IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 3, 2020, accepted June 20, 2020, date of publication June 24, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004625

Light Field Image Coding Based on Hybrid Data Representation

RICARDO J. S. MONTEIRO[1,2], (Member, IEEE), NUNO M. M. RODRIGUES[3,4], (Member, IEEE), SÉRGIO M. M. FARIA[3,4], (Senior Member, IEEE), AND PAULO J. L. NUNES[1,2], (Member, IEEE)

[1]Instituto de Telecomunicações, 1049-001 Lisbon, Portugal
[2]Instituto Universitário de Lisboa (ISCTE-IUL), 1649-026 Lisbon, Portugal
[3]Instituto de Telecomunicações, 2411-901 Leiria, Portugal
[4]Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, 2411-901 Leiria, Portugal

Corresponding author: Ricardo J. S. Monteiro (ricardo.monteiro@lx.it.pt)

This work was supported by the FCT/MCTES through the National Funds and when applicable co-funded EU Funds under Project UIDB/EEA/50008/2020. The work of Ricardo J. S. Monteiro was supported by the Fundação para a Ciência e Tecnologia (FCT) under Grant SFRH/BD/136953/2018.

ABSTRACT This paper proposes a novel efficient light field coding approach based on a hybrid data representation. Current state-of-the-art light field coding solutions either operate on micro-images or sub-aperture images. Consequently, the intrinsic redundancy that exists in light field images is not fully exploited, as is demonstrated. This novel hybrid data representation approach allows to simultaneously exploit four types of redundancies: i) sub-aperture image intra spatial redundancy, ii) sub-aperture image inter-view redundancy, iii) intra-micro-image redundancy, and iv) inter-micro-image redundancy between neighboring micro-images. The proposed light field coding solution allows flexibility for several types of baselines, by adaptively exploiting the most predominant type of redundancy on a coding block basis. To demonstrate the efficiency of using a hybrid representation, this paper proposes a set of efficient pixel prediction methods combined with a pseudo-video sequence coding approach, based on the HEVC standard. Experimental results show consistent average bitrate savings when the proposed codec is compared to relevant state-of-the-art benchmarks. For lenslet light field content, the proposed coding algorithm outperforms the HEVC-based pseudo-video sequence coding benchmark by an average bitrate savings of 23%. It is shown for the same light field content that the proposed solution outperforms JPEG Pleno verification models MuLE and WaSP, as these codecs are only able to achieve 11% and −14% bitrate savings over the same HEVC-based benchmark, respectively. The performance of the proposed coding approach is also validated for light fields with wider baselines, captured with high-density camera arrays, being able to outperform both the HEVC-based benchmark, as well as MuLE and WaSP.

INDEX TERMS Light field representation, light field image coding, HEVC, pseudo-video sequence, spatial pixel prediction, least squares prediction.

I. INTRODUCTION

The light field (LF) imaging technology allows to jointly capture the scene radiance and angular information using single-tier lenslet LF cameras, i.e., with narrow baseline, or by using, for example, a high-density camera array (HDCA), i.e., with a wider baseline. A lenslet LF camera is composed of the standard main lens and sensor, common to 2D cameras, with the addition of a third element: the microlens array (MLA) [1]. The MLA allows the LF camera to capture both spatial and angular information about the light reaching the sensor [2]. Depending on the LF capturing device, different degrees of freedom are available in terms of both spatial and angular resolution [3]. Nonetheless, the captured LF information can convey 3D information about the scene, instead of representing just a single 2D perspective.

By capturing the angular information, several a posteriori image processing manipulations may be performed, such as changing the perspective and refocusing after taking the picture [1]. The richer content capturing technology based on LF also has applications in image recognition, medical imaging [4] and 3D television [5], since by rendering

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

several views from different perspectives, 2D, 3D and multiview (MV) signals can be created. This imaging technology allows for interactive media applications, such as interactive MV video [6], [7], free viewpoint video streaming [8], and interactive streaming of light field images captured by HDCAs [9] or lenslet LF cameras [10].

The LF technology has recently attracted many research groups and standardization bodies such as JPEG and MPEG, not only because of its new appealing features and the necessity to normalize the LF data representation, but also due to the very large amount of data generated by these LF capturing devices, that demands efficient compression techniques. Moreover, coding solutions that allow for compatibility with legacy displays, i.e., view scalable, and that allow for viewpoint random access to facilitate content navigation, also need to be developed. Thus, these groups are currently developing coding standards for emerging imaging technologies like LF, point cloud, holographic and 360° video content, whose activities are known as JPEG Pleno [11] and MPEG-I [12].

State-of-the-art 2D image and video coding algorithms struggle to cope with the new features and the large amount of data generated by lenslet LFs, lacking in terms of coding efficiency. One of the major limitations of these algorithms is the inefficient exploitation of the intra and inter micro-image (MI) redundancy exhibited in LF content, which corresponds to the redundancy within each MI and across neighboring MIs, respectively. Instead of straightforwardly applying state-of-the-art image and video codecs, three more efficient alternative approaches are commonly used to encode LF images [13]:

1) apply pre- and post-processing tools to convert the LF image into a so-called pseudo-video sequence (PVS), and encode it using a standard video codec;

2) add novel prediction tools to an image codec that are able to exploit the MI redundancies;

3) develop alternative coding approaches, specifically designed for LF images.

Some techniques have been proposed for the three mentioned coding approaches, however the common denominator between most of the current contributions in the literature consists on exploiting only one specific type of LF data representation. Most techniques either rely on MIs or sub-aperture images (SAIs), i.e., views generated by extracting at least one pixel in a fixed position from each MI and organizing them into a matrix. Exclusive MI-based or SAI-based techniques limit the amount of redundancy that the LF image coding can exploit, thus limiting its overall efficiency. Additionally, most of the coding approaches proposed in the literature are specifically tailored to encode either narrow or wide baseline LFs, limiting their practical application.

This paper proposes a hybrid LF data representation, i.e., that uses both the SAI and MI representations, enabling a more exhaustive exploitation of the inherent LF redundancy and improving the LF compression efficiency. The use of this hybrid approach enables the exploitation of four main types of redundancy: i) intra spatial redundancy within each SAI, ii) inter-view redundancy between SAIs, iii) intra-MI redundancy within each MI, and iv) inter-MI between neighboring MIs. The efficiency of the proposed hybrid representation is demonstrated for both narrow and wide baseline LF images, by incorporating this hybrid representation in a standard HEVC PVS codec [14], where the spatial redundancy within each SAI can be exploited by standard intra coding tools, and the inter-view redundancy between SAIs and the intra- and inter-MI redundancies are exploited by using a new hybrid reference picture list (HRPL). The proposed HRPL, allows the already encoded LF information to be stored simultaneously in a SAI- and a MI-based manner, which allows all types of redundancy to be exploited, by applying different prediction modes adaptively, using rate-distortion (RD) optimization.

The proposed codec uses an optimized set of pixelwise prediction methods such as: DC predictor, median edge detector (MED) [15], gradient adjusted predictor (GAP) [16] and accurate gradient selective prediction (AGSP) [17]. Additionally, new prediction methods based on least squares prediction (LSP) [18] are proposed to further improve the coding efficiency.

The first major contribution of this paper is the introduction of a novel hybrid representation coding approach, based on a HRPL, as previously explained. The second major contribution is the adaptation of the DC, MED, GAP, AGSP and LSP prediction methods to exploit the intra- and inter-MI redundancy using the MI representation domain, within the HRPL. Finally, the third major contribution is the validation of the proposed hybrid representation solution, against current relevant benchmarks, for a vast number of coding scenarios, which include different baseline types (i.e., narrow and wide), LF representation types (i.e., SAI and MI), and color formats (i.e., YUV 4:4:4 10 bit and YUV 4:2:0 8 bit).

The remainder of this paper is organized as follows: Section II reviews the state-of-the-art on LF coding approaches based on MIs and SAIs; Section III presents the proposed hybrid LF data representation; Section IV describes the new intra-MI prediction modes proposed in this paper; Section V presents the new inter-MI prediction modes proposed in this paper; Section VI evaluates the performance of the proposed LF coding solution against the most relevant state-of-the-art solutions; and, finally, Section VII concludes the paper.

II. RELATED WORK ON LIGHT FIELD CODING

This section briefly reviews the state-of-the-art methods for lenslet LF image encoding (a more thorough review can be found in [13]). The available schemes in the literature rely generally on one of two LF data representations: MI- or SAI-based approaches. MI-based approaches usually use state-of-the-art image codecs with additional prediction tools, specifically tailored for LF images. SAI-based approaches commonly rely on interpreting the LF image as a HDCA signal and encoding the LF as a PVS or as a MV sequence.

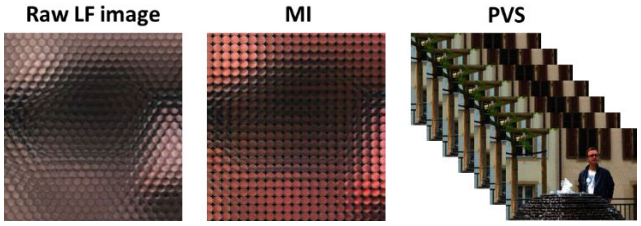

FIGURE 1. Example of a lenslet LF image using the raw LF image, MI and PVS image representations.

A. MAIN LF DATA REPRESENTATIONS

When a LF is captured using a lenslet LF camera, the raw LF image is naturally organized in MIs. This raw LF image can be converted to SAIs using a reversible or irreversible process, resulting in different organization formats for the SAIs [19]. The adopted LF representation type is critical since it directly affects the available coding options and tools. The LF representation types shown in Fig. 1 include:

- **MI-based representation**– A single frame with all the concatenated MIs that represent either the raw LF image, or a variation of the 4D LF [20]. The raw LF image is represented by variable-size MIs that can be organized in different grid styles, e.g., hexagonal or squared. The 4D LF variation is represented by squared, equal-size MIs in a squared grid;
- **SAI-based representation**– The generated SAIs are concatenated in a single frame or organized into a PVS. In this latter case, each SAI corresponds to a frame of a sequence of frames organized according to a certain scanning strategy, e.g., raster or spiral.

The existing coding solutions rely usually on the MI or the SAI representation. The MI representation facilitates a more efficient exploitation of intra- and inter-MI redundancies, while the SAI representation facilitates a more efficient exploitation of intra- and inter-SAI redundancies.

B. MI-BASED RELATED WORK

Several methods to exploit the inter-MI redundancy, also known as non-local spatial redundancy, were proposed in the literature [13]. Most approaches are based on the use of additional coding tools for state-of-the-art video coding standards, like HEVC. When using the MI-based LF representation, the non-local spatial redundancy is normally much higher than the typical image spatial redundancy, therefore, most methods consist of block-matching algorithms that try to exploit the inter-MI similarity. These algorithms can have different degrees of freedom and may use one or multiple references [21]–[28]. In [21], a self-similarity (SS) compensated prediction is proposed that takes advantage of the flexible partition patterns used by HEVC. The authors in [25] extended this approach by developing a multi-hypothesis coding method using up to two hypotheses for prediction in spatial and time domain. The approaches based on SS can be considered low order prediction (LOP) approaches because they are limited to translations, i.e., two degrees of freedom (DoF) transforms. LOP prediction methods have a limited ability to describe perspective changes between adjacent MIs. These perspective changes require geometric transformations (GT) with more (up to 8) DoF. A method that uses a high order prediction approach was added to a HEVC framework in [27]. Additionally, in [28], an alternative non-local spatial prediction method has been investigated, relying on a prediction mode based on locally linear embedding (LLE) integrated in HEVC. This prediction mode can adaptively use a different number of references for each block, varying from one, up to eight hypotheses in the spatial domain.

