# Improved Inter-Layer Prediction for Light Field Content Coding with Display Scalability

Caroline Conti[*a,b], Luís Ducla Soares[a,b], Paulo Nunes[a,b]
[a]Instituto de Telecomunicações Av. Rovisco Pais 1, 1049-001 Lisbon;
[b]Instituto Universitário de Lisboa (ISCTE-IUL), Av. das Forças Armadas,1649-026 Lisbon, Portugal

## ABSTRACT

Light field imaging based on microlens arrays – also known as plenoptic, holoscopic and integral imaging – has recently risen up as feasible and prospective technology due to its ability to support functionalities not straightforwardly available in conventional imaging systems, such as: post-production refocusing and depth of field changing. However, to gradually reach the consumer market and to provide interoperability with current 2D and 3D representations, a display scalable coding solution is essential.

In this context, this paper proposes an improved display scalable light field codec comprising a three-layer hierarchical coding architecture (previously proposed by the authors) that provides interoperability with 2D (Base Layer) and 3D stereo and multiview (First Layer) representations, while the Second Layer supports the complete light field content. For further improving the compression performance, novel exemplar-based inter-layer coding tools are proposed here for the Second Layer, namely: (i) an inter-layer reference picture construction relying on an exemplar-based optimization algorithm for texture synthesis, and (ii) a direct prediction mode based on exemplar texture samples from lower layers.

Experimental results show that the proposed solution performs better than the tested benchmark solutions, including the authors' previous scalable codec.

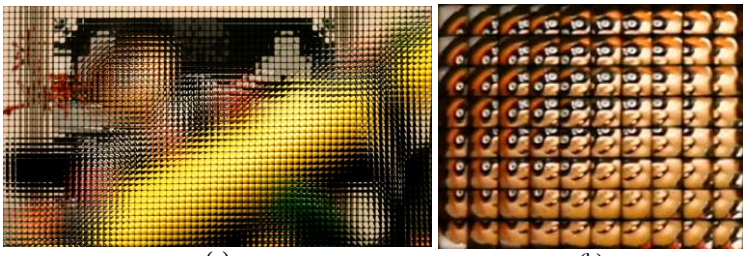
**Keywords:** Light field, plenoptic, holoscopic, scalable coding, HEVC, MV-HEVC

## 1 INTRODUCTION

The recent advances in optical and sensor manufacturing allow having richer forms of visual data, where not only the spatial information about the three-dimensional (3D) scene is represented but also angular viewing direction – the so-called four-dimensional (4D) light field/radiance sampling[1]. In the context of light field (LF) imaging technologies, the approach based on a single-tier camera equipped with a microlens array[2] (also known as holoscopic[3], plenoptic[4], and integral imaging[5,6]) has become a promising approach, being applicable in many different areas of research, such as, 3D television[3,7], richer photography capturing[8,9], biometric recognition[10], and medical imaging[6].

Among the advantages of employing a LF imaging approach based on microlens arrays is the ability to open new degrees of freedom in terms of content production and manipulation, supporting functionalities not straightforwardly available in conventional imaging systems, namely: post-production refocusing, changing depth-of-field, and changing viewing perspective. Recognizing these new and exciting possibilities, novel initiatives on LF image and video coding standardization are also emerging; notably, the JPEG committee has recently started the JPEG Pleno standardization initiative[11], and the MPEG group has also started the third phase of Free-viewpoint Television (FTV) targeting free navigation and full parallax imaging applications[12]. One of the objectives of these new initiatives[11,12] is to identify the requirements and challenges in light field systems, as well as to understand the users' needs in terms of light field visualization and content interaction. The challenge to provide a light field representation with convenient spatial resolution and viewing angles requires handling a huge amount of data and thus efficient coding is of utmost importance. In addition to this, as the imaging technology moves toward richer representations, novel data representations are essential to support the new applications and functionalities that arise[11]. In this sense, a scalable coding architecture is desirable to accommodate in a single compressed bitstream a variety of sub-bitstreams appropriate for users with different preferences/requirements and various application scenarios: from the user who wants to have a simple 2D version of the light field content to be visualized in a conventional 2D display; to the user who wants immersive and interactive 3D

[*] caroline.conti@lx.it.pt; phone +351 218 418 164, www.it.pt/cconti



(a) (b)

Figure 1 Light field image captured with a 250 µm pitch microlens array: (a) Full image with resolution of 1904×1064; (b) Enlargement of 280×224 pixels showing a sub-array of micro-images

visualization by using a more advanced display technology, such as an integral imaging display[3,7] or a head mounted display for augmented and virtual reality[13,14].

Therefore, this paper proposes an improved display scalable light field codec comprising a three-layer hierarchical coding architecture (previously proposed by the authors[15]) that provides interoperability with 2D (Base Layer) and 3D stereo and multiview (First Layer) representations, while the Second Layer supports the complete LF content. For further improving the compression performance in the Second Layer, novel exemplar-based inter-layer coding tools are proposed here for the Second Layer, namely: (i) an inter-layer compensated prediction using an inter-layer reference (ILR) picture that is constructed relying on an exemplar-based[16] optimization algorithm for texture synthesis, and (ii) a direct prediction mode based on exemplar texture samples from lower layers.

The remainder of this paper is organized as follow: Section 2 reviews the relevant work on coding solutions for LF content; Section 3 describes the used display scalable coding architecture; Section 4 presents the two novel exemplar-based inter-layer coding tools for improving the compression performance; Section 5 presents the test conditions and experimental results; and, finally, Section 6 concludes the paper.

## 2 RELATED WORK

Several LF image coding schemes have been recently proposed in the literature which try to take advantage of its particular planar intensity distribution to achieve more efficient compression. Notably, as a result of the used optical system, the LF raw image corresponds to a two-dimensional (2D) array of micro-images (MIs) (see Figure 1a), and a significant cross-correlation exists between neighboring MIs (see Figure 1b). In terms of the possible different ways to organize the LF data for coding and transmission, these LF image coding solutions can be categorized in three main types of approaches: i) LF raw data-based approach[17–22], ii) multiview-based LF coding[23–26], and iii) sub-sampled grid of MIs plus disparity approach[27–29].

The LF raw data-based approach corresponds to encoding and transmitting the light field image in its entirety, represented as a 2D grid of MIs. For this, a special prediction scheme is needed to exploit the non-local spatial redundancy between different MIs. Following this approach, the authors[18,19] proposed to include a scheme for self-similarity (SS) estimation and compensation[17] to improve the performance of HEVC standard for LF coding, while taking advantage of the flexible partition patterns used in this type of video codecs. More recently[20], the SS estimation and compensation was extended for bi-prediction to further improve the coding efficiency. A similar scheme was also proposed in Li *et al.*[21] to support SS estimation with multiple prediction hypothesis. In Lucas *et al.*[22], a prediction framework based on locally linear embedding was included into HEVC for light field image coding. However, although these coding schemes achieved significant compression gains when compared to the existing state-of-the-art image coding alternatives, transmitting the entire light field data without a scalable bitstream may represent a serious problem since the user needs to wait until the entire content of each picture arrives before it can be visualized, independently of the users' display type.

