding frame 22, the default RPS uses, besides frame 21, frames 20, 16 and 12, instead of frames 6, 7 and 8, which are closer and would be, in principle, better spatially correlated references.

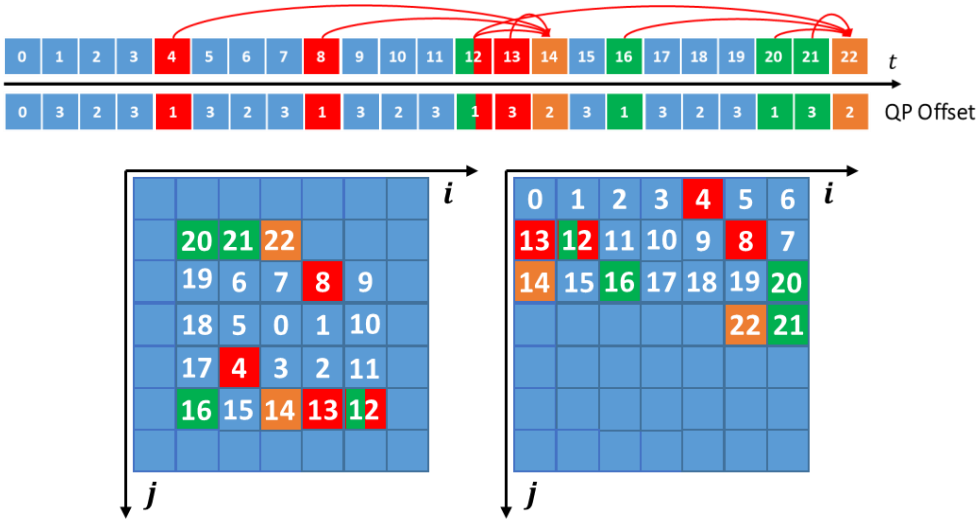


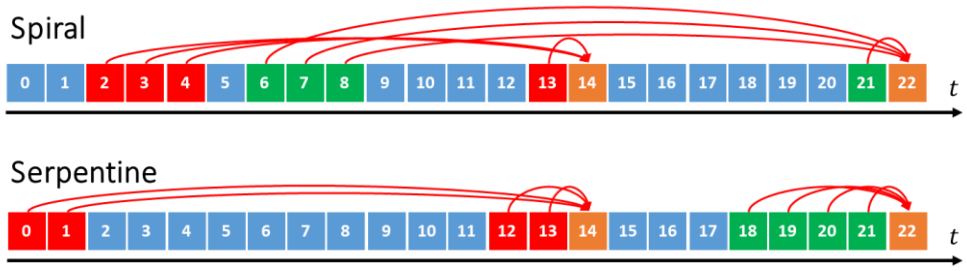
Fig. 2. RPS for frames 14 (red) and 22 (green) when using the "Low Delay" configuration, represented in the temporal domain (top) and corresponding $N \times N$ matrix of viewpoints for spiral (left) and serpentine (right) scanning orders (bottom).

To create an optimized RPS, it is desirable to maximize the correlation between the current viewpoint and the selected reference pictures. Since the current viewpoint tends to be similar to its neighboring viewpoints, when compared to the viewpoints that are furthest away, the proposed RPS approach in this paper is to select

the viewpoints that are located in close proximity to the current viewpoint to be encoded. Therefore, the R spatially closest reference viewpoints, in terms of Euclidean distance, d (1), to the current viewpoint are selected, with d computed as:

$$d(P_{ji}, P_{ji}^{r}) = \sqrt{(j - j_r)^2 + (i - i_r)^2}, \quad (1)$$

where $r = [0,1, \dots, R-1]$, P_{ji} is the current viewpoint spatial position and P_{ji}^{r} the spatial position of each reference viewpoint. Once the R reference viewpoints for each viewpoint to be encoded are found, they are organized in an ascending order of distance in its reference picture list. A possible alternative metric would be the Manhattan distance, however, since R tends to be relatively low, i.e., $R \leq 4$, the selected reference viewpoints tend to be mostly the same. When two or more reference pictures have the same distance, the one with the lowest frame number is selected first. Fig. 3 illustrates the optimized RPS for $R = 4$, for two frames and for the spiral and serpentine scanning, after minimizing the Euclidean distance (1). From Fig. 3, for the examples of frames 14 and 22 previously shown in Fig. 2, it is possible to see that, in the temporal domain, there is no longer a regular RPS pattern, however, it is expected that the correlation between the reference viewpoints and the viewpoint to be encoded is higher than in the traditional approach, with benefits in terms of coding efficiency. One important consequence of this method in that the RPS is variable and adaptively defined for each scanning order. This can be observed in Fig. 3, where, in contrast to Fig. 2, the temporal representations are dependent on the scanning order, i.e., in Fig. 2 the temporal representation is the same for both scanning orders.



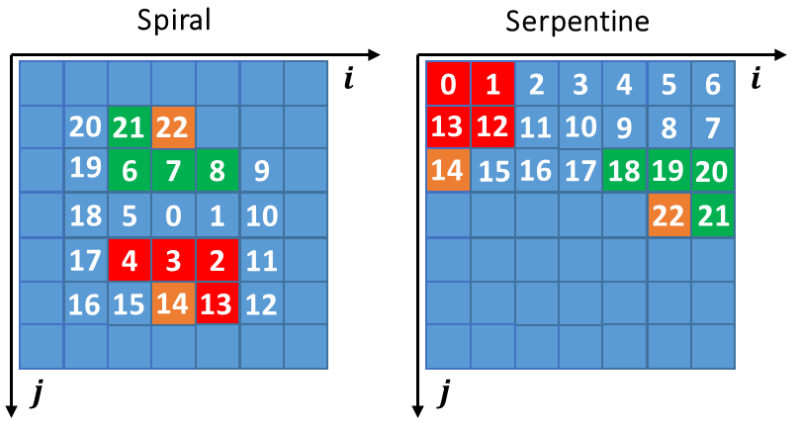


Fig. 3. Optimized RPS for frames 14 and 22 represented in the pseudo-temporal domain (top) and the corresponding $N \times N$ matrix of viewpoints (bottom).

IV. VIEWPOINT SCALABILITY

In this section, the coding approach described in the previous sections is evolved into an efficient viewpoint scalable framework motivated by its ability to adapt to changes in the PVS scanning order, which is particularly useful to achieve viewpoint scalability. This way enabling LF content to be captured and displayed using various types of devices that range from conventional 2D up to full-fledged LF cameras and displays. The following sections present the major functionalities enabled by viewpoint scalability, as well as the proposed scalability structure to enable them.

A. LF viewpoint scalability features

Viewpoint scalability allows the LF content to be represented in several layers, where each layer comprises a group of viewpoints, ultimately allowing compatibility with acquisition and display devices with different capabilities, like spatial resolution, angular resolution and processing power. Additionally, it allows to adaptively stream LF content over networks with changing conditions, making this a valuable feature at several stages of the LF transmission pipeline. The main advantages for the capturing and encoding stages are:

- **Support for legacy capturing devices** – This allows the scalable representation to be compatible with 2D and 3D/Stereo capturing devices;
- **Support for both lenslet and HDCA LFs** – When both types of LFs are represented by viewpoints the major difference between them is the baseline between the several viewpoints, therefore the coding process for both types of LFs should be seamless.
- **Support for flexible encoding profiles** – This allows to adjust the transmitted data depending on criteria such as the available processing power, the available storage space and the network conditions.

Viewpoint scalability allows each consecutive layer to be decoded cumulatively, achieving progressively higher angular resolution as more layers are decoded. This brings some important features for the decoding and displaying stages:

- **Support for legacy display devices** – Non-LF displays, e.g., 2D and 3D/Stereo displays, may receive and display a subset of the whole LF, namely the first layers that may include the central view and the first side views.
- **Support for LF displays with varying capabilities** – The decoded subset of the LF may have an angular resolution or processing power requirements suited to the display device capabilities; more advanced LF displays can still present the whole LF by receiving and decoding all the LF layers.

B. Proposed viewpoint scalability structure

Various configurations of scalability layers can be used to support the above-mentioned features and ultimately the correct combination depends on the practical application, i.e., different applications may require a different number or disposition of layers. Regardless, a hierarchical structure was adapted from [20], which provides a well distributed structure of viewpoints throughout the coding process for the viewpoint scalability features mentioned in this section. The proposed scalability structure is shown in Fig. 4, where the orange and the yellow blocks represent, the viewpoints from the current and previous layers, respectively, and the blue blocks stand for the viewpoints not considered yet, i.e., belonging to higher scalability layers.