The authors in [29] proposed an alternative search algorithm with a reduced search area, allowing only horizontal and vertical directions, to find the N nearest neighbor templates in the causal area. The normalized cross correlation is used to assess the reliability of the obtained templates. The prediction from the N templates is modeled as a non-linear gaussian process and gaussian process regression is used for estimating the prediction block. More recently in order to improve de prediction accuracy for non-homogenous textures and to reduce the computational complexity in this work, in [30], the authors proposed to apply a classification method that can segment the non-homogeneous texture areas improving the prediction accuracy. Moreover, the computational complexity is improved by using different prediction modes for each specific area of the lenslet LF image, i.e., content-based prediction. Coding efficiency is comparable to high order prediction method described in [27] however, no comparisons were performed against SAI-based related work.

In [31], the authors explore scalability features for LF, proposing a two-layer LF coding approach for the focused LF camera model. The chosen LF representation uses a first layer which consists of a sparse set of MIs and associated disparity maps. Based on these data, a reference prediction LF image is obtained through disparity-based interpolation and inpainting. This reconstructed LF image is then used as a reference to encode the original LF image (second layer), by encoding the prediction residue. This approach was later extended [32] with a third layer of scalability and the use of lossy encoded disparity maps, in contrast with the lossless transmission of the disparity maps used in the first approach.

Alternatively to the block-matching approaches, LF coding schemes can also rely on a spatial transform, e.g., the discrete cosine transform (DCT) [33], [34]. In [34], a 3D-DCT is applied to a stack of MIs, to exploit the existent redundancy between the several MIs and within the same MI.

C. SAI-BASED RELATED WORK

Most state-of-the-art video codecs rely on motion estimation tools, like, for example, block-matching algorithms, to exploit the temporal redundancy of video data [13]. As a similar redundancy exists between SAIs, these tools can also

be used to exploit the inter-view redundancy. To this end, several scanning strategies [14], [35]–[37] have been proposed to transform SAIs into a PVS which is then encoded as a regular video sequence. Alternatively, in [27] and [28] SAIs are interpreted as a HDCA signal and encoded with MV-HEVC where two-dimensional weighted prediction and rate allocation is available. In [40], a MV-HEVC based coding solution was proposed, that allows diagonal viewpoint prediction instead of exclusively the 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. More recently, in [41], this method was improved by using a Structural Similarity Index (SSIM) assisted approach to determine the intra frame selection and prediction structure, allowing for competitive random access capabilities and improved coding efficiency.

Narrow baseline LFs, e.g., when the LF is captured by a hand-held camera, lead to low disparity between SAIs. Consequently, several authors have proposed to only encode and transmit some SAIs, normally referred to as structural key views (SKVs), and then transmitting additional information in the bitstream to the decoder to generate the remaining non-SKVs [42]–[48]. These approaches are normally, structurally similar, but the type of additional information varies. In [42], the non-SKVs are generated using a convolutional neural network (CNN) based on an angular super resolution algorithm. In [43], the coefficients are generated through a linear approximation, which are used to generate the non-SKVs as a weighted sum of the SKVs. In [44], non-SKVs are generated using approximated disparity maps that are transmitted to the decoder. In [45], the non-SKVs are generated using depth-image-based rendering (DIBR). In [46], a graph-based transform derived from a coherent super-pixel over-segmentation of the several views is used to encode non-SKVs. In [47], the non-SKVs are encoded using a graph learning approach which estimates the disparity among the views composing the LF. Finally, in [48], the non-SKVs are generated using a shearlet-transform-based prediction scheme which is shown to be efficient when reconstructing densely sampled LFs under low bitrates. Although these approaches are capable of high coding efficiency, their performance is heavily dependent on the SKVs selection.

As in MI-based methods, transforms, such as a three-dimensional discrete wavelet transform (3D-DWT), can also be used to exploit the LF redundancy [49]. In this case, the LF image is decomposed into SAIs, and a 3D-DWT is applied to the stack of SAIs. The lower frequency bands are transformed using a two-dimensional discrete wavelet transform (2D-DWT), while the remaining higher frequency coefficients are simply quantized and arithmetically encoded.

Alternatively, in [50], the authors propose a hierarchical approach to LF image coding which is based on warping, merging, and sparse prediction. The reference viewpoints are warped to the location of the current viewpoint; the warped reference viewpoints merged using one optimal least squares merger; finally, the overall merged image to the original viewpoint is adjusted using a sparse predictor.

D. OTHER RELATED WORK

Alternatively to the MI- and SAI-based techniques, the authors in [51] proposed to exploit the full 4D redundancy of the LF image. The proposed codec works 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 then grouped using hexadeca-trees generating a stream. Finally, the generated stream is encoded using adaptive arithmetic coding. In [52], the LF image is also encoded as a continuous representation of the 4D LF function, which is modeled as a space-continuous Gaussian Mixture Model. This compact model considers different regions of the scene, their edges, and their evolution along the spatial and disparity dimensions. This work was more recently improved in [53] with faster and more robust modeling using minibatches, however the coding efficiency is dependent on the data dimensionality. Consequently, this work is outperformed by a multiview-based coding method when encoding LF images, i.e., 4D data, however it can outperform the same approach for LF videos, i.e., 5D data. The authors in [54], represent the 4D LF as weighted binary images. Several binary images and corresponding weight values are to be chosen to optimally approximate the LF. This approach allows for competitive results against PVS-based approaches using HEVC.

III. PROPOSED HYBRID LIGHT FIELD DATA REPRESENTATION

This section presents the main contribution of this work. The new hybrid LF data representation is explained, as well as a new coding method that integrates a set of prediction modes in a HEVC-PVS LF coding scheme.

A. COMBINING LF DATA REPRESENTATIONS

As mentioned in the previous section, the lenslet LF data is typically expressed through MIs or SAIs. Converting the LF raw image into an equivalent 4D LF image representation (e.g., by means of the LF toolbox [55]), allows the use of both representations interchangeably. The 4D LF data representation organizes the LF image using four dimensions as $LF(h, v, x, y)$, containing a stack of SAIs that is generated from the raw LF image. The first two dimensions index the location, (h, v), of the SAI in the LF, using horizontal and vertical coordinates, and the remaining two dimensions index the spatial position, (x, y), of a pixel within each SAI. The main feature of the 4D LF data representation is that it allows the conversion between MI- and SAI-based data representations to become seamless and reversible.

Once the LF image is converted to the 4D LF representation, similarly to the raw LF image, it can be sampled using MI- or SAI-based data representations. Equivalent data representations based on 4D LF are referred to as MI and

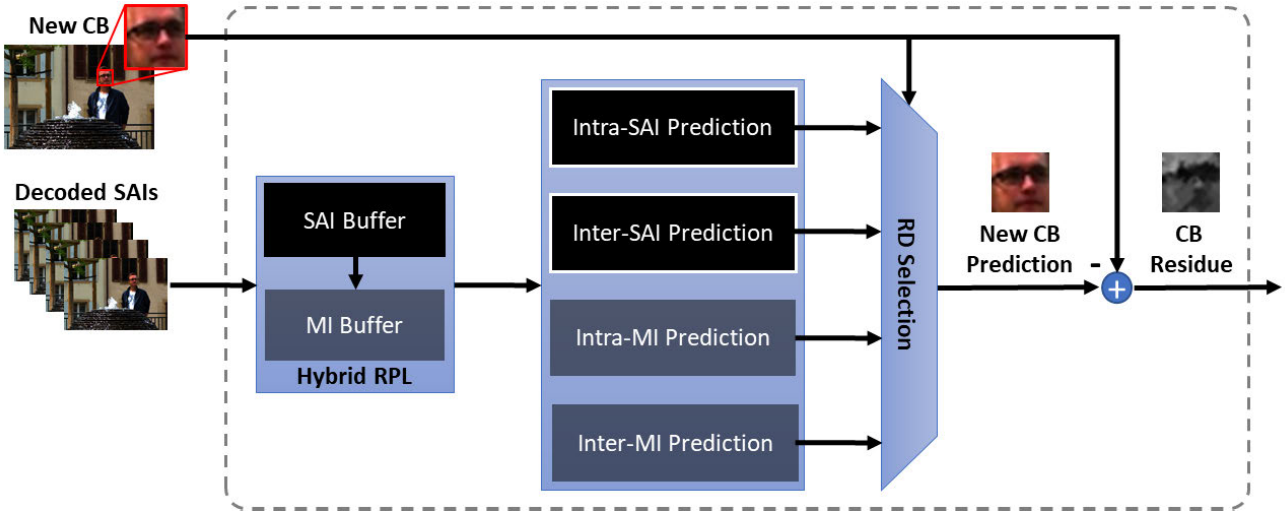

FIGURE 2. Proposed LF prediction module being applied to a new coding block (CB).

PVS. The PVS is generated by applying a scanning strategy to the generated SAIs.

Each of the LF data representations has its own advantages:

- **MI-based representation**– Allows for a more efficient exploitation of the spatial and inter-MI redundancy;
- **PVS-based representation**– Allows for more efficient exploitation of the spatial and inter-view redundancy within each SAI.

Because traditional coding solutions usually rely in a single representation (MI or PVS), they are unable to take full advantage of the various types of redundancy that exist in LF data. This limitation is circumvented in this work by using a hybrid LF data representation.

B. PROPOSED HYBRID DATA REPRESENTATION

The hybrid LF data representation proposed in this paper uses a combination of the PVS and MI representations, taking advantage of the seamless and reversible conversion between PVS and MI representations, enabling, this way, the use of more prediction modes in the compression of LF data. Fig. 2 shows the proposed LF prediction module, which integrates the proposed hybrid LF data representation in a HEVC-like encoder. The current SAI is partitioned into coding blocks (CB), which are then passed as input to the prediction module. This prediction module allows the creation of a prediction block for each new input CB that minimizes the RD cost (New CB prediction). The resultant CB residue, i.e., the difference between the generated prediction block and the new input CB, follows the regular HEVC-like processing chain that includes being transformed, quantized and entropy encoded together with signaling data.

From Fig. 2 it is possible to see the use of two decoded picture buffers: i) the SAI buffer, which is the standard HEVC decoded picture buffer, and ii) the new MI buffer that stores the full LF image using the MI representation, which is gradually updated from the decoded SAIs. The combination of both picture buffers defines the HRPL.

The use of the HRPL enables four prediction types (as shown in Fig. 2) that exploit different types of redundancy available in each data representation model:

- **Intra-SAI prediction** – Corresponds to the intra-picture prediction modes, i.e., DC, Planar, and Directional modes in HEVC, which are used to exploit the spatial redundancy of SAIs;
- **Inter-SAI prediction** – Corresponds to the inter-picture prediction modes, i.e., Motion Compensation, Merge/Skip modes in HEVC, which are used to exploit the inter-view redundancy between SAIs. These prediction modes make use of the SAI decoded picture buffer;
- **Intra-MI prediction** – New prediction modes that exploit the intra-MI redundancy, by using the MI decoded picture buffer. Such intra-MI prediction modes are explained in more detail in Section IV;
- **Inter-MI prediction** – New prediction modes that exploit the inter-MI redundancy, by using the MI decoded picture buffer. Such inter-MI prediction modes are explained in more detail in Section V.