Alternatively, other schemes followed a multiview-based approach and proposed to extract the viewpoint images from the LF content to be represented as multiview content[23,24], or a pseudo video sequence[25,26]. A viewpoint image represents an orthographic projection of the complete captured scene in a particular direction, and can be constructed by simply extracting one pixel with the same relative position from each MI. Following this approach, the set of viewpoint images was then encoded using multiview video coding (MVC)[23,24], H.264/AVC[25], or HEVC[26] standards. Since rendering viewpoint images usually produces very low resolution images with aliasing[30], an alternative to the multiview

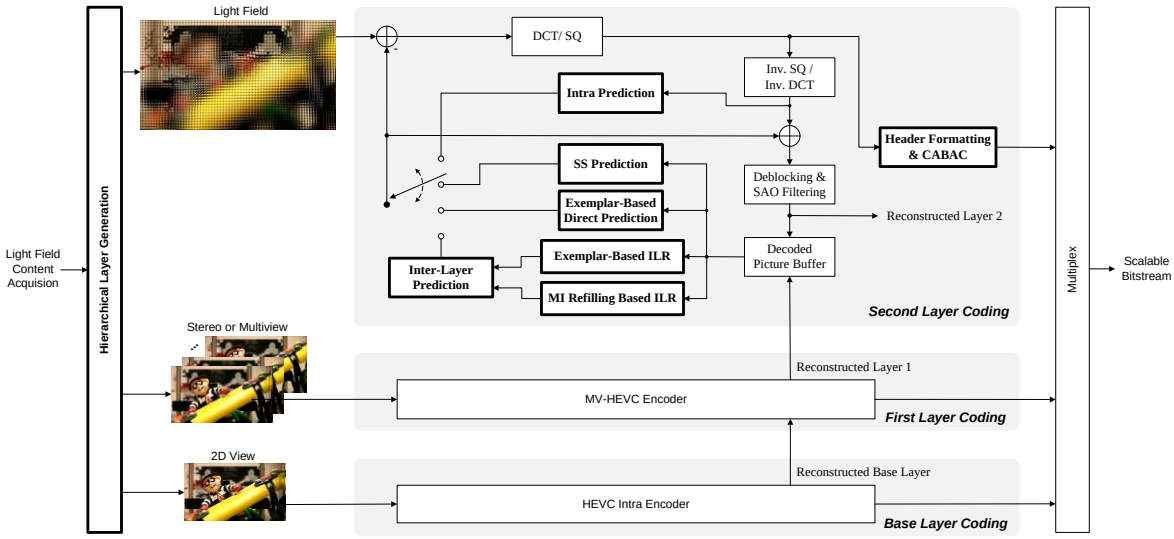



Figure 2 Block diagram of the proposed scalable LF image codec. The Second Layer is coded with the proposed LF enhancement layer encoder

representation based on these viewpoint images was proposed in the authors' previous work[15,31]. The architecture and coding scheme of this solution will be reviewed in Section 3.

Other coding schemes proposed to represent the light field data by a sub-sampled set of MIs with their associated disparity information [27–29]. In this case, the grid of MIs was sub-sampled to remove the redundancy between neighboring MIs and to achieve compression. Thus, only the remainder set of MIs and associated disparity were encoded and transmitted. At the decoder side, the light field data was reconstructed by simply applying a disparity shift[27,29] or by using a Depth Image Based Rendering (DIBR) algorithm modified to support the multiple MIs as input views[28], and followed by an inpainting algorithm to fill in the missing areas. However, in real-world images, the disparity/depth information is estimated from the acquired LF raw data, which introduces inaccuracies. Hence, the quality of the reconstructed MIs – and, consequently, the quality of rendered views – is severely affected by these inaccuracies.

## 3 SCALABLE LIGHT FIELD CODING ARCHITECTURE

The coding architecture adopted in this paper, which has been previously proposed by the authors[15], is built upon a predictive and three-layered scalable approach, as depicted in Figure 2. The Base Layer contains a sub-sampled portion of the light field raw data that represents a 2D version of the LF content, which can be then used to deliver LF content to 2D displays. This Base Layer is coded with a conventional HEVC Intra encoder to provide backward compatibility with a state-of-the-art coding solution, and the reconstructed frames are then used for coding the higher layers. Following this, the First Layer represents the necessary information to obtain an additional view (representing a stereo pair) or various views (representing multiview content) from the LF content, which can be visualized by using a stereo or an autostereoscopic display. This First Layer can be encoded by using a standard stereo or multiview coding solution, such as MVC[32] or the 3D video coding extensions of HEVC[33]. With these solutions, inter-view prediction can be used to improve the coding efficiency between the Base Layer and the First Layer as well as within the First Layer. For the work presented in this paper, the multiview extension of HEVC, MV-HEVC, is adopted. Finally, the Second Layer represents the additional information necessary to support immersive LF content visualization. This Second Layer is then encoded with the proposed LF enhancement layer encoder depicted in Figure 2, which is also based on the HEVC coding framework.

The basic blocks (emphasized in bold in Figure 2) of the proposed scalable light field codec (SLFC) are explained in the following subsections.

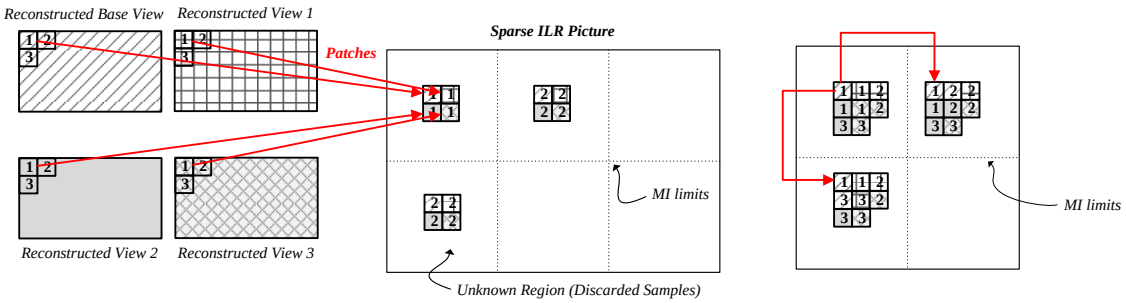



Figure 3 Inter-Layer Prediction using the previously proposed MI refilling based ILR construction[15]: Patch Remapping step; and (b) MI Refilling step

### 3.1 Hierarchical Layer Generation

Several algorithms to generate 2D view images from a LF image have been proposed in the literature, mainly in the context of richer 2D image capturing systems[4,34–36]. In this paper, the algorithm referred to as Basic Rendering[4] is considered to generate the content for the first two hierarchical layers.

Briefly, since each micro-image can be seen as a low resolution view of the scene, the idea behind the Basic Rendering algorithm is to choose suitable portions (patches) from each micro-image which can be stitched to properly compose a 2D view image. Then, as explained by Georgiev and Lumsdaine[4], the process of generating a 2D view image can be controlled by the following two main parameters:

- *Size of the patch*: It is possible to control the plane of focus in the generated 2D view image (i.e., which objects will appear in sharp focus) by choosing a suitable patch size to be extract from each MI. Therefore, by varying the patch size, different content will be generated for the first two hierarchical layers;
- *Position of the patch*: By varying the relative position of the patch in the MI, it is possible to generate multiple 2D views with different horizontal and vertical viewing angles (i.e., different scene perspectives).

### 3.2 Intra Prediction

HEVC Intra prediction is available as an alternative prediction when selecting the most efficient mode for encoding a coding block in the Second Layer (i.e., in the LF picture). The decision between the different available prediction modes is made in a rate-distortion (RD) optimization manner, as in conventional HEVC.

### 3.3 SS Prediction

Since the Second Layer contains the full LF image, the SS compensated prediction[19] can be also used to exploit the existing redundancy and to improve coding efficiency within the Second Layer. For this, a block matching estimation is used to find the best prediction for a coding block within the previously coded and reconstructed area of the current frame. As a result, the residual information and a displacement vector are coded and sent to the decoder.

### 3.4 Inter Layer Prediction

This prediction mode is used to further improve the Second Layer coding efficiency by removing redundancy between the multiview and the LF content. For this, two ILR are constructed, which can be then used as new reference pictures for employing a compensated prediction when encoding the LF image. These are the previously proposed MI refilling based ILR[15] and the new exemplar-based ILR, as further explained in the following.

### 3.5 MI Refilling Based ILR

As previously proposed by the authors[15], the MI refilling based ILR picture is built by using the following two steps.

#### Patch Remapping

The input for this step is the coded and reconstructed views from the two lower layers as well as the parameters used for acquiring these views (such as the resolution of micro-images, patch sizes, and positions).