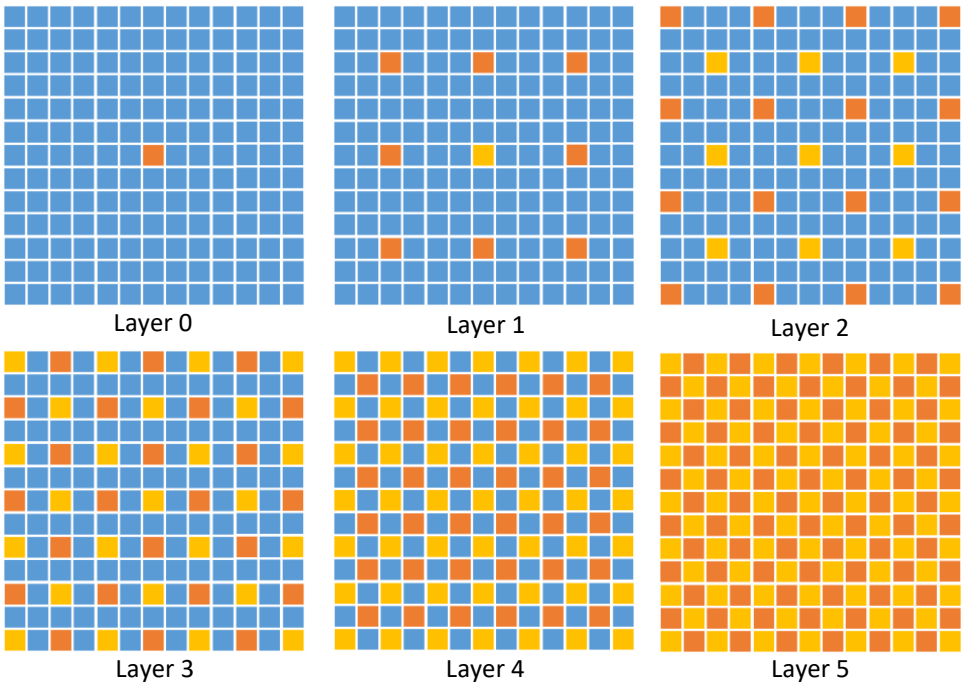


Fig. 4. Proposed scalability structure (each square represents a different viewpoint): the orange squares are the viewpoints that compose the current layer and the yellow squares compose perviously encoded layers.

The viewpoint distribution across the several layers roughly follows a $(2n+1) \times (2n+1)$ pattern, where n is the layer number. Layer 0 allows for the central view to be encoded/decoded independently, e.g., for compatibility with 2D displays. Layer 1 allows a 3×3 LF to be displayed, or a pair of views can be selected to be displayed in a stereo display. After decoding layers 2, 3 and 4, the resulting LF is roughly a 5×5, 7×7 and 9×9 LF, respectively. These intermediate layers allow some granularity in terms of angular resolution and processing capabilities, as well as network resources required. Finally, Layer 5 includes the remaining viewpoints that represent a 13×13 LF image. Being a regular pattern, it can be expanded and adjusted to the number of viewpoints required by the application.

In order to encode and decode the various scalability layers, the chosen scanning order needs to be adapted to the proposed scalability structure. Furthermore, the use of the optimized RPS approach proposed in Section III is fundamental to achieve high coding efficiency and to make it approximately independent of the adopted scanning order, as the RPS adapts to any scanning order that may be used.

The adaptation of the scanning order to the proposed scalability structure is performed using a scalability layer mask, as shown in Fig. 5. Each number represents the layer that each viewpoint is assigned to. The spiral scan adaptation to the proposed scalability structure is generated by applying the spiral scan to viewpoints within a given layer, e.g., the spiral marked in black in Fig. 5 corresponds to the Layer 4 spiral scan. This scalability layer mask allows the creation of a scalable version of the spiral scanning order as well as a configurable scalable structure that can be adapted to the application requirements. When encoding the LF image using the scalable spiral scan, the optimized RPS coding technique will adapt to the new scanning order. This means that for each frame that is encoded, the RPS will be optimized by minimizing the Euclidean distance (1) between the frame to be encoded and its possible references. This is illustrated in Fig. 6, where Layer 0 (black) is composed by frame 0 (being the first frame it will be encoded as Intra), Layer 1 (green) is composed by frames 1 to 8 and Layer 2 (grey) is composed by frames 9 to 14. In the case of frame 11, it is possible to see that the closest available reference viewpoints, according to the Euclidean distance, are frame 0 (Layer 0), and frames 5 to 7 (Layer 1), therefore these viewpoints compose the optimized RPS, since they will be already available at decoding time of frame 11. The same process is applied to frame 14, where, in this example, the closest available reference viewpoints are frames 1 and 2 (Layer 1) and frames 9 and 13 (Layer 2). It is therefore expected that the scalable spiral scanning order combined with the optimized RPS will allow for high coding efficiency, while adding support for viewpoint scalability. Additionally, as shown

in the temporal domain in Fig. 6, each layer is being encoded cumulatively, therefore each layer only has dependencies on the previous and current scalability layers.

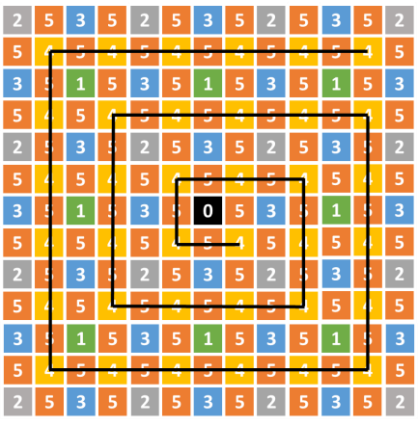


Fig. 5. Scalability layer mask: each numbered square represents a different viewpoint that bellongs to a specific layer. The black line is the result of appliying the spiral scanning order to Layer 4.

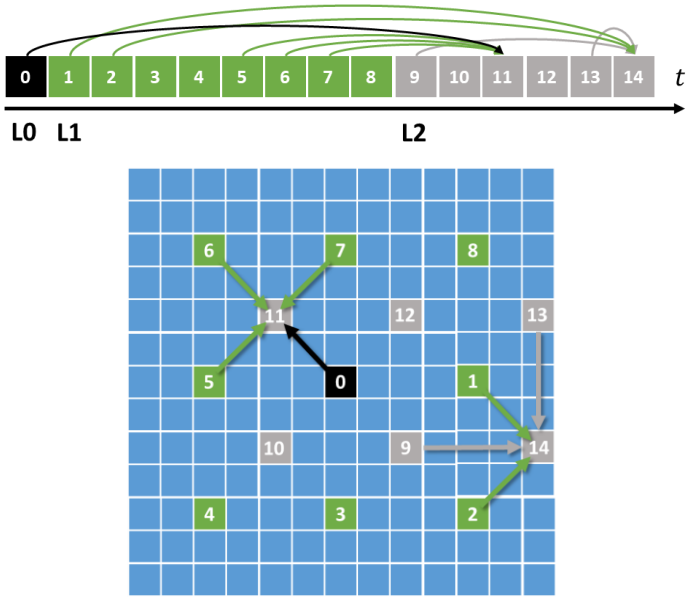


Fig. 6. Optimized RPS when applied to a scalable spiral scanning order: the black, green and grey squares correspond to Layer 0, 1 and 2, respectively.

V. VIEWPOINT RANDOM ACCESS

Random access points are used in video coding to facilitate the interaction with a video sequence. This way, it is possible to navigate in the video sequence without having to decode the entire bitstream. In such case, Intra frames are used as random access points because they can be decoded independently from the remaining frames, as only intra prediction modes are used.

For the scalable LF framework, the viewpoint navigation has to be considered for different scenarios, namely when using a 2D/3D display, i.e., one or two viewpoints; a LF display, i.e., $N \times N$ viewpoints; or even a head mounted display (HMD) [50], i.e., one (2D) or two (3D) viewpoints. This means that during visualization the user must be able to access any desired viewpoint or group of viewpoints. Nevertheless, due to inter-viewpoint predictions used in the coding process, several dependencies are created, which increase the number of viewpoints that must be decoded in order to visualize a specific viewpoint. The aim here is to minimize the number of viewpoint dependencies, i.e., to maximize viewpoint random access, allowing to:

- **Improve LF navigation efficiency** – The number of necessary viewpoints to decode the desired viewpoints should be kept as low as possible.

- **Reduce decoding delay** – The lower the number of required decoded viewpoints the lower de decoding delay for decoding a given viewpoint; the viewpoint decoding delay should be kept as low as possible.

- **Reduce computational complexity** – The lower the number of required decoded viewpoints the lower de computational power needed for decoding a given viewpoint, facilitating access for decoders/displays with limited processing power.

Viewpoint random access is therefore an important functional feature but this comes, in principle, at the cost of a reduction in coding efficiency. Consequently, it is important to have fine control over the tradeoff between viewpoint random access and coding efficiency that can be adjusted depending on the envisaged application scenario. For this purpose, the next sections will discuss, firstly, how to quantify the random access capabilities, and, secondly, the proposed random access control parameters that allow for fine control over the described tradeoff.

A. Quantifying the random access capabilities

The measurement of the random access capability of a given coding solution can be accessed using the random access penalty (RAP) metric as suggested by JPEG Pleno [51]. This metric is defined by (2).

$$RAP = \frac{\#\ encoded\ bits\ required\ to\ access\ a\ RoI}{\#\ encoded\ bits\ to\ decode\ the\ full\ LF} \tag{2}$$

The definition of the Region of Interest (RoI) depends on the application scenario and the coding algorithm, some examples include, specific viewpoints or specific pixels. In this case, since the coding algorithm uses

a viewpoint-based representation, the RoI corresponds to a specific viewpoint. The RAP for a specific viewpoint is proportional to the amount of bits required to decode that viewpoint, which includes also the amount of bits required to decode its reference viewpoints. Since these reference viewpoints may also have other dependencies, it may happen that the full LF needs to be decoded in order to decode a given RoI, which results in a $RAP = 1$. However, if only part of the LF image needs to be decoded, then $0 < RAP < 1$. Since the RAP depends on the selected RoI, in order to make a more conservative measurement, only the maximum value will be considered which corresponds to the worst-case scenario, i.e. the viewpoint that requires the largest amount of bits to be decoded.

*B. **Proposed viewpoint random access control parameters***

The main factor that influences the RAP is the amount of inter-viewpoint dependencies created during the coding process and the amount of bits associated with each dependency. In the proposed scalable framework, there are several elements that influence these dependencies, therefore, three random access control parameters are proposed, which include the reference picture list (RPL) size, the maximum dependency layer (MDL) and the number of viewpoint regions (NVPR). These control parameters allow for a large number of interactions with the inter-viewpoint dependencies when used in combination.