1) GENERATION OF THE MI DECODED PICTURE BUFFER

The proposed LF coding solution, both input and output data use the PVS data representation. Therefore, the full LF image using the MI data representation must be generated from the decoded SAIs, as show in Fig. 2, by using the correspondence between the PVS and MI representations. The pixel position on the MI image, (i, j), can be defined as a function of the

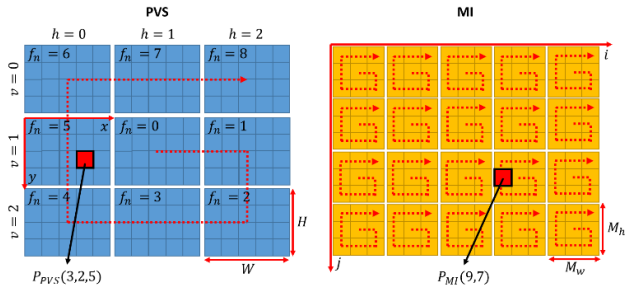

FIGURE 3. Pixel correspondence between PVS and MI data representations.

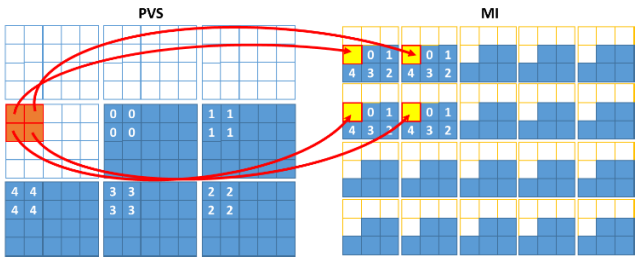

FIGURE 4. Converting a reference 2 × 2 block in the PVS representation to the MI representation.

PVS pixel position as follows:

$$\begin{pmatrix} i \\ j \end{pmatrix}_{MI} = \begin{pmatrix} h\,(f_n) + xM_w \\ v\,(f_n) + yM_h \end{pmatrix}_{PVS} \tag{1}$$

where, M_w and M_h correspond to the MI width and height, respectively, and f_n is the PVS frame number. As mentioned before, the (x, y) coordinates index the pixel position within each SAI and (h, v) coordinates index the SAI position within the LF. Since the SAIs are organized in a PVS the (h, v) coordinates depend on the frame number f_n. Fig. 3 illustrates how each MI image is generated from the 9 decoded SAIs when a spiral scan is used. Note that the dashed red arrow shows the scanning order being applied on the PVS representation and the consequent scanning order from the MI representation point of view. When applying (1) to the example in Fig. 3 the pixel at the $P_{PVS}(x, y, f_n)$ coordinate in the PVS representation is copied to the $P_{MI}(i, j)$ coordinate on the MI representation. This allows the MI decoded picture buffer to be gradually filled after encoding each full SAI. In this paper a spiral scan has been adopted as it is more efficient than, for example, the more straightforward raster scan as shown in [14], [35]. A serpentine scan could also be used instead of the spiral scan, as both have been shown to achieve similar coding efficiency [13]. Additionally, the spiral scan was chosen because the first frame is the central viewpoint and the last frames include the outer viewpoints. This characteristic is a favorable property for the addition of features such as viewpoint scalability and the fact that in lenslet LF images the outer viewpoints normally exhibit illumination issues. Regardless, the proposed LF coding solution based on a hybrid representation can be used with any scanning order.

Since the conversion to the MI representation is performed progressively the MI decoded picture buffer resembles a sparse LF image. Fig. 4 shows the conversion of the first 2×2 block when encoding the sixth SAI of the PVS from the example shown before in Fig. 3. As it is possible to see from the example in Fig. 4, the reference 2×2 (orange) block in the PVS representation, leads to 4 individual (yellow) pixels in 4 different MIs, in the MI representation. Because of this characteristic, the intra-MI and inter-MI prediction modes are applied pixelwise, instead of blockwise, as in the PVS representation case. In the example of Fig. 4, each one of the 4 individual (yellow) pixels in the MI representation is predicted using either the causal area in the same MI (intra-MI) or the causal area in neighboring MIs (inter-MI). After pixelwise pixel prediction, the predicted pixels are mapped back to the PVS representation positions forming a prediction block for the reference 2×2 block.

2) SELECTION OF THE PREDICTION MODE

The HEVC encoder decides between both intra and inter prediction modes, i.e., intra-SAI and inter-SAI prediction modes, respectively, by generating a prediction block for each prediction mode and the prediction mode that minimizes a Lagrangian RD cost function (represented as RD selection in Fig. 2), given by $J = D + \lambda R$, is selected. The distortion, D, is calculated by comparing the original block to the block generated by each prediction mode using a distortion metric such as the sum of absolute difference. The rate, R, is the number of bits required to signal such prediction mode to the decoder and λ is the Lagrangian multiplier dependent on the quantization parameter (QP) value. In the proposed coding approach, the same process is extended to the proposed prediction modes to be used in the MI data representation. The prediction modes in the SAI and MI representations also generate prediction blocks, which will then compete, in terms of RD cost, with the ones generated by the standard HEVC prediction modes.

3) PREDICTION MODE SIGNALING

Since in the proposed coding architecture four different types of redundancies can be exploited by the different prediction modes, it is also necessary to efficiently signal the usage of such prediction modes. The standard intra and inter modes, i.e., intra-SAI and inter-SAI, from HEVC are signaled as in the HEVC standard. The proposed intra-MI and inter-MI prediction modes also use the same signaling logic, from the 35 possible intra-SAI prediction modes, 8 directional mode indexes are allocated for intra-MI modes and inter-MI modes. The substituted modes are the suggested ones in [28], which include intra directions that are seldomly used. Table 1 shows the list of used modes (name/number) for each of the four prediction types.

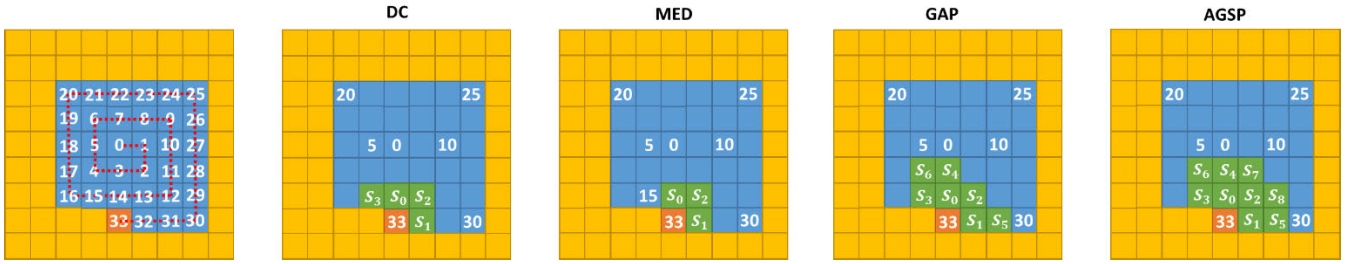

FIGURE 5. Pixel support for the pixel predictors DC, MED, GAP and AGSP, when using a clockwise spiral scan.

FIGURE 6. Generic pixel support and filling pattern.

IV. INTRA-MI PREDICTION

The intra-MI prediction modes aim to exploit intra-MI redundancy within each MI. For this, several pixel prediction modes are adopted, i.e., DC, MED [15], GAP [16], and AGSP [17]. These modes, with the exception of AGSP, are used in popular image coding approaches, such as HEVC, JPEG-LS and CALIC, respectively. AGSP was selected because it is able to outperform MED and GAP when encoding natural images [17]. However, such prediction modes cannot be used directly in the proposed codec. An adaptation of the prediction area is necessary because the causal area may differ, depending on the scanning strategy adopted. The following subsections describe the proposed pixel prediction modes, including the necessary adaptations to the spiral scanning strategy.

A. SPIRAL SCANNING PREDICTOR ADAPTATIONS

In the proposed hybrid representation, a clockwise spiral PVS scan is performed, causing the causal area to grow differently from a raster scan. When using a raster scan, as in the case of the original DC, MED, GAP and AGSP predictors, the causal area is always on the top and left of the current pixel. The predicted pixel generated by each of these predictors is a combination of part or all of the available pixels in the causal area. For example, the predicted pixel that is generated by the DC predictor is an average, i.e., a linear combination, of the surrounding pixels. The set of pixels that are selected within the causal area of pixels are hereafter referred to as pixel support. Differently from the raster scan, for the clockwise spiral scan, the causal area of each new pixel is not always available in the upper-left area relative to the new pixel. Thus, the pixel support is dynamically adapted to maximize the number of available pixels for prediction, which is illustrated in Fig. 5 for Frame 33, where each individual predictor is used to predict the orange pixel.

The spiral scanning is divided into four phases, i.e., Left, Up, Right and Down, named according to the direction of the spiral scan. Fig. 5 illustrates a pixel prediction, showing the causal area and the pixel support, when the scan direction is left (i.e., Left phase). For the following phases (Up, Right and Down) the same pixel prediction structure is used, but rotated relatively to the Left phase as follows:

- 90° for the Up phase;
- 180° for the Right phase;
- 270° for the Down phase.

At some positions of the scan, one or more pixels of the pixel support area may not exist. When some predictor pixel value is not available, due to being part of the non-causal area, it is copied from a neighboring location. The filling pattern for these pixels is shown in Fig. 6 with red arrows.

TABLE 1. List of mode name/number for the proposed codec.

Prediction type	Mode name/index
Intra-SAI	0 (Planar), 1 (DC), 2, 4, 5, 6, 8, 9, 10, 12, 13, 14, 16, 17, 18, 20, 21, 22, 24, 25, 26, 28, 29, 30, 32, 33, 34
Inter-SAI	Skip, Merge, Motion Compensation
Intra-MI	3, 7, 11, 15
Inter-MI	19, 23, 27, 31

B. DC PREDICTION MODE

The DC prediction mode consists in applying an average of the available pixels in a 3×3 template centered on the current pixel. In the example shown in Fig. 5 the predictor is an average of the values of pixels $S_0, \cdots, S_3$. If N is the number of available pixels, the prediction value $\hat{P}$ is generically determined by (2):

$$\hat{P} = \frac{\sum_0^{N-1} S_n}{N} \tag{2}$$

The number of available pixels varies between 1 and 4. Notice that the decoder has access to the same predictor values, since a causal prediction is used.

C. MED PREDICTION MODE

The MED [15] prediction mode consists of a 3-pixel template, as shown in Fig. 5. The prediction value is calculated by (3):

$$\hat{P} = \begin{cases} \min(S_1, S_0), & \text{if } S_2 \geq \max(S_1, S_0) \\ \max(S_1, S_0), & \text{if } S_2 \leq \min(S_1, S_0) \\ S_1 + S_0 - S_2, & \text{otherwise.} \end{cases} \tag{3}$$

TABLE 2. GAP prediction value calculation based on the vertical and horizontal gradients [16].