Although most of the LF information is discarded when rendering each view in the hierarchical layer generation block (see Figure 2), it is still possible to re-organize the reconstructed texture information into its original positions in the LF image. Therefore, the patch remapping simply corresponds to an inverse process of the Basic Rendering algorithm[4]. More

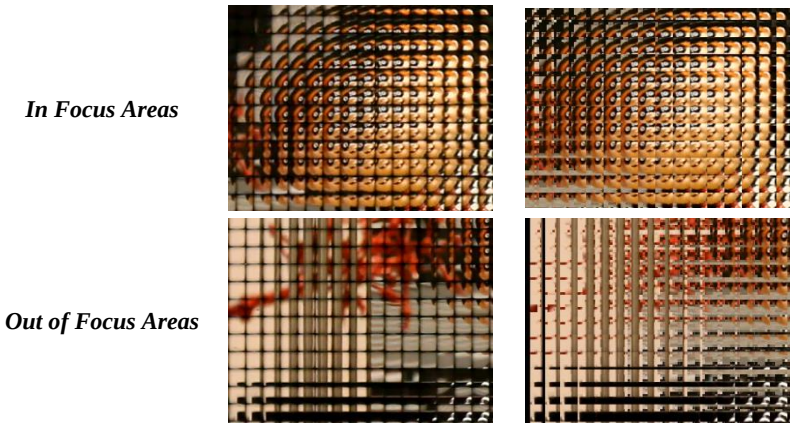



Figure 4 Comparison of the original LF image (a) and the resulting MI refilling based ILR picture (b) for a particular patch size. The reconstruction is much more accurate for areas of the rendered image where the selected patch size is focused (top)

specifically, it corresponds to an inverse mapping (referred to as remapping) of the patches from all rendered and reconstructed views to their original positions in the LF image, as illustrated in Figure 3a. A template for the LF image assembles all patches, and the output is then the sparse ILR picture, as illustrated Figure 3.

**MI Refilling**

The input for this step is the sparse ILR picture generated by the Patch Remapping. Basically, the MI refilling aims at emulating the significant cross-correlation existing between neighboring MIs so as to fill the holes in the sparse ILR picture (see Figure 3a) as much as possible.

Since there is no information about the disparity/depth between objects in neighboring MIs, the disparity is defined in a patch-based manner. Then, for each MI in the sparse ILR picture, an available set of pixels (a patch) is copied to a suitable position in a neighboring micro-image that is shifted by the size of the patch, as illustrated in Figure 3b. Additionally, the number of neighboring micro-images where the patch may be copied to depends on the size of the micro-image and the size of the patch.

### 3.6 Exemplar-based ILR

A characteristic of the MI refilling process is that the constructed ILR picture is considerably more accurate in areas where the generated 2D views are in focus compared to out of focus areas. This is illustrated in Figure 4, and happens since the actual disparity is not known in the out of focus areas and, consequently, is assumed to be given by the patch size. Motivated by this fact, an improved ILR picture construction process, referred to as exemplar-based ILR, is proposed in this paper to further improve the Inter-Layer Prediction efficiency. Details are given in Section 4.1.

### 3.7 Exemplar-Based Direct Prediction

This new prediction mode aims at exploiting the redundancy between the First and Second Layers to find a prediction block and, then, to implicitly derive an inter-layer vector for encoding the current block in the Second Layer picture. As a result, no vector needs to be transmitted and the decoder can simply use the same process for inferring the vectors to carry out the compensated prediction using the decoded residual samples. Similarly to the conventional HEVC merge mode[37], an index is transmitted (together with the coded residual information), which is used to distinguish the Exemplar-based Direct Prediction from the conventional HEVC merge mode. The process to derive the implicit inter-layer vector is presented with further detail in Section 4.2.

### 3.8 Header formatting & Context-Adaptive Binary Arithmetic Coding (CABAC)

Additional high level syntax elements are carried through the HEVC bitstream to support this type of scalability. These are basically acquisition information (e.g., MI resolution, patch sizes and positions) and dependency information (for signaling the use of novel reference pictures). Finally, residual and prediction mode signaling are entropy coded using CABAC.

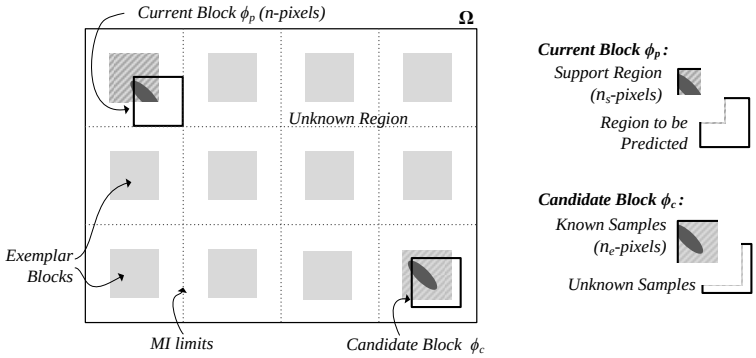



Figure 5 Exemplar-based ILR Picture Construction: For each block $\phi_P$ in the sparse ILR picture, the best candidate block, $\phi_c^{best}$, is derived by solving the optimization problem in (1).

# 4 NOVEL EXEMPLAR-BASED INTER-LAYER CODING TOOLS

Two exemplar-based coding tools are proposed in this section to improve the coding performance in the Second Layer. These are: i) the exemplar-based ILR picture construction; and ii) the exemplar-based direct prediction.

## 4.1 Exemplar-based ILR Picture Construction

This section describes the process that is carried out in the exemplar-based ILR block depicted in Figure 2.

**Input Information**

Similarly to the previously proposed MI refilling, the input information for this process is the sparse ILR picture that is built by using the Patch Remapping shown in Figure 3a.

**Problem Formulation**

The basic idea for constructing the improved ILR picture is to find a good estimation to fill in the unknown data in the sparse ILR picture. This is clearly an ill-posed problem; however, it is still possible to obtain a realistic approximate solution by imposing additional constraints coming from the physics of the problem. This is done here by using the prior knowledge that neighboring MI samples present significant cross-correlation, and for this reason, it is likely to find the unknown region of a particular MI in an area of neighboring MIs.

In the given problem, the unknown pixels are initially set to zero. Hence, given the sparse ILR picture shown in Figure 3a, divide it into blocks with $n$-pixels. Each block $\phi_P$ is formed by a $n_s$-pixel set of known samples – referred to as the support region – and a $(n - n_s)$-pixel set of unknown samples to be predicted as shown in Figure 5. Hence, each block can be represented as the product of a texture column vector, **y**, by a binary mask, **S**, in which all but $(n - n_s)$ samples have value equal to one. The mask **S** is here represented as an $n \times n$ identity matrix with the respective $(n - n_s)$ unknown diagonal samples set to zero.

To fill in the unknown region of $\phi_P$ the first step is to find a candidate block that best agrees with the support samples of $\phi_P$. Thus, let $\phi_c$ be a $n$-pixel candidate block that is inside a neighborhood area, **Ω**, comprising samples from $K$ neighbor MIs (i.e., $\mathbf{\Omega} = \{\mathrm{MI}_k\}_{k=1\ldots K}$). The candidate block $\phi_c$ might be also formed by known and unknown samples. Considering that $\phi_c$ comprises a $n_e$-pixel region of known samples, the candidate block can be similarly represented as the product of a texture column vector, **x**, by an identity matrix, **E**, with $(n - n_e)$ diagonal samples set to zero.