1) Reference picture list size

Increasing the RPL size per viewpoint increases the maximum number of reference pictures, which is likely to improve the coding efficiency up to a certain limit. However, since more inter-viewpoint dependencies are created, which also have their own dependencies, the RAP increases as well.

2) Maximum dependency layer

This parameter sets the scalability layers containing viewpoints that can be used as reference viewpoints. As shown in Fig. 7; when $MDL = 2$, only the first 9 viewpoints (orange blocks) can be used as references for the remaining viewpoints, which corresponds to Layers 0 and 1. A lower coding efficiency but a better (lower) value of RAP is expected by using a lower MDL value. In Fig. 7, several examples are shown, organized from more restrictive to less restrictive, in terms of inter-viewpoint dependencies, where the corresponding MDL is set from 2 to 6 (all references available, i.e., no restrictions).

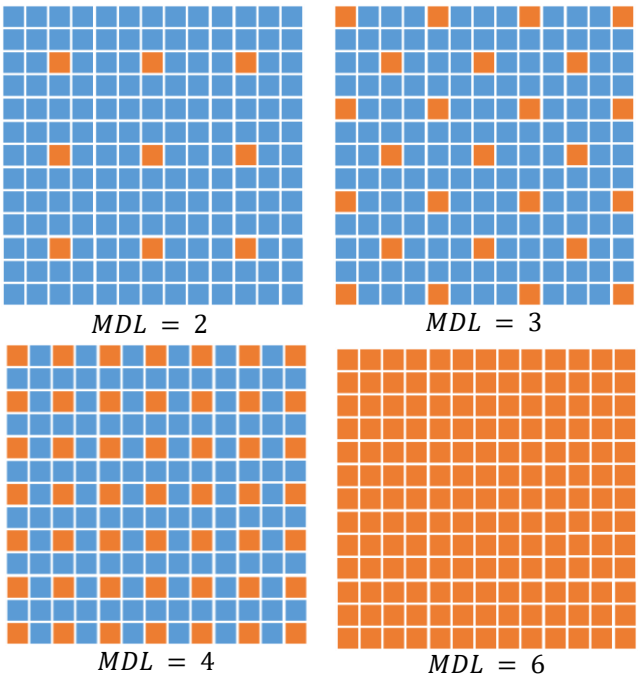


Fig. 7. Different MDL values and consequent selection of possible reference viewpoints. For each case, the orange blocks represent the viewpoints that may be selected as references.

3) Number of viewpoint regions

When the user navigates interactively along the LF, either through a 2D/3D display (LF display with different capabilities) or an HMD, from a given viewpoint location, certain spatial regions of the LF will be more likely to be accessed next than others [52]. Therefore, it makes sense to separate/cluster certain spatial regions of the LF during the coding and decoding process, to ensure that only spatially close viewpoints are used as reference viewpoints. These spatial regions ensure that each region can be encoded and decoded independently from the remaining regions. The RAP is, therefore, expected to improve (lower value) for a higher NVPR. Additionally, these spatial regions can be overlapping or non-overlapping regions. In order to create non-overlapping spatial regions, each region is required to have at least one Intra frame. Fig. 8, illustrates this concept for different configurations of overlapping and non-overlapping spatial regions. The top examples use two and four overlapping regions and the bottom examples use five and nine non-overlapping regions. The red lines represent the limit of the spatial region. The yellow blocks show the viewpoints that belong to more than one region, in the case of the overlapping regions, e.g., in $NVPR = 4$, the block to the left of the I frame (central viewpoint) belongs to Region C and B. As mentioned above, when using non-overlapping regions, at least on Intra frame is required per region, which is the reason why five and nine Intra frames are used roughly in the center of each region in the bottom examples of Fig. 8. In this case, the higher number of Intra frames is expected to also decrease the RAP at the cost of a slight decrease

in coding efficiency.

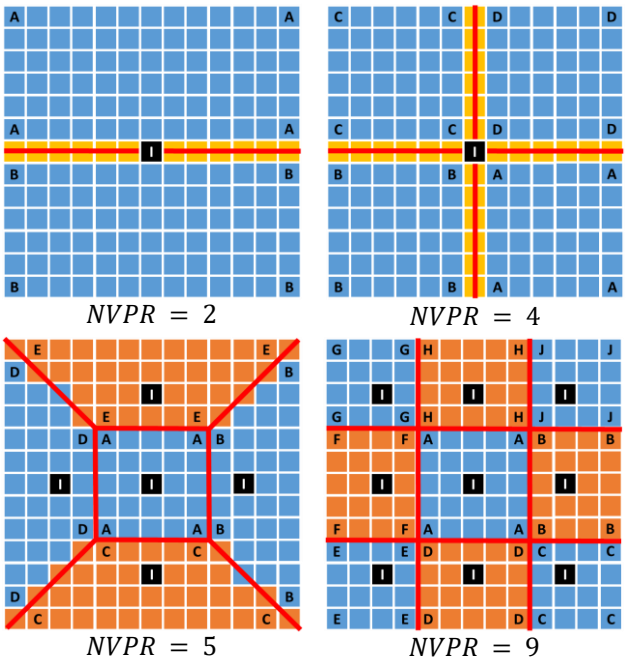


Fig. 8. Definition of viewpoint regions and corresponding I frames for different NVPR values per LF.

4) Control parameter combination

The proposed control parameters can be used in several combinations, which is expected to provide a fine control over the tradeoff between random access capabilities and coding efficiency. A list of random access profiles are suggested in the end of Section VI resulting from the achieved maximum RAP and coding efficiency results. These profiles include combinations of control parameters that can be used to maximize the coding efficiency or the random access capabilities, or balanced profiles, biased towards one metric or the other.

VI. EXPERIMENTAL RESULTS

In this section the performance of the proposed scalable framework is evaluated against several state-of-the-art LF coding solutions. First, the testing methodology, including the processing chain for objective quality assessment, is explained. Then, experimental results comparing the RD performance of the proposed codec using various viewpoint scalability and random access configurations are presented and discussed. Statistical in-depth results about the computational complexity and the random access penalty metric are shown to support the experimental results analysis.

A. Test methodology

In order to evaluate the RD performance of the proposed LF coding solution, the EPFL dataset, comprised of 12 lenslet LF images acquired using a Lytro Illum camera, is used [53]. The "RAW" lenslet LF images are first converted into the 4D LF representation using the LF Toolbox [54] and then color and gamma correction is applied, as suggested by JPEG Pleno Common Test Conditions (CTCs) document [51]. The LF images are converted to the PVS representation, prior to being encoded and decoded. After the decoding step, 13×13 viewpoints are generated with a resolution of 625×434 pixels, using the YUV 4:4:4 10-bit color format [51]. Additionally, three HDCA LFs are also used, namely, *Greek*, *Sideboard* and *Set2*. The HDCA images *Greek* and *Sideboard* are composed by, 9×9 viewpoints (512×512 pixels each) and *Set2* is composed by 33×11 viewpoints (1920×1080 pixels each), all in the YUV 4:4:4 10-bit color format.

Several LF image coding solutions were selected as benchmarks, namely: HEVC-PVS [22], HEVC-PVS-RA [22], HEVC-OPT [14], WaSP [20] and MuLE [19]. HEVC-OPT, described in Section III, is the non-scalable basis of the proposed scalable codec HEVC-SLF. HEVC-PVS and HEVC-PVS-RA use HEVC with standard (non-optimized) PVS with two configurations, "Low Delay" and "Random Access", respectively. Several "Random Access" configurations have been tested using different intra period (IP) values ranging from 8 to 64. The proposed codec is tested under two variants:

- **HEVC-SLF** – The codec that includes the scalable functionalities described in Section IV.
- **HEVC-SLF-RA** – The codec that includes both the scalable functionalities described in Section IV, as well as the viewpoint random access functionalities described in Section V.

A spiral scanning order was used for the proposed codecs as well as HEVC-PVS, HEVC-PVS-RA and HEVC-OPT, since it has been demonstrated in previous works to have overall similar performance to the serpentine scanning [22]. Additionally, when using the optimized RPS, i.e., HEVC-OPT, the authors verified for lenslet LF images that the spiral scanning performs better than the serpentine scanning. The proposed codecs use the optimized RPS method applied to a scalable spiral scanning order as shown in Fig. 6. A list of all tested codecs with the respective coding parameters is given in Table I, where the HEVC-based benchmarks and proposed codecs are based on HM-16.9. The different QPs and λ values (see Table I) allow the use of a common bitrate range for all tested codecs, enabling to compare their results.

TABLE I – LIST OF TESTED CODECS AND RESPECTIVE CODING PARAMETERS

Codec	Coding parameters
HEVC-PVS HEVC-PVS-RA HEVC-OPT HEVC-SLF HEVC-SLF-RA	$QP = [17, 22, 27, 32, 37, 42]$
MuLE	$\lambda = [270, 3\,880, 30\,000, 310\,000, 4\,600\,000]$
WaSP	$Target\ bpp = [0.001, 0.005, 0.02, 0.1, 0.75]$

For each codec, the RD analysis is performed by comparing the size of the bitstream (rate) and the average PSNR-YUV (distortion) of all viewpoints generated on the decoder. The average PSNR-YUV for the viewpoints is calculated by comparing the decoded viewpoints of the different codecs with the reference viewpoints generated by the same process as the decoded viewpoints but using the original (not encoded) LF.

When encoding the HDCA LFs using HEVC-SLF and HEVC-SLF-RA, the scalability mask shown in Fig. 5 was truncated according to the number of viewpoints of each respective LF, e.g., for *Greek* it was truncated from 13×13 to 9×9 viewpoints. Additionally, due to the rectangular organization of the viewpoints in *Set2*, this LF was partitioned into three groups of 11×11 viewpoints, which are independently encoded and decoded. However, when calculating the average PSNR-YUV and the size of the bitstream all the 33×11 viewpoints are considered.