Edge	Threshold ($g_v - g_h$)	Prediction ($\hat{P}$)
Sharp Hor	> 80	S_1
Sharp Ver	< -80	S_0
Regular Hor	> 32	$\frac{S_1+S_0}{4}+\frac{S_3-S_2}{8}+\frac{S_1}{2}$
Regular Ver	< -32	$\frac{S_1+S_0}{4}+\frac{S_3-S_2}{8}+\frac{S_0}{2}$
Weak Hor	> 8	$\frac{3(S_1+S_0)}{8}+\frac{3(S_3-S_2)}{16}+\frac{S_1}{4}$
Weak Ver	< -8	$\frac{3(S_1+S_0)}{8}+\frac{3(S_3-S_2)}{16}+\frac{S_0}{4}$
Smooth	otherwise	$\frac{S_1+S_0}{2}+\frac{S_3-S_2}{4}$

D. GAP PREDICTION MODE

The GAP [16] prediction mode consists of a 7-pixel template, as shown in Fig. 5. Firstly, the vertical (g_v) and horizontal (g_h) gradients are estimated using (4):

$$\begin{aligned} g_h &= |S_1 - S_5| + |S_0 - S_2| + |S_0 - S_3| \\ g_v &= |S_1 - S_2| + |S_0 - S_4| + |S_3 - S_6|\,, \end{aligned} \tag{4}$$

Secondly, depending on the values of g_v and g_h, GAP will recognize weak, regular and sharp vertical (Ver) and horizontal (Hor) axis as well as smooth edges. The prediction value $\hat{P}$ is determined by the thresholds [16] and equations shown in Table 2.

E. AGSP PREDICTION MODE

The AGSP [17] prediction mode uses a 9-pixel predictor as, shown in Fig. 5. AGSP is able to determine horizontal, vertical and diagonal edges. In order to determine the direction of the edge, four gradients are calculated as defined in (5), corresponding to the horizontal (g_h), vertical (g_v), 45° diagonal (g_{45}) and −45° diagonal (g_{-45}):

$$\begin{aligned} g_h &= (2\,|S_1 - S_5| + 2\,|S_0 - S_2| + 2\,|S_0 - S_3| \\ &\quad + |S_4 - S_7| + |S_4 - S_6| \\ &\quad + |S_2 - S_8|)/9 + 1 \\ g_v &= (2\,|S_1 - S_2| + 2\,|S_0 - S_4| + |S_3 - S_6| \\ &\quad + |S_5 - S_8| + |S_2 - S_7|)/7 + 1 \\ g_{45} &= (2\,|S_1 - S_0| + 2\,|S_0 - S_6| + |S_5 - S_2| \\ &\quad + |S_2 - S_4|)/6 + 1 \\ g_{-45} &= (2\,|S_1 - S_8| + 2\,|S_0 - S_7| \\ &\quad + |S_3 - S_4|)/5 + 1 \end{aligned} \tag{5}$$

After calculating the four gradients, the two lowest ones are selected as g_{min} and g_{min2}. Additionally, the causal pixels P_{min} and P_{min2} that correspond to the direction of each of the selected gradients, i.e., $g_{\min}$ and $g_{\min 2}$, are selected. The correspondent causal pixel for the gradients g_v, g_h, g_{45} and g_{-45}, is S_0, S_1, S_2 and S_3, respectively. For example, if $g_{\min} = g_v$ and $g_{\min 2} = g_{45}$, then $P_{min} = S_0$ and $P_{min2} = S_2$. The final prediction value is calculated as defined by (6):

$$\hat{P} = \frac{g_{min}P_{min2} + g_{min2}P_{min}}{g_{min} + g_{min2}} \tag{6}$$

V. INTER-MI PREDICTION

The inter-MI prediction modes aim to exploit the inter-MI redundancy which is also known, in the state-of-the-art, as non-local spatial redundancy [13]. The similarities between the neighboring MIs in the LF image using the MI representation can be exploited in several ways [26]–[28]. However, most approaches available in the literature are block-based instead of pixel-based [13]. As an alternative to the block-based approaches, this paper adopts an LSP-based prediction mode, which is applied in a pixelwise manner, in order to exploit the inter-MI redundancy.

A. LSP- BASED PREDICTION MODE

LSP is a prediction method that adaptively estimates optimal linear coefficients using least squares training. Thus, the main advantage of the LSP-based prediction mode, when compared to the previously presented predictors, is its ability to dynamically adapt to the available causal area and determine the best prediction direction adapted to this area [18]. The least squares training step based on a least squares minimization problem is defined as (7):

$$\min_{\boldsymbol{a}} (\|\boldsymbol{y} - \boldsymbol{C}\boldsymbol{a}\|_2^2), \tag{7}$$

where $\boldsymbol{a} = [a_0, \ldots, a_{M-1}]^T$, a $M \times 1$ column vector, corresponds to the linear coefficients to estimate. The closed-form solution for (7) is given by (8):

$$\boldsymbol{a} = \left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}^T\boldsymbol{y}\right). \tag{8}$$

The matrix $\boldsymbol{C}$ is a $T \times M$ matrix, where M is the order of the LSP pixel support, i.e., the number of pixels that compose the pixel support, and T is the number of causal neighbors used for training, which can include the causal area in the current MI or several neighboring MIs; $\boldsymbol{y} = [y_0, \ldots, y_{T-1}]^T$ is a $T \times 1$ column vector where T is computed as (9):

$$T = (f_n - 1) \times M_T, \tag{9}$$

where f_n is the frame number of the PVS and M_T is the number of MIs used for training. Note that if $M_T = 1$ this mode can be considered intra-MI, because the training step only uses pixels from the current MI. If $M_T > 1$ then, this prediction mode is considered an inter-MI prediction mode. Additionally, the training size increases with the frame number, because more pixels are available for training.

B. ADAPTIVE PIXEL SUPPORT AND TRAINING

In order to perform the LSP training, the pixel support for the predictor (i.e., the pixels used for prediction after the training step is performed) needs to be determined, as well as matrix $\boldsymbol{C}$ and vector $\boldsymbol{y}$ need to be constructed. In the approach

FIGURE 7. Example of an adaptive pixel support generation by minimizing the Manhattan distance (M=5 and M=9).

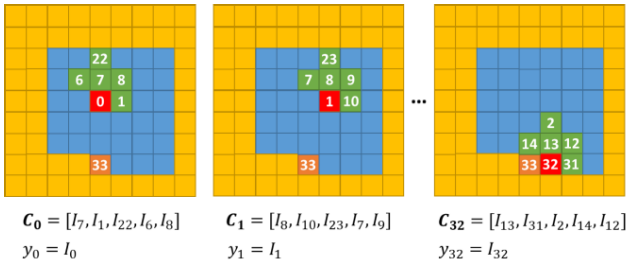

FIGURE 8. Composition of matrix C and vector y for the current MI necessary for the training step of LSP.

proposed in this paper, since the causal area grows in a spiral order, the pixel support needs to be adapted accordingly, by selecting the M available pixels closest to the current pixel. The distance between the current pixel, I_c, and the causal pixel, I_n, is determined by the Manhattan distance, d, as (10):

$$d\,(I_c, I_n) = |I_{c_x} - I_{n_x}| + |I_{c_y} - I_{n_y}|. \tag{10}$$

For example, an adaptive pixel support generation, by minimizing the Manhattan distance, for $M = 5$ and $M = 9$, is shown in Fig. 7; the numbers displayed in the neighboring pixels (blue pixels) correspond to the Manhattan distance to the current pixel. Using as an example the pixel 33 shown in Fig. 5, $T = 32 \times M_T$, because $F_n = 33$. High values of M_T may heavily increase the computational complexity, therefore, in this work, only values between 1 and 9 have been considered. As illustrated in Fig. 8, vector $\boldsymbol{y}$ comprises every single causal pixel inside the MIs that contain the training area, with the exception of the current pixel. The matrix $\boldsymbol{C}$ is composed of the pixel support determined in the previous step, centered on each neighboring pixel included in vector $\boldsymbol{y}$. Using the example shown in Fig. 8, if using a 5$^{\text{th}}$ order pixel support, i.e., LSP using $M = 5$, this support is centered on pixel 0 and, therefore, the first row of matrix $\boldsymbol{C}$ is $\boldsymbol{C_0} = [I_7, I_1, I_{22}, I_6, I_8]$ and $y_0 = I_0$.

Since matrices $\boldsymbol{C}$ and $\boldsymbol{y}$ are already determined, then (8) can be solved. After solving (8), vector $\boldsymbol{a}$ is used to estimate a prediction value $\hat{P}$ for the current pixel as (11):

$$\hat{P} = \sum_{0}^{M-1} S_n \times a_n, \tag{11}$$

where S_n are the M pixels that compose the pixel support.

VI. PERFORMANCE EVALUATION

In this section the performance of the proposed LF coding solution based on a hybrid LF data representation is evaluated against the most relevant state-of-the-art 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 are presented and discussed. A statistical analysis of the prediction mode usage as well as the performance for different encoder configurations are evaluated and discussed.

TABLE 3. Benchmark coding solutions.

Codec	Input Resolution	Color format	LF Data Representation
HEVC-PVS [14]	625×434 (169 SAIs)	YUV 4:4:4 10 bit	SAI
WaSP [50]	625×434 (169 SAIs)	RGB 4:4:4 10 bit	SAI
MuLE [51]	625×434 (169 SAIs)	RGB 4:4:4 10 bit-	4D LF
HEVC-SS [21]	8125×5642 (13×13 MIs)	YUV 4:2:0 8 bit	MI
HEVC-HOP [27]	8125×5642 (13×13 MIs)	YUV 4:2:0 8 bit	MI
HEVC-LLE [28]	8125×5642 (13×13 MIs)	YUV 4:2:0 8 bit	MI
HEVC-HR (proposed)	**625×434 (169 SAIs)**	**YUV 4:4:4 10 bit**	**Hybrid (SAI and MI)**

A. TEST METHODOLOGY

The experimental tests for all coding solutions presented in this section adopted the JPEG Pleno – Light Field Coding Common Test Conditions [56]. The EPFL dataset, comprised of 12 LF images acquired using a Lytro Illum camera, was used [57] to evaluate the several benchmarks. The raw LF images are first converted to the 4D LF representation using the LF Toolbox [55] and then converted to the MI and PVS representations, prior to being encoded and decoded. After the decoding step, 13 × 13 SAIs are generated with a resolution of 625 × 434 pixels, using the YUV 4:4:4 10-bit color format [56].

The state-of-the art LF codecs that were used as benchmarks, include: HEVC-PVS [14], WaSP [50], MuLE [51], HEVC-SS [21], HEVC-HOP [27] and HEVC-LLE [28]. The proposed codec, based on a hybrid LF representation, is referred to as HEVC-HR. The LF input resolution, LF data representation and the supported color format, are presented in Table 3.

As mentioned before, the output color format for objective comparison of all benchmarks is YUV 4:4:4 10-bit. However some codec implementations are limited to the YUV 4:2:0 8-bit color format [58]. This is the case for the HEVC-SS, HEVC-HOP and HEVC-LLE codecs. For these codecs, a preprocessing step is applied at the encoder, to generate the YUV 4:2:0 8-bit input color format, and a post-processing step is performed at the decoder to generate the YUV 4:4:4 10-bit output color format.

Table 4 shows the list of tested codecs including the corresponding control configurations. The different QP and λ values selected allow the use of a common bitrate range for every tested codec, enabling a direct comparison through the

TABLE 4. List of tested codecs and corresponding control configurations.