Therefore, let **A** be a binary diagonal matrix that represents the samples from $\phi_p$ and $\phi_c$ that overlap, simply given by $\mathbf{A} = \mathbf{SE}$. The best candidate block, $\phi_c^{best}$, can then be found by solving the optimization problem in (1), which comprises, respectively, a data-fitting term and a sparseness prior function. The former term tries to find the best match within the region where $\phi_p$ and $\phi_c$ overlap, while the latter term penalizes candidate blocks whose $n_e$-pixel region is too small (note that $\mathbf{I}_n$ corresponds to a $n \times n$ identity matrix).

$$\min_{\mathbf{x},\mathbf{A}} \|\mathbf{A}(\mathbf{y} - \mathbf{x})\|_1 + \lambda \times \|diag(\mathbf{I}_n - \mathbf{A})\|_0 \quad (1)$$

Since the border of the MIs typically exhibits high intensity variations (mainly due to vignetting), a further constraint is imposed to the problem formulated in (1) to guarantee that these high frequency samples from the borders of an MI sample, $\mathrm{MI_k} \subset \mathbf{\Omega}$, do not affect negatively the synthesized patterns, which is to solve the problem in (1), subjected to: $\mathbf{x} \in \mathrm{MI_k} \wedge$

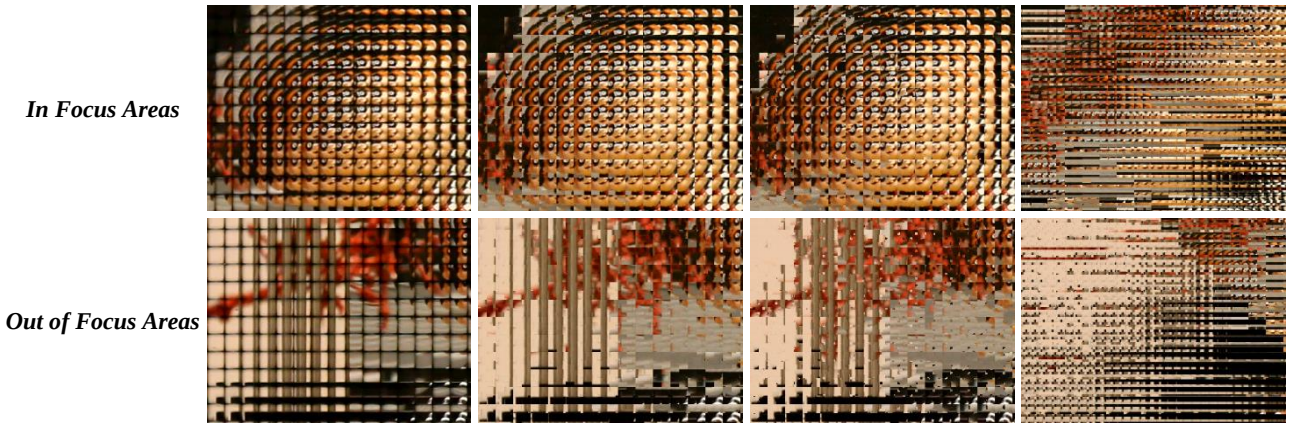



Figure 6 Comparison of the original LF image (a) and the resulting exemplar-based ILR picture when: (b) 50 %, (c) 80 %, and (d) more than 90 % of the LF information is discarded when generating the content for the two lower hierarchical layers. The reconstruction is also shown for in focus and out of focus areas just for comparison with the ILR picture in Figure 4

$\mathbf{x} \notin \mathrm{MI}_{\mathrm{m}\neq\mathrm{k}}$. In the experimental results presented in this paper (Section 5), the λ value is selected empirically and the $\phi_p$ and $\phi_c$ blocks size is selected to be a quarter of the size of an MI size.

**Texture synthesis algorithm**

Once the best patch sample $\phi_c^{best}$ is obtained by solving (1), the synthesized region is derived by simply copying the samples of the region defined by $\mathbf{E} - \mathbf{A}$. If there are still patches with unknown samples, the optimization process is iteratively repeated until all unknown samples are filled in or until the number of unknown samples stabilizes. Thus, at each iteration, the values of **x** and **A** are updated with the found values. If there are still unknown samples at the end of the algorithm, the MI refilling process may be used to fill the remainder holes.

The presented exemplar-based solution is chosen due to its simplicity and effectiveness for tackling the proposed problem. However, it should be noticed that, since the size of the exemplar blocks affects the ill-posedness of the problem, the more information of the LF image is discarded when generating the content for the Base and First Layers, the worst the quality of the built ILR picture will be. As an illustrative example of this fact, Figure 6 presents the resulting ILR picture when varying the amount of information which is discarded from 50 % (see Figure 6b) up to more than 90 % (see Figure 6d), compared with the original LF image (see Figure 6a). Better solutions for dealing with the case in Figure 6d might still be formulated, for instance, by adding an edge-preserving regularizer in (1). However, it will be left for future work.

### 4.2 Exemplar-based Direct Prediction

This section describes the process to implicitly derive an inter-layer vector when encoding the Second Layer by using the texture information from the previously encoded layers, as illustrated in Figure 7. This process can be divided in the following two steps.

**Co-located Block Derivation**

The input for this step is also the sparse ILR picture comprising a sparse set of known samples, referred to as exemplar blocks in Figure 7. Therefore, the block from the sparse ILR picture with the same size and co-located position to the current block being coded is derived and referred to as co-located block in Figure 7.

**Inter-Layer Vector Estimation**

Similarly to the template matching algorithm[38], a matching algorithm is used to find the 'best' candidate predictor for the current block. For this, the candidate block that best agrees with the co-located block determined in the previous step is chosen in the previously coded and reconstructed area of the LF picture being coded. More specifically, the best candidate block is chosen by matching only the known exemplar samples of the co-located block over a causal search window in the LF picture, as shown in Figure 7, since these are the only samples available at the decoder at the corresponding decoding time.

Therefore, let **y** be a column vector containing the $p$-pixel samples of the co-located block in the sparse ILR picture, where only the $p_e$-pixel samples, i.e., the exemplar samples, are known at decoding time (see Figure 7). Also, let **W** be the search

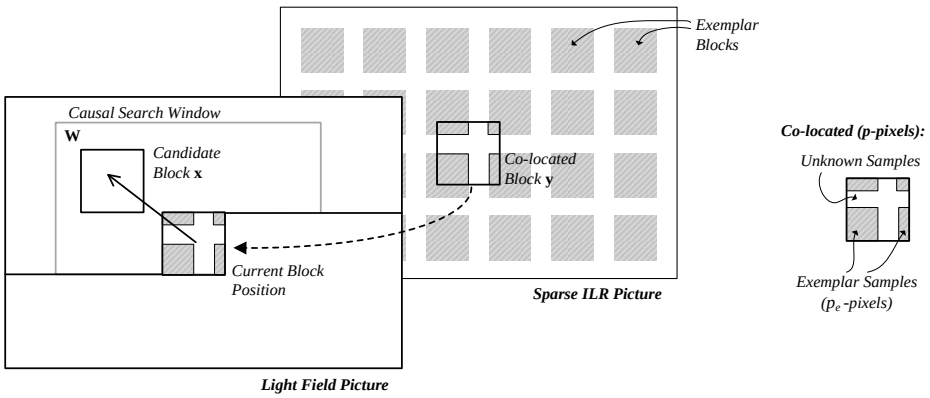



Figure 7 Exemplar-based direct prediction. The best match between the co-located block and a candidate block (within the causal search Window **W** in the LF Picture) allows finding an implicit inter-layer vector for the current coding block.

window, and let $\mathbf{x} \subset \mathbf{W}$ be a column vector containing the $p$-pixel samples of a candidate predictor block in the current layer. Since **y** contains $(p - p_e)$ unknown samples, it can be modeled as $\mathbf{y} = \mathbf{A}\mathbf{x}$, where **A** is a binary mask in which only the corresponding known $p_e$ sample positions are non-zero. Thus, **A** can be represented as an $p \times p$ identity matrix whose $(p - p_e)$ unknown diagonal samples are set to zero. Finally, since the mask **A** is known a priori, the best candidate block, $\mathbf{x}_{best}$, can be simply found by the matching algorithm in (2), where the sum of absolute differences has been used as the matching criteria.

$$\mathbf{x}_{best} = \underset{\mathbf{x} \subset \mathbf{W}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 \tag{2}$$

Finally, once $\mathbf{x}_{best}$ is obtained, the displacement vector between $\mathbf{x}_{best}$ and the current block is derived, which is the output of this process.