B. Viewpoint scalability assessment

The experimental results in Table II show the average BD-PSNR-YUV and average BD-RATE for the 15 lenslet test images, comparing the proposed HEVC-SLF with HEVC-OPT, MuLE, WaSP, HEVC-PVS and HEVC-PVS-RA. In this case an IP of 64 was used for the HEVC-PVS-RA, which is the parameter that allows the highest coding efficiency among the IP values tested. From Table II it is possible to observe that the proposed HEVC-SLF achieves a reduction in the average BD-Rate and an increase in the average quality, outperforming all the other tested benchmarks. The bitrate savings against the tested benchmarks are consistent across both types of LFs. When analyzing each image result individually it is possible to see that HEVC-SLF is only outperformed by HEVC-OPT for LF images I09, I11 and I12, and by MuLE for LF image I02, in terms of bitrate savings. The fact that the proposed scalable codec (HEVC-SLF) outperforms its non-scalable version, HEVC-OPT, is explained by the use of an RPS adapted to the PVS scanning order, as described in Section III, which increases the average viewpoint correlation, notably for the higher layer

viewpoints. It was observed that the coding efficiency of HEVC-SLF does not outperform HEVC-OPT when viewpoints from the first three layers. However, as the remaining layers have a higher number of viewpoints, which are closer to the viewpoint being encoded, the overall encoding efficiency surpasses that of HEVC-OPT. The highest gains in coding efficiency over HEVC-OPT occur when the viewpoint currently being encoded lies in a central position relative to its reference viewpoints, such as viewpoint 11 in Fig. 6. This positioning of references around the encoding viewpoint does not take place in HEVC-OPT because the spiral scan grows outwards from the central viewpoint. Table II also shows a notable drop in performance for MuLE when encoding HDCA LF images, which stems from the fact that MuLE, although being very efficient for lenslet LF images, does not have prediction tools to compensate the wider baselines present in this type of images [19].

In order to analyze the results of the proposed HEVC-SLF codec regarding the viewpoint scalability features, Fig. 9 shows the most important coding indicators, including the average distribution of bits per layer and encoding and decoding times, for a six layer configuration. A QP value of 27 was chosen for this test because it represents a consistent intermediate point in terms of compression ratio and objective quality.

In the first lines of Fig. 9 it is possible to see that the number of bits generated by Layer 0 is fairly high, considering that only one (the central) viewpoint is encoded. Layer 2 is particularly smaller for the HDCA images because the scalability layer is truncated to 9×9 and 11×11 viewpoints as explained in Section VI.A. As expected, scalability Layer 5 normally carries most of the information, as has the highest number of viewpoints, e.g., 84 for the lenslet LF images.

TABLE II – BD-PSNR-YUV AND BD-RATE RESULTS OF THE PROPOSED HEVC-SLF VS HEVC-OPT, MULE, WASP, HEVC-PVS AND HEVC-PVS-RA

Test LF Image	**vs HEVC-OPT [14]**		**vs MuLE [19]**		**vs WaSP [20]**		**vs HEVC-PVS**		**vs HEVC-PVS-RA**	
	BD-PSNR	BD-RATE	BD-PSNR	BD-RATE	BD-PSNR	BD-RATE	BD-PSNR	BD-RATE	BD-PSNR	BD-RATE
I01	0.24 dB	-9.97 %	0.32 dB	-13.65 %	1.54 dB	-46.94 %	1.52 dB	-48.80 %	0.80 dB	-30.99 %
I02	0.41 dB	-16.00 %	0.00 dB	1.21 %	1.04 dB	-29.22 %	1.67 dB	-50.79 %	0.91 dB	-33.22 %
I03	0.22 dB	-9.11 %	0.56 dB	-18.57 %	1.17 dB	-29.81 %	1.58 dB	-48.65 %	0.97 dB	-34.81 %
I04	0.15 dB	-7.61 %	0.40 dB	-20.48 %	0.95 dB	-37.73 %	1.13 dB	-44.97 %	0.81 dB	-34.97 %
I05	0.14 dB	-7.59 %	0.60 dB	-26.05 %	1.27 dB	-42.75%	1.15 dB	-48.33 %	0.66 dB	-32.81 %
I06	0.06 dB	-4.82 %	1.44 dB	-52.35 %	2.45 dB	-73.46 %	0.93 dB	-43.51 %	0.52 dB	-29.36 %
I07	0.05 dB	-2.00 %	0.59 dB	-19.78 %	1.68 dB	-43.61 %	1.03 dB	-39.61 %	0.50 dB	-22.99 %
I08	0.06 dB	-4.04 %	1.48 dB	-50.51 %	2.26 dB	-66.43 %	0.93 dB	-40.71 %	0.48 dB	-26.20 %
I09	-0.12 dB	6.10 %	0.65 dB	-21.31 %	1.30 dB	-32.92 %	1.35 dB	-45.94 %	0.70 dB	-28.20 %
I10	0.01 dB	-0.88 %	0.62 dB	-24.62 %	1.37 dB	-46.33 %	1.26 dB	-45.69 %	0.96 dB	-38.46 %
I11	-0.22 dB	14.60 %	0.59 dB	-22.51 %	2.61 dB	-66.20 %	0.71 dB	-35.57 %	0.25 dB	-14.75 %

I12	-0.16 dB	7.64 %	1.30 dB	-38.51 %	2.21 dB	-52.11 %	1.45 dB	-46.85 %	0.64 dB	-25.92 %
Avg. lenslet	**0.17 dB**	**-2.66 %**	**0.71 dB**	**-25.59 %**	**1.65 dB**	**-47.29 %**	**1.23 dB**	**-44.95 %**	**0.68 dB**	**-29.39 %**
Greek	0.18 dB	-4.95 %	3.71 dB	-71.96 %	1.56 dB	-37.73 %	1.51 dB	-35.64 %	1.18 dB	-28.69 %
Sideboard	0.35 dB	-10.28 %	3.86 dB	-72.90 %	2.06 dB	-49.81 %	1.64 dB	-38.21 %	1.32 dB	-31.88 %
Set2	0.16 dB	-4.03 %	11.53 dB	-97.72 %	2.41 dB	-46.10 %	2.17 dB	-43.22 %	1.65 dB	-34.85 %
Avg. HDCA	**0.23 dB**	**-6.42 %**	**6.37 dB**	**-80.86 %**	**2.01 dB**	**-44.55 %**	**1.77 dB**	**-39.02 %**	**1.38 dB**	**-31.81 %**

In the case of lenslet LF images, it is possible to observe that the size per layer is quite regular, despite increasing number of viewpoints. This is justified by the fact that the last layers are composed by viewpoints that are closer to each other, which reduces the disparity between the viewpoints to be encoded and their references. Therefore, the encoder is able to perform inter-view prediction more efficiently, thus increasing the compression ratio, without necessarily affecting the objective quality.

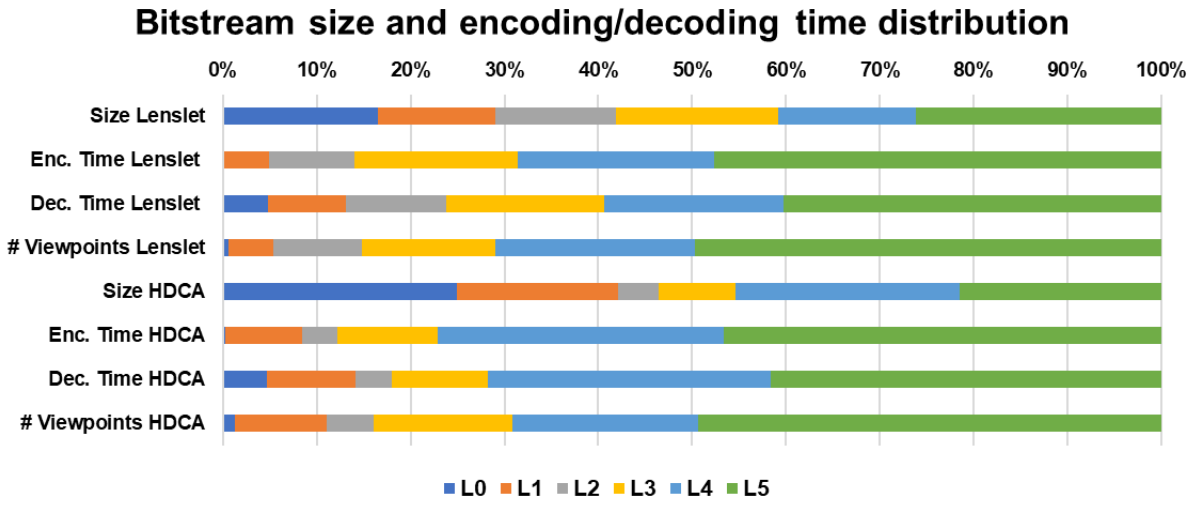


Fig. 9. Bitstream size and encoding and decoding times per layer for HEVC-SLF for QP 27.

The relative average computational complexity, in terms of runtime, of the proposed HEVC-SLF, for every scalability layer for QP 27 is shown in Fig. 9 where it is possible to observe that, differently from the bits distribution per layer, the time distribution required to encode and decode each layer is similar. Consequently, the time required to encode and decode each scalability layer is roughly proportional to its number of viewpoints. The only exception is Layer 0, that is an Intra frame that takes less time to encode than the inter frames, but it takes longer time to decode. It is worth mentioning that Layer 0 encoding time is very small relative to the remaining layers, which is the reason why it is not noticeable on a linear scale as in Fig. 9. The gradual increase of the computational complexity codec along the several scalability layers, as mentioned in Section IV, may be useful in scenarios where the computational power is scarce.