Codec	Configuration
HEVC-HR HEVC-PVS	$QP = [17,22,27,32,37,42]$
HEVC-SS HEVC-HOP HEVC-LLE	$QP = [22,27,32,37,42,47]$
MuLE	$\lambda = [270, 3880, 30000, 310000, 4600000]$
WaSP	$Target\ bpp = [0.001, 0.005, 0.02, 0.1, 0.75]$

Bjontegaard difference (BD) metrics. The HEVC PVS based codecs use the low delay with B slices configuration and the MI based codecs use the intra main configuration. The RD analysis is done by comparing the size of the bitstream (rate) and the average PSNR-YUV of the 13×13 SAIs (distortion) generated at the decoder side for each codec. The average PSNR-YUV of the 13×13 SAIs is calculated by comparing the decoded SAIs when encoded by different codecs and the reference 13×13 SAIs.

B. EXPERIMENTAL RESULTS

The proposed HEVC-HR was tested in three different phases. In the first phase, each intra-MI prediction mode was individually tested, i.e., DC, MED, GAP and AGSP. In the second phase, each inter-MI prediction mode was tested, i.e., several configurations in terms of LSP order and training area of the proposed LSP prediction mode were tested. In the final phase the prediction modes presenting a higher trade-off between coding efficiency and computational complexity were selected to be part of HEVC-HR.

1) INTRA-MI PREDICTION MODES EVALUATION

Table 5 presents the results of each individual intra-MI prediction mode described in Section IV. The Table shows the average BD-PSNR-YUV and average BD-RATE for the 12 images of the EPFL dataset, comparing the HEVC-PVS with the HEVC-HR using each of the represented prediction mode. As can be observed, the prediction mode with highest bitrate savings (12.87%) is AGSP. From the experimental results it can be inferred that increasing the order of the prediction increases the prediction accuracy and, consequently, improves the coding efficiency. The only exception is the MED prediction mode, which achieves lower bitrate savings when compared to the DC prediction mode, which has an order value of 1 to 4, depending on how many support pixels are available. Overall, it is possible to observe that the proposed intra-MI prediction modes improve the LF image coding efficiency. Since their low computational complexity, especially when compared to LSP-based prediction modes, DC, MED, GAP and AGSP were used in the final version of HEVC-HR.

2) INTER-MI PREDICTION MODES EVALUATION

Table 6 presents the experimental results for different configurations of the LSP-based modes described in Section V. Two parameters were tested, corresponding to the LSP Order (M) and the number of MIs used for training (M_T). By varying the LSP order, M, it is possible to compare the performance of an adaptive mode with prediction modes with similar orders, like MED, GAP and AGSP. The value M_T was tested for 1, 5 and 9, which corresponds to use, respectively: the current MI for training (equivalent to an intra-MI prediction mode, as mentioned in Section V); the current MI and the MI on the left, top, right and bottom of the current MI; and the current MI and the 8 surrounding MIs.

From Table 6 it is possible to conclude that, regardless of the LSP order, when the training area (M_T) increases, the coding efficiency also increases. This is especially noticeable when using more than one MI for training. However, increasing the order does not always result into higher bitrate savings. This occurs because the use of higher LSP orders requires larger areas of reconstructed pixels, for LSP training. The size of the available training grows from the first frames to the last ones, affecting the quality of the training step. Thus, the use of higher LSP orders will be more efficient only at later stages of the coding process, while using a lower order may be beneficial since an earlier stage of the coding process. Higher prediction orders and larger training areas also have a negative impact (i.e., increase) on the computational complexity. In Table 6, the best three LSP based prediction methods in terms of bitrate savings vs. computational complexity are represented in bold (LSP3, LSP5 and LSP7, using 5 MIs for training). These modes were included in HEVC-HR. LSP9 modes were excluded, due to their high computational complexity.

TABLE 5. BD-PSNR-YUV and BD-RATE results against HEVC-PVS using different intra-MI prediction modes.

Prediction Mode	Order	BD-PSNR-YUV	BD-RATE
DC	1 to 4	0.25 dB	-10.72 %
MED	3	0.15 dB	-6.77%
GAP	7	0.28 dB	-12.12%
AGSP	9	0.29 dB	-12.87%

TABLE 6. BD-PSNR-YUV and BD-RATE results against HEVC-PVS using different LSP prediction modes configurations.

LSP Order (M)	M_T	BD-PSNR-YUV	BD-RATE
3	*1 (Intra)*	*0.26 dB*	*-11.37 %*
	5 (Inter)	**0.31 dB**	**-13.10 %**
	9 (Inter)	0.31 dB	-13.29 %
5	*1 (Intra)*	*0.22 dB*	*-9.73 %*
	5 (Inter)	**0.32 dB**	**-13.42 %**
	9 (Inter)	0.43 dB	-13.98 %
7	*1 (Intra)*	*0.17 dB*	*-7.47 %*
	5 (Inter)	**0.30 dB**	**-12.55 %**
	9 (Inter)	0.32 dB	-13.50 %
9	*1 (Intra)*	*0.20 dB*	*-6.89 %*
	5 (Inter)	0.32 dB	-13.31 %
	9 (Inter)	0.34 dB	-13.69 %

3) HEVC-HR USING INTRA-MI AND INTER-MI PREDICTION MODES

Table 7 presents the experimental results achieved by two different configurations of HEVC-HR: a) using all of the prediction modes selected in the previous sections (DC,

TABLE 7. BD-PSNR-YUV and BD-RATE results against HEVC-PVS using HEVC-HR, MuLE and WaSP codecs.

Image	**HEVC-HR** vs HEVC-PVS		**HEVC-HR (Intra-MI)** vs HEVC-PVS		MuLE vs HEVC-PVS		WaSP vs HEVC-PVS	
	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)
I01	0.69	-24.65	0.48	-18.04	**0.97**	**-33.15**	-0.24	9.14
I02	0.92	-30.37	0.63	-21.72	**1.54**	**-45.55**	0.56	-23.07
I03	0.72	-25.34	0.50	-18.30	**0.81**	**-28.74**	0.25	-17.04
I04	**0.57**	**-25.49**	0.44	-20.20	0.48	-22.29	-0.06	-2.05
I05	**0.49**	**-23.40**	0.35	-17.48	0.40	-20.79	-0.19	1.93
I06	**0.33**	**-17.59**	0.25	-13.49	-0.58	31.16	-0.85	64.00
I07	**0.50**	**-20.44**	0.36	-15.30	0.23	-12.20	-0.79	25.53
I08	**0.36**	**-17.68**	0.27	-13.58	-0.70	38.08	-1.05	39.75
I09	0.49	-19.00	0.31	-12.46	**0.49**	**-19.94**	-0.01	-7.13
I10	**0.60**	**-24.91**	0.48	-20.35	0.48	-21.34	-0.07	9.44
I11	**0.37**	**-18.20**	0.26	-12.92	0.07	-5.32	-0.99	36.33
I12	**0.69**	**-25.23**	0.45	-17.26	-0.19	5.02	-1.02	34.86
AVG.	**0.56**	**-22.69**	0.40	-16.76	0.33	-11.26	-0.37	14.31

TABLE 8. Average prediction mode usage across the six qps, in percentage of pixels for the HEVC-HR codec.

Image	inter-SAI	intra-SAI	intra-MI				inter-MI		
			DC	MED	GAP	AGSP	LSP3	LSP5	LSP7
I01	**72.3**	1.5	3.7	2.6	*1.4*	2.4	3.2	5.3	**7.5**
I02	**73.0**	1.4	3.1	2.2	*1.2*	2.2	3.3	4.9	**8.8**
I03	**78.7**	1.3	2.3	1.5	*0.8*	1.9	2.5	4.0	**7.1**
I04	**75.5**	*1.3*	3.3	1.8	1.3	2.1	2.4	4.5	**7.8**
I05	**76.0**	2.0	3.5	2.4	*1.0*	1.8	3.2	4.3	**5.8**
I06	**70.6**	**6.9**	5.5	3.0	2.2	*1.6*	3.0	3.4	3.7
I07	**79.1**	2.2	2.7	2.0	*1.4*	1.5	2.5	3.6	**5.0**
I08	**68.9**	**7.9**	5.0	3.3	2.0	*1.8*	3.1	4.0	4.1
I09	**69.8**	2.3	3.9	3.7	*1.1*	1.8	3.2	6.3	**8.0**
I10	**75.6**	*1.4*	4.1	2.4	1.8	2.2	2.9	3.9	**5.7**
I11	**69.2**	4.0	4.4	**4.9**	2.5	*1.6*	4.1	4.8	4.6
I12	**70.0**	2.2	3.3	3.3	*1.5*	2.2	3.3	6.2	**7.8**

MED, GAP, AGSP, LSP3, LSP5 and LSP7); and b) using only the intra-MI (DC, MED, GAP and AGSP) prediction modes; as well as the c) MuLE and d) WaSP LF image codecs. The results compare the performance of these four methods with HEVC-PVS.

From Table 7 it is possible to observe that HEVC-HR consistently outperforms HEVC-PVS in terms of coding efficiency. An average of 22.69% of bitrate savings is achieved by using multiple prediction modes, which is considerably higher than the best performance of an individual prediction mode, i.e., LSP5, with 13.98%. Additionally, HEVC-HR using only intra-MI prediction modes, achieves 16.76% bitrate savings. From this table it is also possible to observe that MuLE exhibits an average bitrate savings, of 11.26%, while WaSP exhibits an average increase in bitrate of 14.31%, relatively to HEVC-PVS. This means that the proposed HEVC-HR is on average more efficient than both MuLE and WaSP for the 12 LF images. HEVC-HR is more efficient than WaSP for every test image, however, WaSP is able to provide better SAI random access than the remaining benchmarks, which may be an important feature for LF content navigation.

From the 12 test LF images used, HEVC-HR is more efficient than MuLE for 8 out of the 12 images. Although HEVC-HR and MuLE are conceptually very different LF codecs, both rely on the DCT as their primary spatial transform, which is characterized by its strong energy compaction property. The major difference between the two DCT approaches, apart from the number of dimensions, is the fact that MuLE applies the DCT directly on the coding block samples [46] and HEVC-HR applies the DCT on the residue of the coding block. HEVC-HR thus tends to be more efficient on less textured images, such as studio images (I06, I07 and I08) or color/ISO charts (I11 and I12), for which the prediction techniques are more efficient (i.e. able to produce very low energy residue blocks). This observation justifies the use of flexible and efficient prediction techniques for different types of redundancy, as is proposed in this work, which have a high impact on the overall compression efficiency. MuLE on the other hand tends to be more efficient on more textured images, where prediction is not a clear advantage. Consequently, MuLE is able to outperform HEVC-HR for I01, I02, I03 and I09, although being less efficient for all other images.

The RD curves for the four LF image codecs when encoding images I02, I04, I06 and I08 are shown in Fig. 9. From these RD curves it is possible to observe that HEVC-HR outperforms HEVC-PVS consistently for every image. Additionally, HEVC-HR tends to be much more efficient than MuLE in images such as I06 and I08. On images such as I02 and I04 although HEVC-HR is in general less efficient than MuLE it is still able to produce very competitive results.