# 5 PERFORMANCE ASSESSMENT

This section assesses the performance of the proposed SLFC codec. For this purpose, the test conditions and tested coding solutions are firstly introduced and, then, the obtained results are presented and discussed.

## 5.1 Experimental Setup

To evaluate the performance of the proposed SLFC codec, six light field images with different spatial and micro-image resolutions are considered so as to achieve representative RD results. These are (see Figure 8): *Fredo*[39], *Seagull*[39], *Laura*[39], *Demichelis Spark*[40] (first frame of a video sequence with identical name), *Robot 3D*[40], and *Plane and Toy*[40] (frame number 123 from a video sequence with identical name). The original tested images were rectified so as to have all micro-images with integer number of pixels, and they were then converted to the YCbCr 4:2:0 color format.

To generate the content for the 2D, stereo and multiview layers, the six test images were processed with the Basic Rendering algorithm[4]. In this process, a set of regular spaced 2D view images were generated – one for the Base Layer and the remainder for the First Layer. In addition, the patch size is selected to represent the case where the main object of the scene is in focus. This way, the chosen patch sizes and positions for each LF test image are summarized in Table 1. As can be seen (Table 1), two sets with a different number of views are considered for each LF image, one with 9×1 views (only in horizontal positions) and another one with 9×3 views.

For these tests, the reference software for the multiview extension of HEVC (MV-HEVC) version 12.0[41] was used as the benchmark, as well as the base software for implementing the proposed SLFC codec. Each of the abovementioned test image was thus encoded using four different quantization parameter (QP) values: 22, 27, 32, and 37, according to the common test conditions defined in[42]. The same QP value was used for coding all hierarchical layers.

For evaluating the RD performance of the proposed LF enhancement layer encoder, the distortion, in terms of PSNR, of only the Second Layer is considered. The rate is presented in bits per pixel (bpp), which is calculated as the total number of bits needed for encoding all scalable layers divided by the number of pixels in the LF raw image.

*Fredo* – 7104 × 5328
$MI_{resol}$ = 74 × 74, 96 × 72 MI-grid

*Seagull* – 7104 × 5328
$MI_{resol}$ = 74 × 74, 96 × 72 MI-grid

*Laura* – 7104 × 5328
$MI_{resol}$ = 74 × 74, 96 × 72 MI-grid

*Demichelis Spark* – 2850 × 1558
$MI_{resol}$ = 38 × 38, 75 × 41 MI-grid


*Robot 3D* – 1904 × 1064
$MI_{resol}$ = 28 × 28, 68 × 38 MI-grid

*Plane and Toy* – 1904 × 1064
$MI_{resol}$ = 28 × 28, 68 × 38 MI-grid

Figure 8 Example of a central view rendered from each LF test image (with the corresponding characteristics below each image)

Alternatively, to analyze the performance in terms of the quality for additional views synthesized from the reconstructed LF content in the Second Layer, the distortion is also measured in terms of the average PSNR and the average SSIM values over a set of views rendered from the reconstructed and from the original LF image. These two quality metrics are referred to as, respectively, Rendering-dependent PSNR and Rendering-dependent SSIM. To have a representative number of rendered views, a set of 9 views were rendered from viewpoint positions equally distributed horizontal and vertical directions. The standard deviation for each result was also used to measure the confidence of the presented average values. For rendering the views, the same algorithm used for generating content for each hierarchical layer was used (i.e., Basic Rendering[4]).

### 5.2 Benchmark Coding Solutions

The next subsections present and analyze the performance of the SLFC (proposed) codec against the following benchmark solutions:

- *MV-based (HEVC Inter P)*: In this case, a multiview-based coding approach is considered, where all viewpoint images are extracted from the light field image and, then, are encoded as a pseudo video sequence[26]. In this case, HEVC is used with the "Low delay P, main" configuration[42] and the largest coding unit size was set to 16×16, since the resolution of each viewpoint image is considerably smaller than what is typically encoded with HEVC. After testing various orders for scanning the viewpoint images (i.e., raster, parallel, zigzag, and spiral), the spiral order is presented as it achieved the best performance among the ones that were tested.
- *MV-based (HEVC Inter B)*: The pseudo video sequence of viewpoint images (scanned in spiral order) is also encoded using HEVC with "Random Access, main" configuration[42].
- *SLFC (Simulcast)*: This scalable codec corresponds to the benchmark for the simulcast case of the proposed SLFC coding architecture. Hence, all pictures from each hierarchical layer are coded independently as Intra frames with standard MV-HEVC solution, using "All Intra, Main" configuration[42].
- *SLFC (Previous Solution)*: In this case, the picture from each hierarchical layer is coded with the previously proposed SLFC codec proposed[31] where only the conventional HEVC Intra, SS and the previous Inter-Layer prediction with MI refilling are enabled. In this case, the Base and First Layers are encoded as Intra Frames and the Second Layer is encoded as an Inter P frame.

For the *SLFC (Proposed Solution)*, all the views are encoded as Intra frames and the Second Layer is encoded as an Inter B frame. Notice that, other configurations for encoding the content in the First Layer are still possible, notably, by enabling inter-view prediction (coding as P or B frames). However, due to the large number of possible test condition combinations, these additional results will be left for future work, Furthermore, a study of the influence of varying the coding configuration in the lower layer on the performance of the proposed solution will be also performed in the future.

Table 1 Tested Conditions: Patch Sizes and Patch Positions (related to the MI center) for generating views for the lower hierarchical layers

| Images | Patch Size (Focus Plane)s | Views Grid | Patch Positions (Views Perspective) |
|---|---|---|---|
| *Fredo* | 10 | 9×1 | {(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)} |
| | | 9×3 | {(-24,-10), (-18,-10), (-12,-10), (-6, -10), (0, -10), (6, -10), (12, -10), (18, -10), (24, -10)}<br>{(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)}<br>{(-24,10), (-18,10), (-12,10), (-6,10), (0,10), (6,10), (12,10), (18,10), (24,10)} |
| *Seagull* | 9 | 9×1 | {(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)} |
| | | 9×3 | {(-24,-10), (-18,-10), (-12,-10), (-6, -10), (0, -10), (6, -10), (12, -10), (18, -10), (24, -10)}<br>{(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)}<br>{(-24,10), (-18,10), (-12,10), (-6,10), (0,10), (6,10), (12,10), (18,10), (24,10)} |
| *Laura* | 10 | 9×1 | {(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)} |
| | | 9×3 | {(-24,-10), (-18,-10), (-12,-10), (-6, -10), (0, -10), (6, -10), (12, -10), (18, -10), (24, -10)}<br>{(-24,0), (-18,0), (-12,0), (-6,0), (0,0), (6,0), (12,0), (18,0), (24,0)}<br>{(-24,10), (-18,10), (-12,10), (-6,10), (0,10), (6,10), (12,10), (18,10), (24,10)} |
| *Demichelis Spark* | 12 | 9×1 | {(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)} |
| | | 9×3 | {(-8,-10), (-6, -10), (-4, -10), (-2, -10), (0, -10), (2, -10), (4, -10), (6, -10), (8, -10)}<br>{(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)}<br>{(-8,10), (-6,10), (-4,10), (-2,10), (0,10), (2,10), (4,10), (6,10), (8,10)} |
| *Robot 3D* | 4 | 9×1 | {(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)} |
| | | 9×3 | {(-8,-4), (-6, -4), (-4, -4), (-2, -4), (0, -4), (2, -4), (4, -4), (6, -4), (8, -4)}<br>{(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)}<br>{(-8,4), (-6,4), (-4,4), (-2,4), (0,4), (2,4), (4,4), (6,4), (8,4)} |
| *Plane and Toy* | 4 | 9×1 | {(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)} |
| | | 9×3 | {(-8,-4), (-6, -4), (-4, -4), (-2, -4), (0, -4), (2, -4), (4, -4), (6, -4), (8, -4)}<br>{(-8,0), (-6,0), (-4,0), (-2,0), (0,0), (2,0), (4,0), (6,0), (8,0)}<br>{(-8,4), (-6,4), (-4,4), (-2,4), (0,4), (2,4), (4,4), (6,4), (8,4)} |

### 5.3 Overall Rate-Distortion Performance

Tables 2 and 3 present the RD performance in terms of the Bjøntegaard Delta[43] in PSNR (BD-PSNR) and rate (BD-BR) with respect to the benchmarks solutions for all test images in Figure 8.