The computational complexity of all tested methods is shown in Table III for a representative test image, I04.

These tests were performed using a PC equipped with an Intel Core i7 CPU 4790K@4.0GHz and 32GB of RAM, running Ubuntu 16.04 The runtimes for MuLE and WaSP were obtained for cases where the resultant objective quality is similar to the HEVC-based codecs, i.e., MuLE using a lambda of 3880 and WaSP using a target bitrate of 0.1 bpp. From Table III it is possible to observe that all the HEVC-based codecs have similar encoding and decoding runtimes. MuLE and WaSP have faster encoders than the HEVC-based codecs, however, their decoders are slower than the HEVC-based ones.

TABLE III – CODEC COMPUTATIONAL COMPLEXITY COMPARISON

Codec	Encoder		Decoder	
	Run Time [s]	vs HEVC-PVS	Run Time [s]	vs HEVC-PVS
HEVC-PVS	1192	-	2.83	-
HEVC-PVS-RA	1013	0.85	**2.05**	**0.72**
HEVC-OPT	1018	0.85	2.96	1.05
MuLE	**209**	**0.18**	15.67	5.54
WaSP*	214	0.18	32.68	11.55
HEVC-SLF	1245	1.04	3.83	1.35

*using multithread (8 threads)

C. Viewpoint random access assessment

In this Section the performance of the proposed viewpoint random access solution (named HEVC-SLF-RA) is evaluated and the effects of different configurations of the control parameters described in Section V are discussed. These control parameters include the RPL size per viewpoint (2 or 4); the MDL (2, 3, 4 and 6 – all, as presented in Fig. 7); and the NVPR (1 to 9, as presented in Fig. 8), which includes a configuration with a single region (NVPR of 1), configurations with overlapping regions (NVPR of 2 and 4) and configurations with non-overlapping regions (NVPR of 5 and 9).

Fig. 10 presents the variation of maximum and average RAP values for several QPs, for images I10 and *Sideboard*. These maximum and average RAP values were computed by calculating the RAP, as explained in (2), for each individual viewpoint respectively from image I10 and *Sideboard* after being encoded with the several presented QPs. This generates a set of RAP values, for each QP, which is the same size as the number of viewpoints in each respective LF image, e.g. 13×13, in the case of I10. Finally, the maximum and average RAP value is calculated for each set and plotted as a function of the QP, as shown in Fig. 10. It is possible to observe that both maximum and average RAP consistently decrease for lower QPs, i.e., higher objective quality. From Fig. 10 it is also possible to see that although both maximum and average RAP follow the same trend for the several QPs, the maximum RAP is significantly higher than the average RAP, i.e., the maximum RAP corresponds to an outlier case. Although the average RAP better represents the entire coding

process, the maximum RAP is used in the following discussion because it corresponds to the worst-case scenario among the lenslet and HDCA LFs, which is important when defining practical application requirements. Based on the results from Fig. 10, a QP value of 27 will be used as reference to evaluate the proposed codec, since it corresponds to an intermediate point not only in terms of compression ratio and objective quality, but also in terms of expected RAP values.

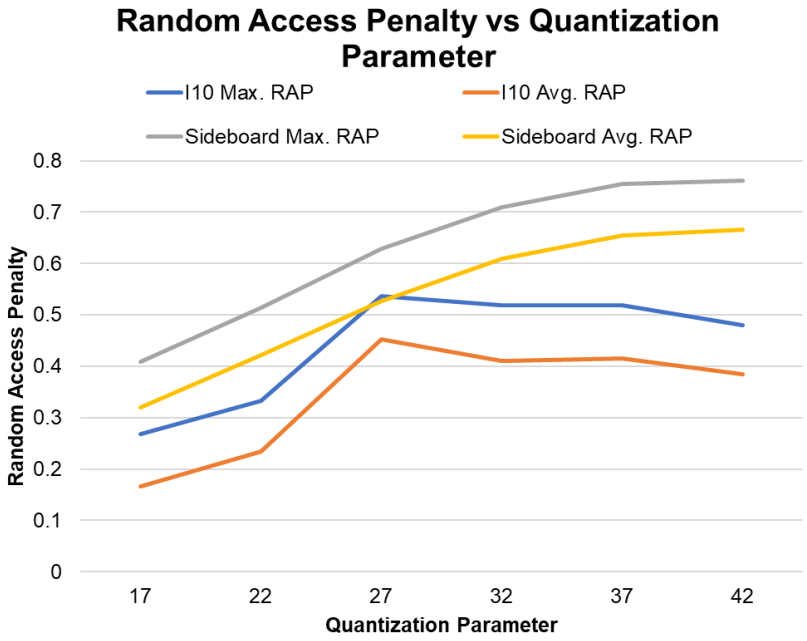


Fig. 10. Maximum and average RAP for images I10 and Sideboard for six different QPs.

The experimental assessment of HEVC-SLF-RA is shown in Table IV and Fig. 11 and Fig. 12, for both types of LFs. Each codec configuration uses a specific combination of the proposed control parameters (RPL, MDL and NVPR) highlighted in italic in Table IV and annotated for each point in Fig. 11 and Fig. 12 that plot the BD-RATE over HEVC-PVS as a function of the maximum and average RAP values, respectively. These maximum and average RAP values were computed as explained for Fig. 10, however, this time all the lenslet and HDCA LF images are considered and only the set of RAP values correspondent to QP 27 was used.

TABLE IV – HEVC-SLF-RA AVG. BD-RATE VS HEVC-PVS AND RAP FOR THE DIFFERENT CONTROL PARAMETER COMBINATIONS FOR LENSLET AND HDCA LFS

		Lenslet						**HDCA**					
		RPL size = 2			*RPL size = 4*			*RPL size = 2*			*RPL size = 4*		
NVPR	*MDL*	Avg. RAP	Max RAP	Avg. BD-RATE	Avg. RAP	Max RAP	Avg. BD-RATE	Avg. RAP	Max RAP	Avg. BD-RATE	Avg. RAP	Max RAP	Avg. BD-RATE
1	6	**0.24**	**0.49**	**-36.44 %**	***0.36***	***0.59***	***-44.95%****	**0.20**	**0.57**	**-34.69 %**	***0.29***	***0.67***	***-39.02 %****
	4	**0.24**	**0.47**	**-35.65 %**	**0.32**	**0.55**	**-42.15 %**	**0.20**	**0.57**	**-33.59 %**	**0.24**	**0.65**	**-35.33 %**
	3	**0.18**	**0.37**	**-17.59 %**	**0.23**	**0.42**	**-23.76 %**	**0.16**	**0.50**	**-23.90 %**	**0.19**	**0.51**	**-26.81 %**
	2	0.14	0.31	2.69 %	**0.16**	**0.32**	**-0.65 %**	**0.14**	**0.42**	**-13.72 %**	**0.16**	**0.43**	**-16.00 %**
2	6	**0.24**	**0.49**	**-36.44 %**	**0.35**	**0.57**	**-44.57 %**	**0.20**	**0.57**	**-34.69 %**	**0.28**	**0.67**	**-38.92 %**
	4	**0.24**	**0.47**	**-35.65 %**	**0.31**	**0.53**	**-41.80 %**	**0.20**	**0.57**	**-33.59 %**	**0.24**	**0.65**	**-35.34 %**
	3	**0.18**	**0.37**	**-17.59 %**	**0.22**	**0.41**	**-23.20 %**	**0.16**	**0.50**	**-23.90 %**	**0.19**	**0.51**	**-26.61 %**

	2	0.14	0.31	2.69 %	**0.15**	**0.32**	**-0.09 %**	**0.14**	**0.42**	**-13.72 %**	**0.16**	**0.43**	**-15.97 %**
4	6	**0.24**	**0.49**	**-36.44 %**	**0.34**	**0.56**	**-44.39 %**	**0.20**	**0.57**	**-34.69 %**	**0.28**	**0.67**	**-38.78 %**
	4	**0.24**	**0.47**	**-35.65 %**	**0.30**	**0.52**	**-41.61 %**	**0.20**	**0.57**	**-33.59 %**	**0.24**	**0.65**	**-35.22 %**
	3	**0.18**	**0.37**	**-17.59 %**	**0.22**	**0.41**	**-22.81 %**	**0.16**	**0.50**	**-23.90 %**	**0.19**	**0.51**	**-26.49 %**
	2	0.14	0.31	2.69 %	0.15	0.32	0.43 %	**0.14**	**0.42**	**-13.72 %**	**0.16**	**0.43**	**-15.75 %**
5	6	**0.10**	**0.22**	**-0.01 %**	**0.13**	**0.23**	**-6.81 %**	0.07	0.23	37.20 %	0.09	0.25	32.25 %
	4	0.10	0.22	2.30 %	**0.11**	**0.23**	**-0.85 %**	0.07	0.23	42.64 %	0.08	0.24	41.66 %
	3	0.08	0.19	25.05 %	0.09	0.20	21.50 %	0.06	0.19	59.17 %	0.06	0.20	58.19 %
	2	0.06	0.16	52.78 %	0.06	0.16	52.78 %	0.06	0.16	68.70 %	0.06	0.16	68.70 %
9	6	0.07	0.17	15.51 %	0.09	0.17	9.84 %	0.05	0.17	78.31 %	0.06	0.19	74.89 %
	4	0.07	0.17	20.10 %	0.08	0.17	16.71 %	0.05	0.17	84.18 %	0.05	0.18	83.30 %
	3	0.06	0.15	42.71 %	0.06	0.15	40.89 %	0.04	0.15	98.00 %	0.05	0.16	96.88 %
	2	0.05	0.14	62.13 %	0.05	0.14	62.13 %	0.04	0.12	107.64 %	0.04	0.12	107.64 %