In order to further analyze the usefulness of each prediction mode in HEVC-HR, the average prediction mode usage across the six QPs is shown in Table 8. The values in bold correspond to the most used prediction modes and italic signals the least used ones. It is possible to observe that most of the pixels are encoded using inter-SAI prediction modes, because the inter-view redundancy is very high in this type of LF content. However, when analyzing the prediction mode usage for intra-MI and inter-MI prediction, which include the new 7 modes, it is possible to conclude that, for most images,

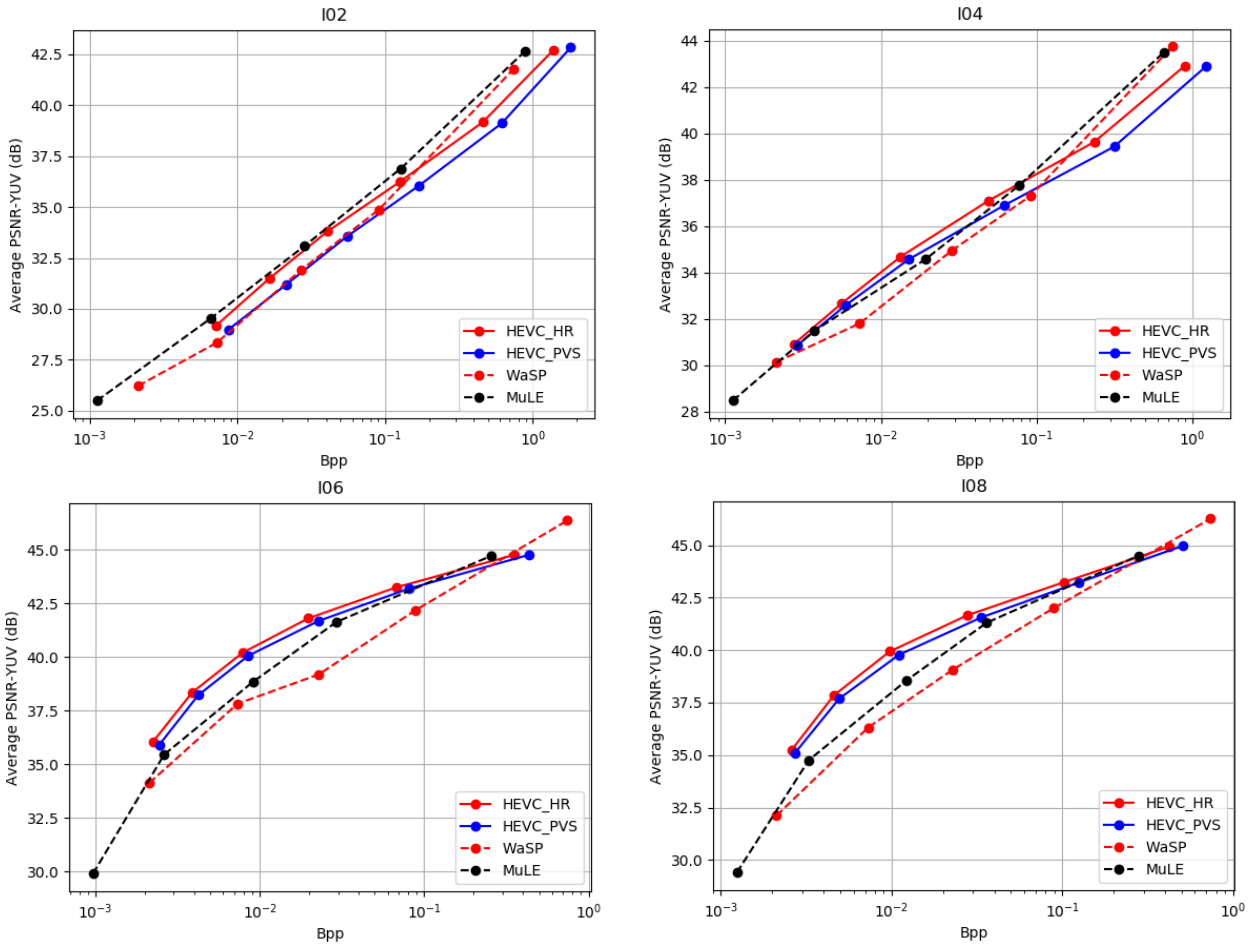

FIGURE 9. RD Curves comparing the proposed hybrid representation LF coding approach (HEVC-HR) and HEVC-PVS, MuLE and WaSP, for four test images.

the new modes are more often used than the intra-SAI modes, i.e., DC, Planar and the 26 remaining directional modes. These statistics allow to conclude that exploiting the intra- and inter-MI redundancy results in more coding efficiency than exploiting the spatial redundancy within each SAI.

Amongst the new prediction modes, proposed in this work, the most used prediction mode is LSP7, which verifies the assumption made about the usefulness of LSP-based prediction modes. LSP7 when tested individually is not as efficient as LSP5, because of the higher requirements in terms of training area in the initial phase of encoding process. Nevertheless, the use of 3 LSP-based prediction modes, with different orders and, therefore, different requirements in terms of training area, allows the encoder to choose the more suitable prediction mode for every phase of the coding process.

4) EXPERIMENTAL EVALUATION FOR YUV 4:2:0 8-BIT COLOR FORMAT

In order to compare HEVC-HR with MI data representation LF coding approaches, a set of similar tests using the YUV 4:2:0 8-bit color format were performed, since the available implementations of HEVC-SS, HEVC-HOP and HEVC-LLE are only compatible with the YUV 4:2:0 8-bit color format. This allows for a fair comparison between PVS, MI and the proposed hybrid approach for LF image coding. The RD curves for the images I02, I04, I06, I08 are shown in Fig. 10. These RD curves are used to compare all the codecs listed in Table 3 using the YUV 4:2:0 8-bit color format.

The RD curves in Fig. 10 show that coding approaches based on PVS outperform approaches based on MI [59]. This is explained by the fact that the inter-view redundancy between SAIs is very high and easily exploited by the inter-prediction tools of HEVC. An extensive comparison between both LF data representations using two different color formats can be seen in [19].

Although HEVC-PVS is more efficient than approaches based on MI, the proposed HEVC-HR, based on a hybrid LF data representation, can achieve the highest coding efficiency, outperforming HEVC-PVS. The achieved average bitrate savings when compared to HEVC-PVS, for the 12 LF images in the YUV 4:2:0 8-bit color format, is 9.36%. This shows that the proposed hybrid data representation and prediction modes are able to increase the coding efficiency, regardless of the color format.

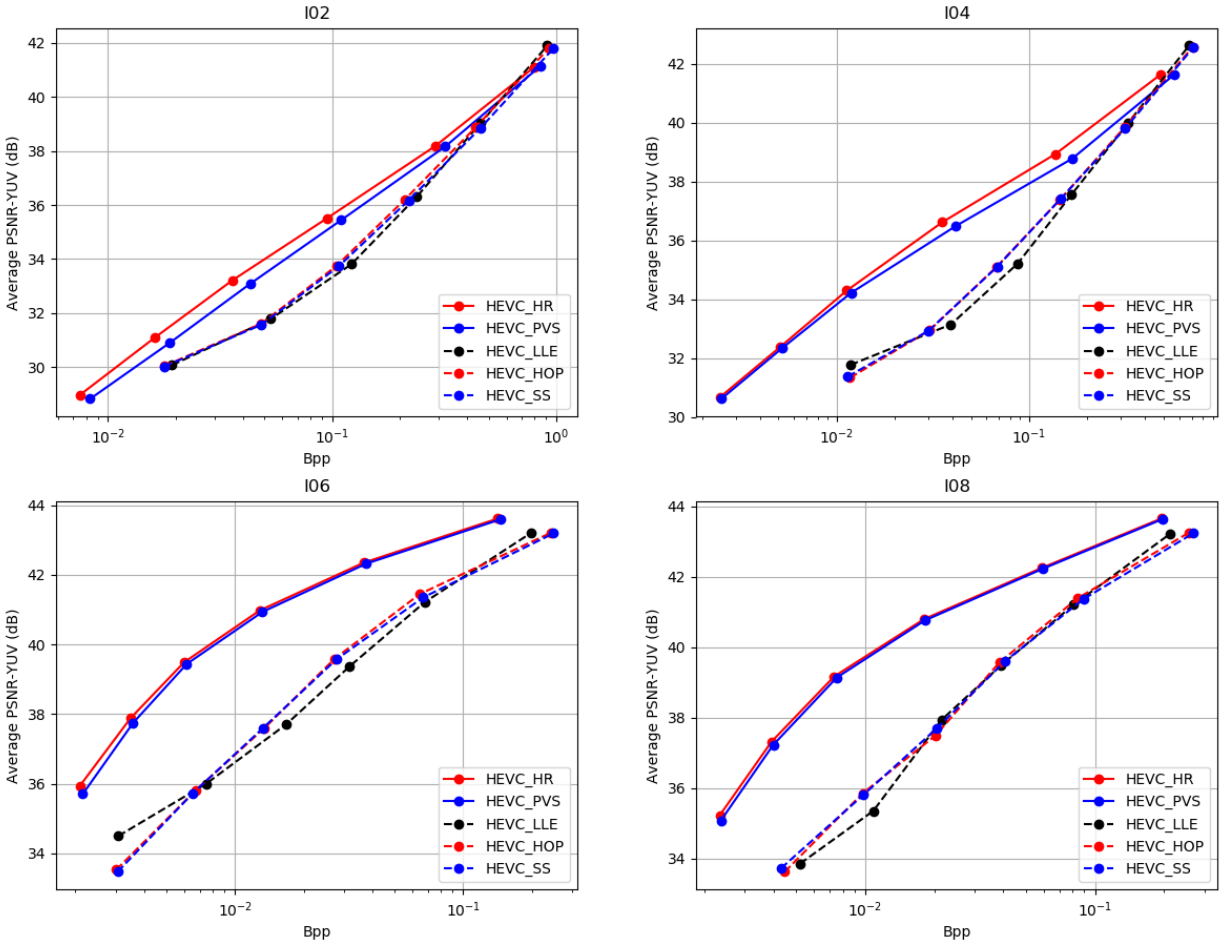

FIGURE 10. RD performance of the proposed hybrid representation LF coding approach (HEVC-HR) against the MI and PVS representation approaches for selected LF test images.

TABLE 9. BD-PSNR-YUV and BD-RATE results against HEVC-PVS using HEVC-HR, MuLE and WaSP codecs for HDCA LF images.

Image	**HEVC-HR** vs HEVC-PVS		**HEVC-HR (intra-MI)** vs HEVC-PVS		MuLE vs HEVC-PVS		WaSP vs HEVC-PVS	
	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)	BD-PSNR (dB)	BD-RATE (%)
Greek	**0.03**	**-1.07**	0.01	-0.45	-2.38	126.99	-0.11	3.42
Sideboard	**0.20**	**-5.50**	0.07	-1.93	-2.48	125.28	-0.59	20.37
Set2	**0.00**	**-0.22**	0.00	-0.07	-9.95	2364.71	-0.42	2.84
Tarot	**0.10**	**-3.34**	0.06	-2.10	-4.08	253.44	-1.00	40.31
AVG.	**0.08**	**-2.53**	0.04	-1.14	-4.72	717.61	-0.53	16.74

5) EXPERIMENTAL EVALUATION FOR HDCA LF IMAGES

In order to gauge the performance of the proposed HEVC-HR coding solution against MuLE and WaSP for LFs with wider baselines, four HDCA LF images were tested. Once again, the JPEG Pleno – Light Field Coding Common Test Conditions [56] were adopted to evaluate the objective quality of the HDCA LF images. The experimental results for the test images *Greek*, *Sideboard* (both $512 \times 512 \times 9 \times 9$), *Set2* ($1920 \times 1080 \times 33 \times 11$) and *Tarot* ($1024 \times 1024 \times 17 \times 17$) are shown in Fig. 11 and Table 9.