From these results, the following conclusions can be derived:

- *Comparison with MV-based approaches*: It can be seen (Tables 2 and 3) that the proposed SLFC solution architecture presents significantly better RD performance than the multiview arrangement of the viewpoint images, for both tested MV-based configurations (*HEVC Inter P* and *HEVC Inter B*) and view arrangements (9×1 and 9×3 views). The BD gains of the *SLFC (Proposed Solution)* go up to 8.68 dB (PSNR) and -79.69 % (BR) when compared to the *MV-based (HEVC Inter P)*, and 8.46 dB (PSNR) and -77.89 % (BD) when compared to *MV-based (HEVC Inter B)*. For the test image *Demichelis Spark* with 9×3 views in the lower layers (see Table 3), the *MV-based (HEVC Inter B)* performs better than the *SLFC (Proposed Solution)*. However, it should be noticed that the worse performance of the *SLFC (Proposed Solution)* in this case is due to the larger set of views (9×3) that are independently encoded as Intra Frames, instead of enabling the inter-view prediction to improve the performance as in the *MV-based (HEVC Inter B)* (in this case, the viewpoint images are coded as B frames).
- *Comparison with SLFC (Simulcast)*: The proposed SLFC RD performance is significantly better than the *SLFC (Simulcast)* independently of the used view arrangements in the lower layers (see Tables 2 and 3). The gains are much more expressive for test images with higher MI resolution, where the BD-PSNR gain goes up to 3.00 dB and the BD-BR to -44.54 % (for *Seagull*). These gains are justified by the efficiency in exploiting the redundancy between the layers (using the proposed inter-layer coding tools), as well as the efficiency in exploiting the correlations within the Second Layer (using the SS prediction).

Table 2 BD-PSNR and BD-BR performance of the proposed SLFC codec against the benchmarks when considering 9×1 views in the lower hierarchical layers

| *Test Image* | *MV-based (HEVC Inter P)* | | *MV-based (HEVC Inter B)* | | *SLFC (Simulcast).* | | *SLFC (Previous Solution)*[31] | |
|---|---|---|---|---|---|---|---|---|
| | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** |
| *Fredo* | 8.68 dB | -79.69 % | 8.46 dB | -77.89 % | 2.91 dB | -41.30 % | 0.23 dB | -4.40 % |
| *Seagull* | 5.66 dB | -72.13 % | 5.43 dB | -71.52 % | 3.00 dB | -44.54 % | 0.45 dB | -9.47 % |
| *Laura* | 6.32 dB | -63.45 % | 5.70 dB | -63.60 % | 2.51 dB | -33.12 % | 0.22 dB | -3.78 % |
| *Demichelis Spark* | 4.15 dB | -69.33 % | 2.78 dB | -53.88 % | 1.17 dB | -28.90 % | 0.39 dB | -10.48 % |
| *Robot 3D* | 6.90 dB | -56.39 % | 5.45 dB | -52.40 % | 1.07 dB | -12.62 % | 0.08 dB | -1.04 % |
| *Plane and Toy* | 5.75 dB | -58.15 % | 4.02 dB | -48.54 % | 1.46 dB | -20.53 % | 0.21 dB | -3.32 % |
| Average | 6.24 dB | -66.52 % | 5.30 dB | -61.31 % | 2.02 dB | -30.17 % | 0.26 dB | -5.42 % |

Table 3 BD-PSNR and BD-BR performance of the proposed SLFC codec against the benchmarks when considering 9×3 views in the lower hierarchical layers

| *Test Image* | *MV-based (HEVC Inter P)* | | *MV-based (HEVC Inter B)* | | *SLFC (Simulcast).* | | *SLFC (Previous Solution)*[31] | |
|---|---|---|---|---|---|---|---|---|
| | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** | **BD-PSNR** | **BD-BR** |
| *Fredo* | 7.42 dB | -73.22 % | 7.10 dB | -70.79 % | 2.85 dB | -39.50 % | 0.41 dB | -7.36 % |
| *Seagull* | 4.60 dB | -63.01 % | 4.37 dB | -62.10 % | 2.74 dB | -40.60 % | 0.40 dB | -7.83 % |
| *Laura* | 5.05 dB | -52.94 % | 4.47 dB | -52.92 % | 2.75 dB | -34.76 % | 0.33 dB | -5.17 % |
| *Demichelis Spark* | 0.47 dB | -11.41 % | -1.05 dB | 33.33 % | 0.77 dB | -17.87 % | 0.27 dB | -6.16 % |
| *Robot 3D* | 4.73 dB | -41.38 % | 3.40 dB | -35.49 % | 1.60 dB | -17.18 % | 0.14 dB | -1.65 % |
| *Plane and Toy* | 4.05 dB | -43.79 % | 2.32 dB | -30.54 % | 1.64 dB | -21.33 % | 0.17 dB | -2.46 % |
| Average | 4.39 dB | -47.63 % | 3.44 dB | -36.42 % | 2.06 dB | -28.54 % | 0.29 dB | -5.10 % |

- *Comparison with SLFC (Previous Solution)*: Comparing this solution with the complete *SLFC (Proposed Solution)*, it can be seen that improved RD performance can be attained by making use of the proposed exemplar-based coding tools. In this case, the BD gains of the *SLFC (Proposed Solution)* go up to 0.45 dB (PSNR) and -9.47 % (BR).

### 5.4 Quality of Additional Rendered Views

In order to assess the performance of the proposed scalable coding architecture regarding the quality of rendered views, the RD performance of the *SLFC (Proposed Solution)* is here presented in terms of the Rendering-dependent PSNR and SSIM metrics (as explained in Section 5.1) and compared to the *SLFC (Simulcast)* and *SLFC (Previous Solution)*. Since similar results were observed independently of considering the 9×1 or the 9×3 views in the lower layers, the results are shown in Figure 9 (in terms of PSNR) and Figure 10 (in terms of SSIM) for the 9×1 arrangement only.

From these results, it can be seen that the *SLFC (Proposed Solution)* outperforms the benchmark solutions and there is a similar trend in terms of coding performance using the different quality metrics. Regarding the standard deviation values presented in Figures 9 and 10, it can be observed that light field images with smaller MI resolution (i.e., *Demichelis Spark*, *Robot 3D* and *Plane and Toy*) present slightly higher variation in PSNR and SSIM values for the three compared solutions.

In addition, it was observed that there is no significant discrepancy between the quality of the compressed views in the lower layers and the quality of rendered views from the compressed Second Layer.