* The configuration $RPL\ size = 4$, $NVPR = 1$, and $MDL = 6$, corresponds to HEVC-SLF

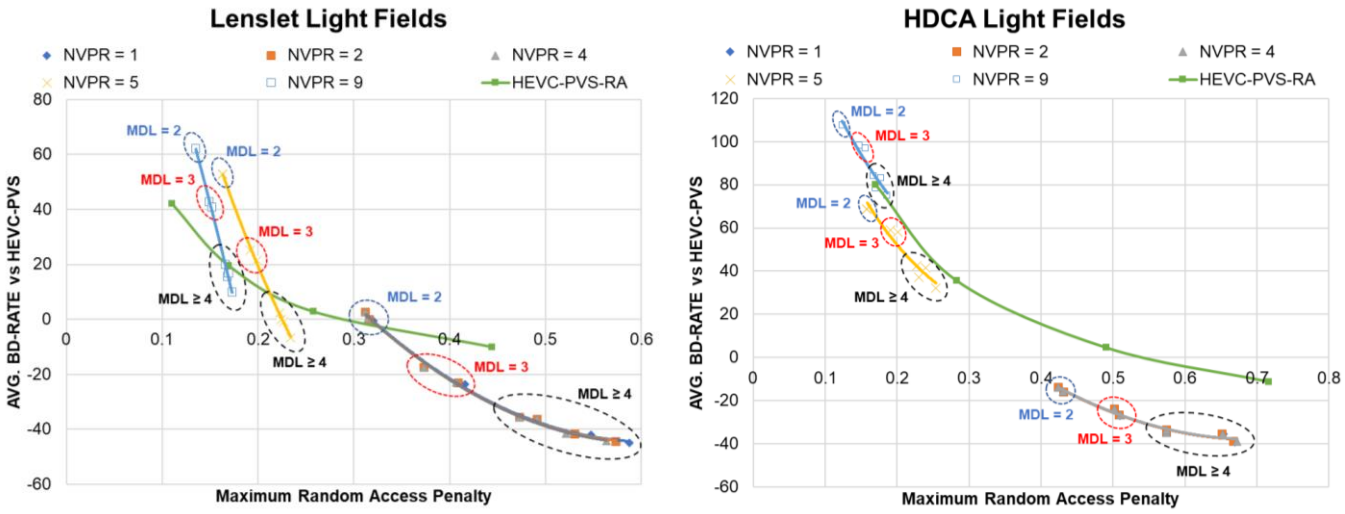


Fig. 11. Average BD-RATE savings vs Maximum RAP for QP 27: 12 lenslet LFs (left) and 3 HDCA LFs (right), for the proposed HEVC-SLF-RA using several combinations of control parameters as well as for HEVC-PVS-RA.

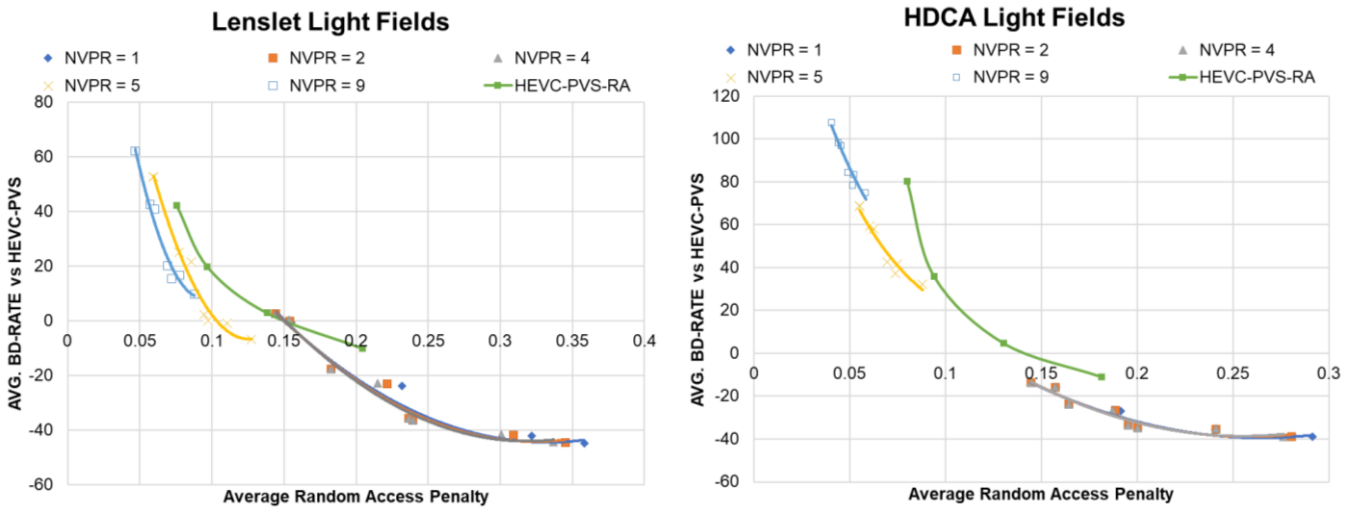


Fig. 12. Average BD-RATE savings vs Average RAP for QP 27: 12 lenslet LFs (left) and 3 HDCA LFs (right), for the proposed HEVC-SLF-RA using several combinations of control parameters as well as for HEVC-PVS-RA.

By analyzing the bitrate variations and the corresponding maximum RAP, it is possible to draw some conclusions in terms of the influence of each control parameter, as elaborated in the following subsections, which also include a comparison with HEVC-PVS-RA and a set of suggested practical usage profiles for both lenslet and HDCA LFs.

1) Reference picture list size

From the results in Table IV it is possible to observe that the proposed RA control parameters allow for a large number of tradeoff options balancing coding efficiency and viewpoint random access capabilities. For some configurations, namely for $RPL\ size = 2$, the results are identical when using $NVPR = 1$, $NVPR = 2$ or $NVPR = 4$. This occurs because inter-viewpoint dependencies are only different after the third reference picture, which is never used when $RPL\ size = 2$, independently of NVPR. From Table IV we may see that when using a lower $RPL\ size$ per viewpoint, i.e., 2 instead of 4, a reduction between 0% and 10% of bitrate savings is observed, but a reduction between 0 and 0.1 of the RAP is also achieved. In the particular example of the first line of Table IV, i.e., when we compare the results from the configuration $[RPL\ size = 4, NVPR = 1, MDL = 6]$ with the configuration $[RPL\ size = 2, NVPR = 1, MDL = 6]$, we can observe a decrease in bitrate savings of 8.5% (from 44.95% to 36.44%) but a better value of RAP, which decreases from 0.59 to 0.49.

2) Number of viewpoint regions

As previously mentioned, Fig. 11 represents graphically the results in Table IV. The maximum RAP is shown in the horizontal axis and the average BD-RATE over HEVC-PVS is shown in the vertical axis. These plots facilitate the experimental results assessment as they allow both random access capabilities and coding efficiency to be visualized for every combination of the proposed control parameters. The NVPR options were plotted in Fig. 11 as several groups of points, signaled with different colors and markers, together with their corresponding estimated tendency curves. It is possible to observe a clear separation of three groups of points (and respective tendency curves), which correspond to $NVPR = 9$ (light blue), $NVPR = 5$ (yellow) and $NVPR = [1, 2, 4]$ (blue, orange and grey). This separation is justified by the type of regions, i.e. overlapped or non-overlapped, that are being used. When increasing the number of overlapping spatial regions, from $NVPR = 1$ to $NVPR = 4$, the variation in terms of both maximum RAP and bitrate savings is lower than when non-overlapping regions are used ($NVPR = 5$ to $NVPR = 9$). These control parameter allows to perform two types of adjustments: 1) a fine adjustment provided by changing the number of overlapping regions, i.e., maintaining the number of Intra frames equal to 1; 2) a coarse adjustment, provided by changing the number of non-overlapping regions and, as a consequence, the number of Intra frames, i.e., 1, 5 and 9, respectively to NVPR of 1, 5 and 9. These conclusions are consistent for lenset and HDCA LF images, as well as when considering the maximum RAP and average RAP, as seen Fig. 11 and Fig. 12, respectively. However, it was observed by the bitrate savings that the use of non-overlapping regions has a higher impact in HDCA than in lenslet LF images.

3) Maximum dependency layer

When analyzing each tendency curve in Fig. 11 it is noticeable that the circles which correspond to the different MDL values are smaller when the MDL value is low, meaning that a low MDL value further restricts the inter-viewpoint dependencies. Additionally, lower MDL values create smaller ranges of values, for bitrate savings and RAP. For example, in the case of lenslet LFs (left graph in Fig. 11), the yellow tendency curve, defined by $NVPR = 5$, includes three circles that correspond to $MDL \geq 4$ (black), $MDL = 3$ (red) and $MDL = 2$ (light blue) which correspond to bitrate savings ranges of roughly [-10%, 5%], [20%, 25%] and 50%, respectively. This reduction in range of obtained values is due to the increasing restrictions applied when choosing the several reference viewpoints, e.g., when $MDL = 2$, only 9 viewpoints are available for reference. This reduction is consistent across the several tendency curves for both types of LFs.

4) HEVC-SLF-RA vs HEVC-PVS-RA

HEVC-PVS-RA was also tested in terms of both coding efficiency and random access capabilities for several IP values, ranging between 8 and 64. This parameter is used to control the GOP size and, therefore, affects the dependency between the several viewpoints. Table V shows the average and maximum RAP for the HEVC-PVS-RA codec across the different IP values. From the results in Table V it is possible to observe that a lower IP will provide HEVC-PVS-RA better viewpoint random access capabilities, however, at the cost of a lower coding efficiency. When using a IP of 32, the coding efficiency is slightly lower than that of HEVC-PVS (2.85% and 4.72% bitrate increase for lenslet and HDCA respectively), however, the maximum RAP is notably smaller (0.26 and 0.49 for lenslet and HDCA, respectively, where HEVC-PVS maximum RAP is 1).