The experimental results depicted in Fig. 11 show that HEVC-HR, despite being only slightly more efficient than HEVC-PVS, it is able to outperform all the remaining coding solutions for higher bitrates, i.e., at bitrates higher than 0.01 bpp. For lower bitrates it is possible to see that WaSP is the most efficient solution, notably for *Set2*. However, the average BD-PSNR-YUV and average BD-RATE results for the HDCA LF images, shown in Table 9, demonstrate that WaSP is outperformed by HEVC-PVS for every tested image, in terms of BD-RATE. HEVC-HR, on the other hand, outperforms HEVC-PVS for every image, although, as mentioned above, with a lower margin when compared to the narrow baselines, achieving 2.53% bitrate savings. This is explained by the lower intra-MI and inter-MI redundancy that is present in HDCA LF images. This lower redundancy also explains the lower performance of MuLE for this type of LF content. Since MuLE does not have prediction tools that are able to compensate the wider baseline [51], its coding performance

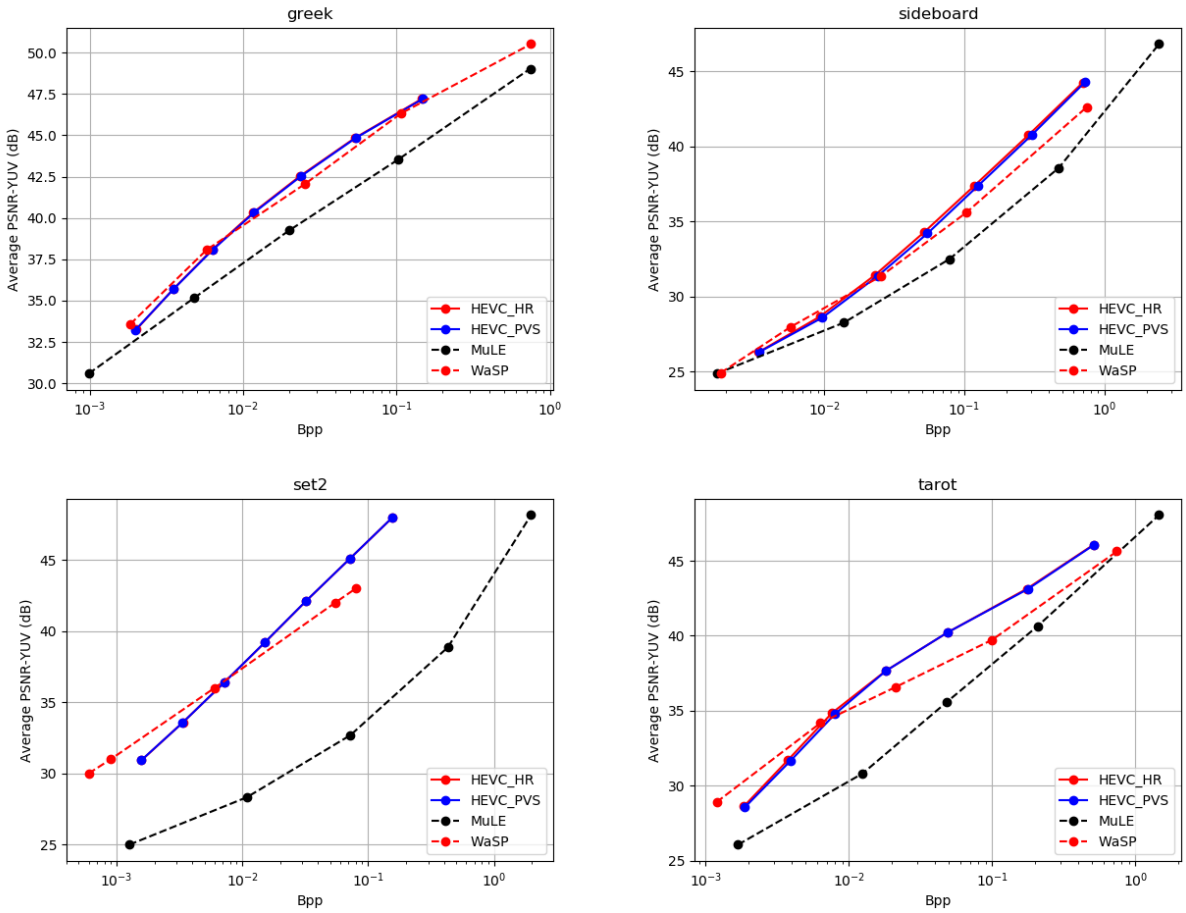

FIGURE 11. RD Curves comparing the proposed hybrid representation LF coding approach (HEVC-HR) and HEVC-PVS, MuLE and WaSP for HDCA LF images.

is strongly affected. Regardless, this experimental evaluation shows that the proposed HEVC-HR outperforms all the tested benchmarks for several types of baselines. This is possible, as explained before, by adaptively exploiting the type of redundancy (SAI or MI) that is more predominant on a coding block basis depending on the RD criterion. Since the most predominant type of redundancy changes based on the captured scene and baseline type, the coding efficiency can be maximized by using the appropriated coding approach for each coding block.

6) COMPUTATIONAL COMPLEXITY ASSESSMENT

The computational complexity of the tested codecs is presented in Table 10. The runtime values shown in this table were measured while encoding and decoding the image I01, with a $QP = 22$, for all the listed HEVC-based codecs. MuLE was tested using a λ value of 270 and WaSP was tested using a target bitrate of 0.75 bpp. 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.

Although the coding efficiency of the proposed HEVC-HR is higher than all the tested benchmarks, this comes at the expense of a higher computational complexity. From Table 10 it is possible to see that HEVC-HR takes a much longer time to encode and decode the same LF image. However, none of the implementations, including the HEVC-HR and the MI coding approaches, are

TABLE 10. Codec single thread computational complexity comparison.

Codec	Encoder		Decoder	
	Run Time (hours)	vs HEVC-PVS (ratio)	Run Time (seconds)	vs HEVC-PVS (ratio)
YUV 4:2:0 8bpp				
HEVC-PVS	0.34	-	1.22	-
HEVC-SS	4.43	13.18	372.10	304.01
HEVC-HOP	6.50	19.32	396.81	324.19
HEVC-LLE	11.74	34.90	1011.89	826.71
HEVC-HR	22.18	65.25	326.10	267.30
HEVC-HR (intra-MI)	**0.38**	**1.12**	**5.25**	**4.30**
YUV 4:4:4 10bpp				
HEVC-PVS	0.54	-	3.21	-
MuLE	0.15	0.28	18.19	5.67
WaSP*	**0.14**	**0.26**	38.46	11.98
HEVC-HR	24.14	44.70	3293.75	1026.09
HEVC-HR (intra-MI)	0.59	1.09	**11.53**	**3.58**

*using multithread (8 threads)

computationally optimized. It is also possible to see that when testing HEVC-HR using only the intra-MI prediction modes, the computational complexity is only marginally higher than HEVC-PVS, i.e., the increase in computational complexity is below 10%, while still being on average more efficient than MuLE and WaSP. Thus, it is possible to conclude that the computational complexity increase of HEVC-HR relative to HEVC-PVS comes mostly from the inter-MI prediction modes, therefore these prediction modes would benefit from a more optimized implementation or parallelization.

VII. CONCLUSIONS

In this paper, a new hybrid LF data representation paradigm for LF data coding is proposed. A HEVC-based codec implementation is described, as well as a set of pixel-based prediction modes to efficiently compress LF images. The hybrid LF data representation comprises both MI- and SAI-based representations to enhance the reference domain for the prediction modes. To efficiently exploit the intra-MI redundancy within each MI, a set of pixel-based prediction methods, i.e., DC, MED, GAP and AGSP were adapted to the proposed codec. Additionally, in order to exploit the inter-MI redundancy, efficient pixel prediction modes based on LSP using different order values were proposed.

The proposed HEVC-HR codec was evaluated against state-of-the-art codecs. When compared to HEVC-PVS, for the YUV 4:4:4 10-bit color format, an average bitrate saving of 22.69% was achieved, while for the YUV 4:2:0 8-bit color format, the average bitrate saving was 9.36%. Additionally, the RD curves show that the proposed HEVC-HR also outperforms the MI-based benchmarks, such as HEVC-SS, HEVC-HOP and HEVC-LLE using the YUV 4:2:0 8-bit color format, for all used test images. Approaches such as MuLE and WaSP, which are integral parts of the JPEG Pleno standard, were also used as benchmarks, being more appropriated for narrow and wide baselines, respectively. Such approaches were outperformed by the proposed HEVC-HR solution, which only achieve overall bitrate savings over HEVC-PVS of 11.26% and −14.31%, respectively.

To validate the flexibility of the proposed HEVC-HR for LFs captured with different baselines, an additional performance evaluation was performed using LFs captured using HDCAs. In this case, when compared to HEVC-PVS, an average bitrate savings of 2.53% was achieved with the proposed HEVC-HR. Although MuLE and WaSP are on average less efficient than HEVC-PVS, WaSP is the most efficient solution for low bitrates.

Future work includes adding SAI scalability and random access capabilities to the proposed HEVC-HR solution, as these functionalities vastly improve compatibility with legacy displays and LF navigation efficiency.

ACKNOWLEDGMENT

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 *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Mar. 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 *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, San Francisco, CA, USA, Apr. 2009, pp. 1–8, doi: 10.1109/ICCPHOT.2009.5559008.
[4] 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.
[5] J. Arai, *Integral Three-Dimensional Television (FTV Seminar)*, document ISO/IEC JTC1/SC29/WG11 MPEG2014/N14552, Sapporo, Japan, Jul. 2014.
[6] L. Toni, G. Cheung, and P. Frossard, "In-network view synthesis for interactive multiview video systems," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 852–864, May 2016, doi: 10.1109/TMM.2016.2537207.
[7] L. Toni and P. Frossard, "Optimal representations for adaptive streaming in interactive multiview video systems," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2775–2787, Dec. 2017, doi: 10.1109/TMM.2017.2713644.
[8] O. Stankiewicz, M. Domanski, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2182–2195, Aug. 2018, doi: 10.1109/TMM.2018.2790162.
[9] P. Ramanathan, M. Kalman, and B. Girod, "Rate-distortion optimized interactive light field streaming," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 813–825, Jun. 2007, doi: 10.1109/TMM.2007.893350.
[10] C. Conti, L. D. Soares, and P. Nunes, "Light field coding with field-of-view scalability and exemplar-based interlayer prediction," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2905–2920, Nov. 2018, doi: 10.1109/TMM.2018.2825882.
[11] *JPEG PLENO Abstract and Executive Summary*, document ISO/IEC JTC 1/SC 29/WG1 N6922, Sydney, NSW, Australia, Feb. 2015.
[12] *MPEG-I Technical Report on Immersive Media*, document ISO/IEC JTC1/SC29/WG11 N17069, Jul. 2017.
[13] C. Conti, L. D. Soares, and P. Nunes, "Dense light field coding: A survey," *IEEE Access*, vol. 8, pp. 49244–49284, 2020, doi: 10.1109/ACCESS.2020.2977767.
[14] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, "Data formats for high efficiency coding of lytro-illum light fields," in *Proc. Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Orleans, France, Nov. 2015, pp. 494–497, doi: 10.1109/IPTA.2015.7367195.
[15] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000, doi: 10.1109/83.855427.
[16] X. Wu and N. Memon, "CALIC-a context based adaptive lossless image codec," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf. Proc.*, vol. 4. Atlanta, GA, USA, May 1996, pp. 1890–1893, doi: 10.1109/ICASSP.1996.544819.
[17] H. Tang, "A gradient based predictive coding for lossless image compression," *IEICE Trans. Inf. Syst.*, vol. 89, no. 7, pp. 2250–2256, Jul. 2006, doi: 10.1093/ietisy/e89-d.7.2250.
[18] X. Li and M. T. Orchard, "Edge-directed prediction for lossless compression of natural images," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 813–817, Jun. 2001, doi: 10.1109/83.923277.
[19] R. J. S. Monteiro, N. M. M. Rodrigues, S. M. M. Faria, and P. J. L. Nunes, "Light field image coding: Objective performance assessment of Lenslet and 4D LF data representations," *Proc. SPIE*, vol. 10752, Sep. 2018, Art. no. 107520D, doi: 10.1117/12.2322713.
[20] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1027–1034, doi: 10.1109/CVPR.2013.137.
[21] 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.