### 5.5 Further ILR Performance Analysis for Bi-Prediction

This section further analyzes the performance of the proposed improved Inter-Layer prediction (by using exemplar-based ILR picture construction) for the specific case where bi-prediction is allowed when encoding the Second Layer. For this, an alternative SLFC solution is compared with the complete *SLFC (Proposed Solution)*:

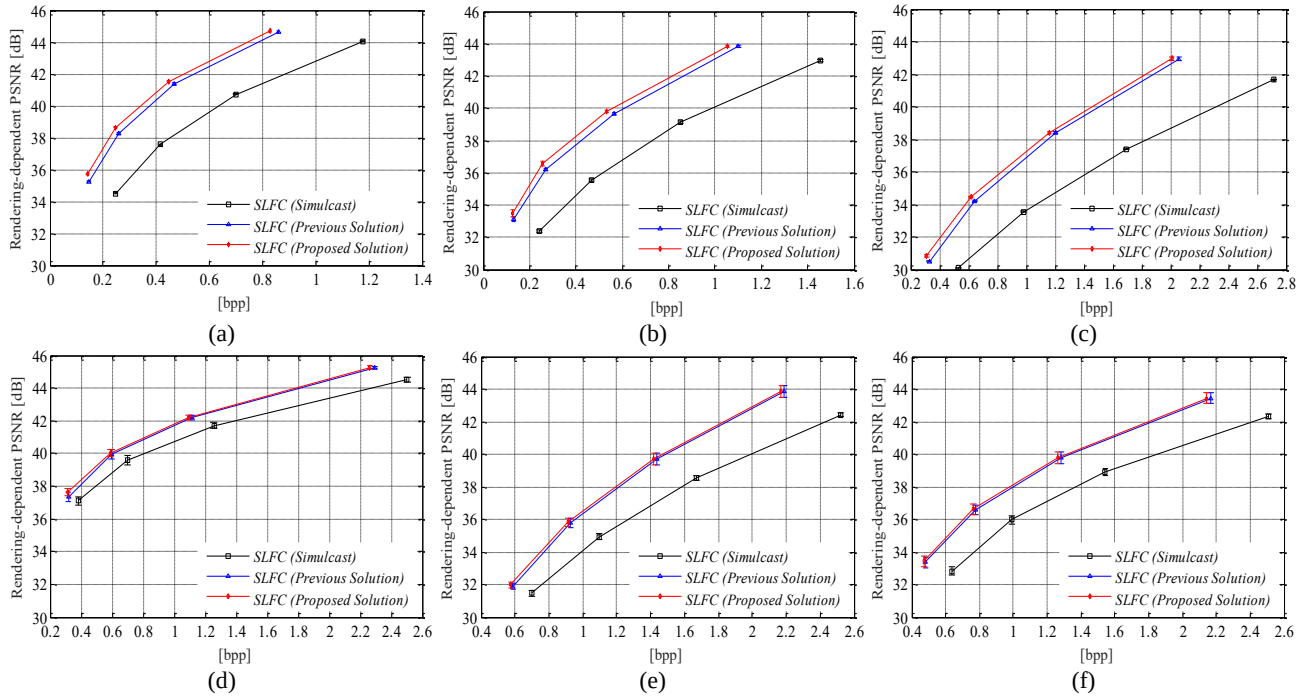



Figure 9 Rate distortion performance (in terms of PSNR against bpp) for a set of 9 views rendered from the compressed light field layer for: (a) Fredo, (b) Seagull, (c) Laura, (d) Demichelis Spark, (e) Robot 3D, and (f) Plane and Toy

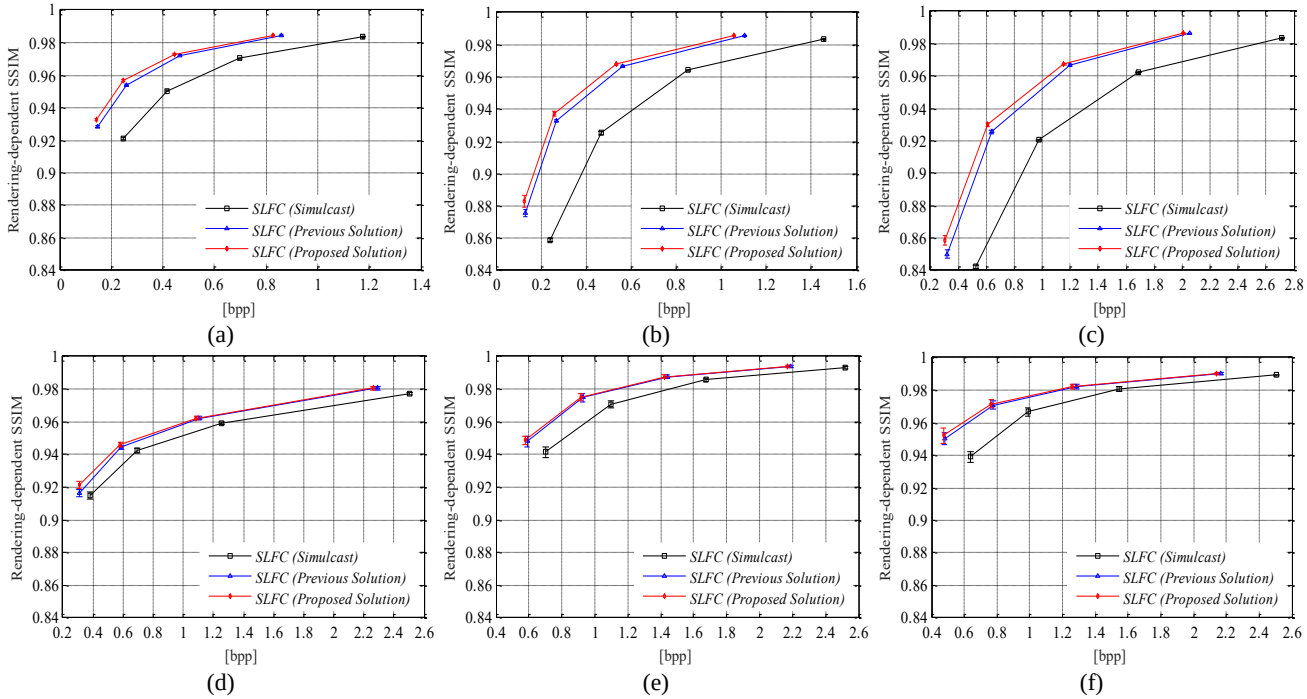



Figure 10 Rate distortion performance (in terms of SSIM against bpp) for a set of 9 views rendered from the compressed light field layer for: (a) Fredo, (b) Seagull, (c) Laura, (d) Demichelis Spark, (e) Robot 3D, and (f) Plane and Toy

- *SLFC (Restricted IL)*: In this case, the Second Layer is encoded with the proposed LF enhancement encoder where bi-prediction is allowed (coded as an Inter B frame). However, Inter-Layer prediction with only the previous MI refilling ILR is enabled.

Table 4 Analysis of the proposed improved inter-layer prediction when considering 9×1 views in the lower hierarchical layers. Prediction usage statistics when encoding each test images with the proposed SLFC solution and comparison of the performance of the improved inter-layer prediction when bi-prediction is used

| **Images** | ***SLFC(Proposed Solution)* vs. *SLFC(Restricted ILR)*** | | ***SLFC (Proposed Solution)* Prediction Usage** | | | |
|---|---|---|---|---|---|---|
| | **BD-PSNR** | **BD-BR** | **SS Reference** | **MI Refilling based ILR** | **Exemplar-based ILR** | **Bi-prediction** |
| *Fredo* | -0.02 dB | 0.33 % | 36.40 % | 13.53 % | 0.32 % | 49.75 % |
| *Seagull* | -0.01 dB | 0.21 % | 29.86 % | 6.92 % | 0.32 % | 62.90 % |
| *Laura* | -0.01 dB | 0.11 % | 35.71 % | 12.13 % | 0.49 % | 51.67 % |
| *Demichelis Spark* | 0.01 dB | -0.31 % | 18.11 % | 17.56 % | 1.62 % | 62.71 % |
| *Robot 3D* | -0.01 dB | 0.13 % | 53.24 % | 25.52 % | 0.84 % | 20.40 % |
| *Plane and Toy* | 0.00 dB | -0.02 % | 41.01 % | 23.37 % | 0.33 % | 35.29 % |
| Average | -0.01 dB | 0.08 % | 35.72 % | 16.50 % | 0.65 % | 47.12 % |