TABLE V – HEVC-PVS-RA AVG. BD-RATE VS HEVC-PVS AND RAP FOR THE DIFFERENT INTRA PERIOD VALUES FOR LENSLET AND HDCA LFS

	Lenslet			**HDCA**		
Intra Period	Avg. RAP	Max RAP	Avg. BD-RATE	Avg. RAP	Max RAP	Avg. BD-RATE
64	**0.20**	**0.44**	**-10.06 %**	**0.18**	**0.72**	**-11.01 %**
32	0.14	0.26	2.85 %	0.13	0.49	4.72 %
16	0.10	0.17	19.71 %	0.09	0.28	35.90 %
8	0.08	0.11	42.30 %	0.08	0.17	80.33 %

When comparing both HEVC-SLF-RA and HEVC-PVS-RA in Fig. 11, the proposed HEVC-SLF-RA achieves better results if both coding efficiency and RAP are considered. For most combinations of control parameters, the proposed codec is able to achieve superior performance, i.e., higher bitrate savings for the same maximum RAP or a lower maximum RAP for the same coding efficiency, which is especially notorious for the HDCA LF images. However, there are a few combinations of control parameters that present inferior performance for HEVC-PVS-RA, notably for lenslet LFs when using $NVPR = 5$ or $NVPR = 9$, combined with low MDL values, e.g., $MDL = 2$ (light blue circle) and $MDL = 3$ (red circle). Nevertheless, for lenslet LF images, HEVC-PVS-RA is more efficient than HEVC-SLF-RA for maximum RAP values lower than 0.15. However, when analyzing the results in Fig. 12, where the average RAP is used instead of the maximum RAP, it is possible to see that the proposed HEVC-SLF-RA outperforms HEVC-PVS-RA in terms of both bitrate savings and RAP. In general, regardless of the type of RAP metric, the proposed codec is more advantageous due to its viewpoint scalability features, which are not available in HEVC-PVS-RA.

5) Suggested encoding profiles for lenslet and HDCA LFs

The previous sections showed a clear advantage of HEVC-SLF-RA for a very large number of configurations but in a practical scenario, the tradeoff between coding efficiency and random access capabilities depend on the application requirements. This section summarizes the results of the experimental assessment and proposes a list of suggested profiles based on four different tradeoff points. This list is shown in Table VI, which includes *a maximum coding efficiency* profile, achieving roughly 40% bitrate savings over HEVC-PVS; two *balanced profiles* biased towards coding efficiency and random access capabilities, achieving roughly 20% bitrate savings over HEVC-PVS and 0.25 maximum RAP, respectively; and, a *maximum random access profile* that values the RAP over the coding efficiency, achieving a maximum RAP of roughly 0.15. Although some profiles are less efficient than HEVC-PVS, the maximum RAP is notably superior, considering that HEVC-PVS maximum RAP is 1.

TABLE VI – PERFORMANCE ASSESSMENT OF THE SUGGESTED RANDOM ACCESS ENCODING PROFILES, FOR LENSLET AND HDCA LFS

	Lenslet						**HDCA**					
Encoding Profiles	RPL size	MDL	NVPR	Avg. RAP	Max RAP	BD-RATE*	RPL size	MDL	NVPR	Avg. RAP	Max RAP	BD-RATE*
Max. Eff.	4	6	1	0.36	0.59	-44.95%	4	6	1	0.29	0.67	-39.02%
Balanced High Eff.	4	3	4	0.22	0.41	-22.81%	2	3	4	0.16	0.50	-23.90%
Balanced High RA	4	6	5	0.13	0.23	-6.81%	4	6	5	0.09	0.25	32.25%
Max. RA	2	4	9	0.07	0.17	20.10%	2	2	5	0.06	0.16	68.70%

*vs HEVC-PVS

VII. CONCLUSIONS

In this paper, a new coding framework that supports viewpoint scalability and random access capabilities for LF content is proposed. Despite presenting a higher flexibility with the new functionalities, the coding efficiency is improved by using a new optimized RPS method that is able to adapt to any PVS scanning order. This technique is very important to accommodate the proposed viewpoint scalability structure, allowing to maintain the coding efficiency comparable to the non-scalable version. Additionally, the proposed control parameters allow to exploit a flexible set of encoding profiles, enabling a fine control of the viewpoint random access capabilities.

When compared to the state-of-the-art HEVC-PVS, the proposed HEVC-SLF codec achieves average bitrate

savings of approximately 45% and 39%, for lenslet and HDCA, respectively, which is 3% and 6% more efficient than its non-scalable version (HEVC-OPT). In comparison to the JPEG Pleno standard MuLE and WaSP, which are used as benchmarks for lenslet and HDCA LFs, respectively, the proposed HEVC-SLF is able to achieve average bitrate savings of 25%, and 45%, respectively.

To enable viewpoint random access capabilities, a scalable codec, HEVC-SLF-RA, was proposed, introducing a flexible set of random access profiles. Depending on the application, these profiles can be used to control the tradeoff between random access and coding efficiency. By acting on the control parameters, the maximum RAP of the proposed HEVC-SLF-RA in relation to HEVC-PVS (maximum RAP equal to 1) ranges from 0.17 to 0.59, respectively, for bit rating savings up from -20% to 45% for lenslet LF images. For HDCA LF images, the maximum RAP ranges from 0.16 to 0.67, allowing bit rating savings from -68% to 39%. The proposed codec was also compared to HEVC-PVS using the "Random Access" profile to evaluate its viewpoint random access capabilities. It was observed for both types of LFs that, for most combinations of the control parameters, the proposed HEVC-SLF-RA solution was able to achieve higher coding efficiency for the same maximum RAP, especially for maximum RAP values higher than 0.15.

ACKNOWLEDGMENT

The authors acknowledge the support of Fundação para a Ciência e Tecnologia, under the grant SFRH/BD/136953/2018, the projects UIDB/50008/2020, PlenoISLA POCI-01-0145-FEDER-028325 and PTDC/EEI-COM/7096/2020.

The authors would like to thank Mr. Pekka Astola for providing the WaSP software and Dr. Eduardo Silva and Dr. Carla Pagliari for providing the MuLE software as well as contributing with insightful discussions.

REFERENCES

[1] T. Georgiev and A. Lumsdaine, "Rich Image Capture with Plenoptic Cameras," in IEEE International Conference on Computational Photography, Cluj-Napoca, Romania, Aug. 2010, pp. 1–8.

[2] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, and J. C. J. Fernández, "Light field geometry of a standard plenoptic camera," Opt Express, vol. 22, no. 22, pp. 26659–26673, Nov. 2014, doi: 10.1364/OE.22.026659.

[3] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in IEEE International Conference on Computational Photography, San Francisco, CA, USA, Apr. 2009, pp. 1–8, doi: 10.1109/ICCPHOT.2009.5559008.

[4] D. G. Dansereau, O. Pizarro, and S. B. Williams, “Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras,” in 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, Jun. 2013, pp. 1027–1034, doi: 10.1109/CVPR.2013.137.

[5] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, “Advances in three-dimensional integral imaging: sensing, display, and applications,” Appl Opt, vol. 52, no. 4, pp. 546–560, Feb. 2013, doi: 10.1364/AO.52.000546.

[6] J. Arai, “Integral three-dimensional television (FTV Seminar),” Sapporo, Japan, ISO/IEC JTC1/SC29/WG11 MPEG2014/N14552, Sapporo, Japan, Jul. 2014.

[7] L. Toni, G. Cheung, and P. Frossard, “In-Network View Synthesis for Interactive Multiview Video Systems,” IEEE Trans. Multimed., vol. 18, no. 5, pp. 852–864, May 2016, doi: 10.1109/TMM.2016.2537207.

[8] L. Toni and P. Frossard, “Optimal Representations for Adaptive Streaming in Interactive Multiview Video Systems,” IEEE Trans. Multimed., vol. 19, no. 12, pp. 2775–2787, Dec. 2017, doi: 10.1109/TMM.2017.2713644.

[9] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, “A Free-Viewpoint Television System for Horizontal Virtual Navigation,” IEEE Trans. Multimed., vol. 20, no. 8, pp. 2182–2195, Aug. 2018, doi: 10.1109/TMM.2018.2790162.

[10] P. Ramanathan, M. Kalman, and B. Girod, “Rate-Distortion Optimized Interactive Light Field Streaming,” IEEE Trans. Multimed., vol. 9, no. 4, pp. 813–825, Jun. 2007, doi: 10.1109/TMM.2007.893350.

[11] C. Conti, L. D. Soares, and P. Nunes, “Light Field Coding with Field of View Scalability and Exemplar-Based Inter-Layer Prediction,” IEEE Trans. Multimed., vol. 20, no. 11, pp. 2905–2920, Nov. 2018, doi: 10.1109/TMM.2018.2825882.

[12] “JPEG PLENO Abstract and Executive Summary,” Sydney, ISO/IEC JTC 1/SC 29/WG1 N6922, Feb. 2015, Sydney, Australia.

[13] “MPEG-I Technical Report on Immersive Media,” Torino, Italy, ISO/IEC JTC1/SC29/WG11 N17069, Jul. 2017, Torino, Italy.

[14] R. J. S. Monteiro, P. J. L. Nunes, S. M. M. Faria, and N. M. M. Rodrigues, “Optimized Reference Picture Selection for Light Field Image Coding,” in 2019 27th European Signal Processing Conference (EUSIPCO), Sep. 2019, pp. 1–4.