[22] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574667.

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

[24] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574673.

[25] 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.

[26] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574670.

[27] R. J. Monteiro, P. Nunes, N. Rodrigues, and S. M. M. de Faria, "Light field image coding using high-order intrablock prediction," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1120–1131, Oct. 2017, doi: 10.1109/JSTSP.2017.2721358.

[28] L. F. R. Lucas, C. Conti, P. Nunes, L. D. Soares, N. M. M. Rodrigues, C. L. Pagliari, E. A. B. Da Silva, and S. M. M. De Fari, "Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 11–15.

[29] D. Liu, P. An, R. Ma, C. Yang, and L. Shen, "3D holoscopic image coding scheme using HEVC with Gaussian process regression," *Signal Process., Image Commun.*, vol. 47, pp. 438–451, Sep. 2016, doi: 10.1016/j.image.2016.08.004.

[30] D. Liu, P. An, R. Ma, W. Zhan, X. Huang, and A. A. Yahya, "Content-based light field image compression method with Gaussian process regression," *IEEE Trans. Multimedia*, vol. 22, no. 4, pp. 846–859, Apr. 2020, doi: 10.1109/TMM.2019.2934426.

[31] Y. Li, M. Sjostrom, and R. Olsson, "Coding of plenoptic images by using a sparse set and disparities," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Turin, Italy, Jun. 2015, pp. 1–6, doi: 10.1109/ICME.2015.7177510.

[32] Y. Li, M. Sjostrom, 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.

[33] M. C. Forman, "Quantisation strategies for 3D-DCT-based compression of full parallax 3D images," in *Proc. 6th Int. Conf. Image Process. its Appl.*, Dublin, Ireland, 1997, pp. 32–35, doi: 10.1049/cp:19970848.

[34] A. A. Mieee, "A 3D DCT compression algorithm for omnidirectional integral images," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process. Proc.*, Toulouse, France, May 2006, pp. 1–4, doi: 10.1109/ICASSP.2006.1660393.

[35] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 4733–4737, doi: 10.1109/ICIP.2015.7351705.

[36] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Seattle, WA, USA, Jul. 2016, pp. 1–4, doi: 10.1109/ICMEW.2016.7574674.

[37] C. Jia, Y. Yang, X. Zhang, X. Zhang, S. Wang, S. Wang, and S. Ma, "Optimized inter-view prediction based light field image compression with adaptive reconstruction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 4572–4576, doi: 10.1109/ICIP.2017.8297148.

[38] W. Ahmad, R. Olsson, and M. Sjostrom, "Towards a generic compression solution for densely and sparsely sampled light field data," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 654–658, doi: 10.1109/ICIP.2018.8451051.

[39] W. Ahmad, R. Olsson, and M. Sjostrom, "Interpreting plenoptic images as multi-view sequences for improved compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 4557–4561, doi: 10.1109/ICIP.2017.8297145.

[40] N. Mehajabin, S. R. Luo, H. Wei Yu, J. Khoury, J. Kaur, and M. T. Pourazad, "An efficient random access light field video compression utilizing diagonal inter-view prediction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 3567–3570, doi: 10.1109/ICIP.2019.8803668.

[41] N. Mehajabin, M. Pourazad, and P. Nasiopoulos, "SSIM assisted Pseudo-sequence-based prediction structure for light field video compression," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2020, pp. 1–2, doi: 10.1109/ICCE46568.2020.9042968.

[42] J. Hou, J. Chen, and L.-P. 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, Feb. 2019, doi: 10.1109/TCSVT.2018.2802943.

[43] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 4562–4566, doi: 10.1109/ICIP.2017.8297146.

[44] 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.

[45] X. Jiang, M. Le Pendu, and C. Guillemot, "Light field compression using depth image based view synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Hong Kong, Jul. 2017, pp. 19–24, doi: 10.1109/ICMEW.2017.8026313.

[46] M. Rizkallah, X. Su, T. Maugey, and C. Guillemot, "Graph-based transforms for predictive light field compression based on super-pixels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 1718–1722, doi: 10.1109/ICASSP.2018.8462288.

[47] T. Ebrahimi, I. Viola, P. Frossard, and H. Petric Maretic, "A graph learning approach for light field image compression," in *Proc. Appl. Digit. Image Process. XLI*, San Diego, CA, USA, Sep. 2018, pp. 126–137, doi: 10.1117/12.2322827.

[48] W. Ahmad, S. Vagharshakyan, M. Sjostrom, A. Gotchev, R. Bregovic, and R. Olsson, "Shearlet transform-based light field compression under low bitrates," *IEEE Trans. Image Process.*, vol. 29, pp. 4269–4280, Jan. 2020, doi: 10.1109/TIP.2020.2969087.

[49] A. Aggoun, "Compression of 3D integral images using 3D wavelet transform," *J. Display Technol.*, vol. 7, no. 11, pp. 586–592, Nov. 2011, doi: 10.1109/JDT.2011.2159359.

[50] P. Astola and I. Tabus, "WaSP: Hierarchical warping, merging, and sparse prediction for light field image compression," in *Proc. 7th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Tampere, Finland, Nov. 2018, pp. 1–6, doi: 10.1109/EUVIP.2018.8611756.

[51] M. B. de Carvalho, M. P. Pereira, G. Alves, E. A. B. da Silva, C. L. Pagliari, F. Pereira, and V. Testoni, "A 4D DCT-based lenslet light field codec," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 435–439, doi: 10.1109/ICIP.2018.8451684.

[52] R. Verhack, T. Sikora, L. Lange, R. Jongebloed, G. Van Wallendael, and P. Lambert, "Steered mixture-of-experts for light field coding, depth estimation, and processing," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kong, Jul. 2017, pp. 1183–1188, doi: 10.1109/ICME.2017.8019442.

[53] R. Verhack, T. Sikora, G. Van Wallendael, and P. Lambert, "Steered Mixture-of-Experts for light field images and video: Representation and coding," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 579–593, Mar. 2020, doi: 10.1109/TMM.2019.2932614.

[54] K. Komatsu, K. Takahashi, and T. Fujii, "Scalable light field coding using weighted binary images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 903–907, doi: 10.1109/ICIP.2018.8451812.

[55] *Light Field Toolbox v0.4*. Accessed: Apr. 2020. [Online]. Available: http://dgd.vision/Tools/LFToolbox/

[56] *JPEG Pleno Light Field Coding Common Test Conditions*, document ISO/IECJTC1/SC29/WG1N83029, Geneva, Switzerland, Mar. 2019.

[57] *EPFL Light-Field Image Dataset*. Accessed: Apr. 2020. [Online]. Available: http://mmspg.epfl.ch/EPFL-light-field-image-dataset

[58] 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.
[59] C. Brites, J. Ascenso, and F. Pereira, "Lenslet light field image coding: Classifying, reviewing and evaluating," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Feb. 27, 2020, doi: 10.1109/TCSVT.2020.2976784.

RICARDO J. S. MONTEIRO (Member, IEEE) received the degree in electrical engineering, in 2012, and the M.Sc. degree from the Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, Leiria, Portugal, in 2014. He is currently pursuing the Ph.D. degree with the University Institute of Lisbon (ISCTE-IUL), Lisbon, Portugal.

Since 2010, he has been a Researcher with the Instituto de Telecomunicações, Portugal. His current research interests include image and video processing, namely, multiview video and light field image processing and coding.

NUNO M. M. RODRIGUES (Member, IEEE) received the degree in electrical engineering, in 1997, the M.Sc. degree from Universidade de Coimbra, Coimbra, Portugal, in 2000, and the Ph.D. degree from Universidade de Coimbra, in 2009, in collaboration with the Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Since 1997, he has been with the Department of Electrical Engineering, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, Leiria, Portugal. Since 1998, he has been with the Instituto de Telecomunicações, Portugal, where he is currently a Senior Researcher. He has coordinated and participated as a Researcher in various national and international funded projects. His current research interests include digital signal and image processing, namely, image and video compression, point cloud and light field image and video compression, medical image compression, and many-core programming.

SÉRGIO M. M. FARIA (Senior Member, IEEE) was born in Portugal, in 1965. He received the Engineering and M.Sc. degrees in electrical engineering from Universidade de Coimbra, Portugal, in 1988 and 1992, respectively, and the Ph.D. degree in electronics and telecommunications from the University of Essex, U.K., in 1996.

He is currently a Full Professor with the Department of Electrical Engineering, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, Portugal, since 1990. He has collaborated in master courses with the Faculty of Science and Technology and the Faculty of Economy, Universidade de Coimbra. He is an Auditor with A3ES organization for Electrical and Electronic Engineering courses in Portugal. He is a Senior Researcher with the Instituto de Telecomunicações. His research interests include 2D/3D image and video processing and coding, motion representation, and medical imaging. In this field, he has published one book, edited two books and authored 13 book chapters, 28 journal articles, 125 referred conference papers, and two patents. He has been participating and he is responsible for several, national and international (EU), funded projects.

Dr. Faria has been a Scientific and Program Committee Member of many international conferences. He is a Reviewer for several international scientific journals and conferences, such as IEEE, IET, and EURASIP. He received the title of Agregado by the Instituto Superior Técnico, University of Lisbon, in 2014. He is an Area Editor of *Signal Processing: Image Communication*.

PAULO J. L. NUNES (Member, IEEE) received the degree in electrical and computers engineering from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1992, and the M.Sc. and Ph.D. degrees in electrical and computers engineering from IST, in 1996 and 2007, respectively.

He is currently an Assistant Professor with the Information Science and Technology Department, University Institute of Lisbon (ISCTE-IUL), Portugal, and a Senior Member of the Research Staff of Instituto de Telecomunicações, Portugal. His current research interests include 2D/3D image and video processing and coding, namely light field image and video processing and coding. He has coordinated and participated in various national and international (EU) funded projects and has acted as a Project Evaluator for the European Commission. He has contributed more than 60 articles.

Dr. Nunes acts often as a Reviewer for various conferences and journals and a member of the program and organizing committees of various international conferences.