[15] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, “Coding of Focused Plenoptic Contents by Displacement Intra Prediction,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 7, pp. 1308–1319, Jul. 2016, doi: 10.1109/TCSVT.2015.2450333.

[16] C. Perra, W. Song, and A. Liotta, “Effects of light field subsampling on the quality of experience in refocusing applications,” in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), May 2018, pp. 1–3, doi: 10.1109/QoMEX.2018.8463393.

[17] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, “Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 1, pp. 20–34, Jan. 2016, doi: 10.1109/TCSVT.2015.2461951.

[18] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.

[19] M. B. de Carvalho et al., “A 4D DCT-Based Lenslet Light Field Codec,” in 2018 25th IEEE International Conference on Image Processing, Athens, Greece, Oct. 2018, pp. 435–439, doi: 10.1109/ICIP.2018.8451684.

[20] P. Astola and I. Tabus, “WaSP: Hierarchical Warping, Merging, and Sparse Prediction for Light Field Image Compression,” in 2018 7th European Workshop on Visual Information Processing (EUVIP), Nov. 2018, pp. 1–6, doi: 10.1109/EUVIP.2018.8611756.

[21] C. Conti, L. D. Soares, and P. Nunes, “Dense Light Field Coding: A Survey,” IEEE Access, vol. 8, pp. 49244 - 49284, Mar. 2020, doi: 10.1109/ACCESS.2020.2977767.

[22] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, “Data formats for high efficiency coding of Lytro-Illum light fields,” in International Conference on Image Processing Theory, Tools and Applications, Orleans, France, Nov. 2015, pp. 494–497, doi: 10.1109/IPTA.2015.7367195.

[23] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, “Lenselet image compression scheme based on subaperture images streaming,” in IEEE International Conference on Image Processing, Quebec, Canada, Sep. 2015, pp. 4733–4737, doi: 10.1109/ICIP.2015.7351705.

[24] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, “Pseudo-sequence-based light field image compression,” in IEEE International Conference on Multimedia Expo Workshops, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574674.

[25] C. Jia et al., “Optimized inter-view prediction based light field image compression with adaptive reconstruction,” in 2017 IEEE International Conference on Image Processing, Beijing, China, Sep. 2017, pp. 4572–4576, doi: 10.1109/ICIP.2017.8297148.

[26] W. Ahmad, R. Olsson, and M. Sjostrom, “Towards a Generic Compression Solution for Densely and Sparsely Sampled Light Field Data,” in 2018 25th IEEE International Conference on Image Processing, Athens, Greece, Oct. 2018, pp. 654–658, doi: 10.1109/ICIP.2018.8451051.

[27] W. Ahmad, R. Olsson, and M. Sjöström, “Interpreting plenoptic images as multi-view sequences for improved compression,” in 2017 IEEE International Conference on Image Processing, Beijing, China, Sep. 2017, pp. 4557–4561, doi: 10.1109/ICIP.2017.8297145.

[28] C. Conti, L. D. Soares, and P. Nunes, “HEVC-based 3D holoscopic video coding using self-similarity compensated prediction,” Signal Process. Image Commun., vol. 42, pp. 59–78, Mar. 2016, doi: 10.1016/j.image.2016.01.008.

[29] C. Conti, P. Nunes, and L. D. Soares, “HEVC-based light field image coding with bi-predicted self-similarity compensation,” in IEEE International Conference on Multimedia Expo Workshops, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574667.

[30] C. Conti, P. Nunes, and L. D. Soares, “Light field image coding with jointly estimated self-similarity bi-prediction,” Signal Process. Image Commun., vol. 60, pp. 144–159, Feb. 2018, doi: https://doi.org/10.1016/j.image.2017.10.006.

[31] Y. Li, R. Olsson, and M. Sjöström, “Compression of unfocused plenoptic images using a displacement prediction,” in IEEE International Conference on Multimedia Expo Workshops, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574673.

[32] R. Monteiro et al., “Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction,” in IEEE International Conference on Multimedia Expo Workshops, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574670.

[33] R. J. Monteiro, P. Nunes, N. Rodrigues, and S. M. M. de Faria, “Light Field Image Coding using High Order Intra Block Prediction,” IEEE J. Sel. Top. Signal Process., vol. 11, no. 7, pp. 1120–1131, Oct. 2017, doi: 10.1109/JSTSP.2017.2721358.

[34] R. J. S. Monteiro, P. J. L. Nunes, S. M. M. Faria, and N. M. M. Rodrigues, “Light Field Image Coding using High Order Prediction Training,” in 2018 26th European Signal Processing Conference (EUSIPCO), Sep. 2018, pp. 1845–1849, doi: 10.23919/EUSIPCO.2018.8553150.

[35] L. F. R. Lucas et al., “Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC,” in European Signal Processing Conference, Lisbon, Portugal, Sep. 2014, pp. 11–15.

[36] J. Hou, J. Chen, and L. Chau, “Light Field Image Compression Based on Bi-Level View Compensation with Rate-Distortion Optimization,” IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 2, pp. 517 - 530, 2018, doi: 10.1109/TCSVT.2018.2802943.

[37] S. Zhao and Z. Chen, “Light field image coding via linear approximation prior,” in 2017 IEEE International Conference on Image Processing, Beijing, China, Sep. 2017, pp. 4562–4566, doi: 10.1109/ICIP.2017.8297146.

[38] J. Chen, J. Hou, and L. P. Chau, “Light Field Compression With Disparity-Guided Sparse Coding Based on Structural Key Views,” IEEE Trans. Image Process., vol. 27, no. 1, pp. 314–324, Jan. 2018, doi: 10.1109/TIP.2017.2750413.

[39] X. Jiang, M. L. Pendu, and C. Guillemot, “Light field compression using depth image based view synthesis,” in 2017 IEEE International Conference on Multimedia Expo Workshops, Hong Kong, China, Jul. 2017, pp. 19–24, doi: 10.1109/ICMEW.2017.8026313.

[40] M. Rizkallah, X. Su, T. Maugey, and C. Guillemot, “Graph-based Transforms for Predictive Light Field Compression based on Super-Pixels,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Canada, Apr. 2018, pp. 1718–1722, doi: 10.1109/ICASSP.2018.8462288.

[41] I. Viola, H. P. Maretic, P. Frossard, and T. Ebrahimi, “A graph learning approach for light field image compression,” in Applications of Digital Image Processing XLI, San Diego, CA, USA, 2018, vol. 10752, pp. 126 – 137, doi: 10.1117/12.2322827.

[42] Y. Li, M. Sjöström, and R. Olsson, “Coding of plenoptic images by using a sparse set and disparities,” in IEEE International Conference on Multimedia and Expo, Jun. 2015, Turin, Italy, pp. 1–6, doi: 10.1109/ICME.2015.7177510.

[43] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, “Scalable Coding of Plenoptic Images by Using a Sparse Set and Disparities,” IEEE Trans. Image Process., vol. 25, no. 1, pp. 80–91, Jan. 2016, doi: 10.1109/TIP.2015.2498406.

[44] C. Conti, P. Nunes, and L. D. Soares, “Inter-Layer Prediction Scheme for Scalable 3-D Holoscopic Video Coding,” IEEE Signal Process. Lett., vol. 20, no. 8, pp. 819–822, Aug. 2013, doi: 10.1109/LSP.2013.2267234.

[45] A. Aaron, P. Ramanathan, and B. Girod, “Wyner-Ziv coding of light fields for random access,” in IEEE 6th Workshop on Multimedia Signal Processing, 2004., Sep. 2004, pp. 323–326, doi: 10.1109/MMSP.2004.1436558.

[46] P. Ramanathan and B. Girod, “Random access for compressed light fields using multiple representations,” in IEEE 6th Workshop on Multimedia Signal Processing, 2004., Sep. 2004, pp. 383–386, doi: 10.1109/MMSP.2004.1436573.

[47] H. Amirpour, A. Pinheiro, M. Pereira, F. Lopes, and M. Ghanbari, “Light Field Image Compression with Random Access,” in 2019 Data Compression Conference (DCC), Mar. 2019, pp. 553–553, doi: 10.1109/DCC.2019.00065.

[48] N. Mehajabin, S. R. Luo, H. W. Yu, J. Khoury, J. Kaur, and M. T. Pourazad, “An Efficient Random Access Light Field Video Compression Utilizing Diagonal Inter-View Prediction,” in 2019 IEEE

International Conference on Image Processing (ICIP), Sep. 2019, pp. 3567–3570, doi: 10.1109/ICIP.2019.8803668.

[49] P. Gomes and L. A. da S. Cruz, “Pseudo-Sequence Light Field Image Scalable Encoding with Improved Random Access,” in 2019 8th European Workshop on Visual Information Processing (EUVIP), Oct. 2019, pp. 16–21, doi: 10.1109/EUVIP47703.2019.8946268.

[50] E. Upenik, I. Viola, and T. Ebrahimi, “A Rendering Solution to Display Light Field in Virtual Reality,” in 2018 26th European Signal Processing Conference (EUSIPCO), Sep. 2018, pp. 246–250, doi: 10.23919/EUSIPCO.2018.8553424.

[51] “JPEG Pleno Light Field Coding Common Test Conditions,” Geneva, Switzerland, ISO /IEC JTC 1/SC 29 /WG 1 N83029, Geneva, Switzerland, Mar. 2019.

[52] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, “Saliency Detection on Light Field,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 8, pp. 1605–1616, Aug. 2017, doi: 10.1109/TPAMI.2016.2610425.

[53] EPFL Light-field image dataset. Accessed on: April, 2020 [Online]. Available: http://mmspg.epfl.ch/EPFL-light-field-image-dataset.

[54] Light Field Toolbox v0.4. Accessed on: April, 2020 [Online]. Available: http://dgd.vision/Tools/LFToolbox/.