Table 5 Analysis of the proposed improved inter-layer prediction when considering 9×3 views in the lower hierarchical layers. Prediction usage statistics when encoding each test images with the proposed SLFC solution and comparison of the performance of the improved inter-layer prediction when bi-prediction is used

| **Images** | ***SLFC(Proposed Solution)* vs. *SLFC(Restricted ILR)*** | | ***SLFC (Proposed Solution)* Prediction Usage** | | | |
|---|---|---|---|---|---|---|
| | **BD-PSNR** | **BD-BR** | **SS Reference** | **MI Refilling based ILR** | **Exemplar-based ILR** | **Bi-prediction** |
| *Fredo* | 0.21 dB | -3.95 % | 32.12 % | 8.01 % | 10.26 % | 49.61 % |
| *Seagull* | 0.07 dB | -1.46 % | 31.13 % | 9.68 % | 3.59 % | 55.60 % |
| *Laura* | 0.18 dB | -2.95 % | 31.74 % | 9.78 % | 9.44 % | 49.04 % |
| *Demichelis Spark* | 0.01 dB | -0.27 % | 24.05 % | 23.11 % | 8.24 % | 44.60 % |
| *Robot 3D* | 0.08 dB | -0.95 % | 39.05 % | 25.37 % | 13.84 % | 21.74 % |
| *Plane and Toy* | 0.02 dB | -0.36 % | 37.61 % | 24.89 % | 8.09 % | 29.41 % |
| Average | 0.10 dB | -1.66 % | 32.62 % | 16.81 % | 8.91 % | 41.67 % |

Therefore, Tables 4 and 5 illustrate the BD performance of the complete *SLFC (Proposed Solution)* against the *SLFC (Restricted IL)* when considering, respectively 9×1 and 9×3 view arrangements. It can be observed that, when considering a smaller number of views in lower hierarchical layers (i.e., in the 9×1 arrangement), there is no significant difference between the performance of both these solutions. On the other hand, the complete *SLFC (Proposed Solution)* presents a slightly better performance when a larger set of views is considered in the lower layers (i.e., in the 9×3 arrangement).

Moreover, Tables 4 and 5 also illustrate the percentage of usage of each reference frame (i.e. the percentage of uni-prediction using only the SS reference, the MI refilling based ILR, or the exemplar-based ILR) when encoding with the complete *SLFC (Proposed Solution)* compared with the percentage of usage of a combination of these different references (by using bi-prediction). It can be observed that, in the case of the 9×1 arrangement, the largest percentage of the coding blocks (half of them) are encoded by using bi-prediction, and the exemplar-based ILR is hardly ever used (when using uni-prediction). This result is consistent with what has been shown in Section 4.1 (see Figure 6) regarding the accuracy of the exemplar-based ILR picture when more or less LF information is discarded when generating the content for the lower layers. Notably, the less accurate ILR picture is, the less used it will be in a RD manner. Nevertheless, as the amount of information in the lower layer increases (seeTable 5), a better exemplar-based ILR is constructed, which may then be used as an alternative to the bi-prediction to improve the coding performance. This is illustrated in Table 5 by the increased percentage of usage of the exemplar based ILR, and consequent decreased percentage of usage of the bi-prediction.

# 6 FINAL REMARKS

This paper has proposed to improve the performance of the authors' previous display scalable light field coding solution by using two new exemplar-based inter-layer coding tools. Notably, an inter-layer compensated prediction using a reference picture that was constructed relying on an exemplar-based algorithm for texture synthesis, and a direct prediction

based on exemplar texture samples from lower layers. Experimental results confirmed the advantage of the proposed scalable architecture compared to various benchmark solutions, and showed that the performance of the proposed exemplar-based inter-layer prediction improves as the number of views in the Base and First Layers increases.

Finally, in terms of future work, the authors plan to investigate opportunities to enhance the proposed exemplar-based Inter-layer prediction and to enlarge the applicability of the proposed solution by incorporating supplementary data (such as depth, ray-space, and 3D model) into the scalable bitstream.

## 7 ACKNOWLEDGEMENT

The authors acknowledge the support of FCT (Fundação para a Ciência e a Tecnologia, Portugal), under the project UID/EEA/50008/2013, and SFRH/BD/79480/2011 grant.

## REFERENCES

[1] Levoy, M., Hanrahan, P., "Light Field Rendering," SIGGRAPH 96, 31–42, ACM, New York, NY, USA (1996).
[2] Ng, R., "Digital Light Field Photography" (2006).
[3] Aggoun, A., Tsekleves, E., Swash, M. R., Zarpalas, D., Dimou, A., Daras, P., Nunes, P.., Soares, L. D., "Immersive 3D Holoscopic Video System," IEEE Multimed. 20(1), 28–37 (2013).
[4] Georgiev, T., Lumsdaine, A., "Focused Plenoptic Camera and Rendering," J. Electron. Imaging 19(2), 021106–021106 (2010).
[5] Lippmann, G., "Épreuves Réversibles Donnant la Sensation du Relief," J. Phys. Théorique Appliquée 7(1), 821–825 (1908).
[6] Xiao, X., Javidi, B., Martinez-Corral, M., Stern, A., "Advances in Three-Dimensional Integral Imaging: Sensing, Display, and Applications [Invited]," Appl. Opt. 52(4), 546–560 (2013).
[7] Arai, J., "Integral Three-Dimensional Television (FTV Seminar)," ISO/IEC JTC1/SC29/WG11 M34199, Sapporo, Japan (2014).
[8] Raytrix, "Raytrix Website," 2012, <http://www.raytrix.de/> (7 July 2014 ).
[9] Georgiev, T., Yu, Z., Lumsdaine, A., Goma, S., "Lytro Camera Technology: Theory, Algorithms, Performance Analysis," Proc. SPIE 8667, Multimed. Content Mob. Devices 8667, 86671J – 86671J – 10, Burlingame, USA (2013).
[10] Raghavendra, R., Raja, K. B., Busch, C., "Presentation Attack Detection for Face Recognition using Light Field Camera.," IEEE Trans. Image Process. 24(3), 1060–1075 (2015).
[11] Schelkens, P., "JPEG PLENO – Scope, Use Cases and Requirements Ver.1.3," ISO/IEC JTC 1/SC 29/WG1 M71003, La Jolla, CA, USA (2016).
[12] Tehrani, M. P., Shimizu, S., Lafruit, G., Senoh, T., Fujii, T., Vetro, A., Tanimoto, M., "Use Cases and Requirements on Free-viewpoint Television (FTV)," ISO/IEC JTC1/SC29/WG11 MPEG N14104, Geneva, Switzerland (2013).
[13] Wang, J., Xiao, X., Hua, H., Javidi, B., "Augmented Reality 3D Displays with Micro Integral Imaging," J. Disp. Technol. 11(11), 889–893, IEEE (2015).
[14] Lanman, D., Luebke, D., "Near-eye light field displays," ACM SIGGRAPH 2013 Emerg. Technol. - SIGGRAPH '13, 1–1, ACM Press, New York, New York, USA (2013).
[15] Conti, C., Nunes, P., Soares, L. D., "Inter-Layer Prediction Scheme for Scalable 3-D Holoscopic Video Coding," IEEE Signal Process. Lett. 20(8), 819–822 (2013).
[16] Criminisi, A., Perez, P., Toyama, K., "Region Filling and Object Removal by Exemplar-Based Image Inpainting," IEEE Trans. Image Process. 13(9), 1200–1212 (2004).
[17] Conti, C., Lino, J., Nunes, P., Soares, L. D.., Correia, P. L., "Spatial Prediction Based on Self-Similarity Compensation for 3D Holoscopic Image and Video Coding," 2011 18th IEEE Int. Conf. Image Process., 961–964, IEEE, Brussels (2011).
[18] Conti, C., Nunes, P., Soares, L. D., "New HEVC Prediction Modes for 3D Holoscopic Video Coding," 2012 19th IEEE Int. Conf. Image Process., 1325–1328, IEEE, Orlando, USA (2012).
[19] Conti, C., Soares, L. D., Nunes, P., "HEVC-Based 3D Holoscopic Video Coding using Self-Similarity Compensated Prediction," Signal Process. Image Commun. 42, 59–78 (2016).
[20] Conti, C., Nunes, P., Soares, L. D., "HEVC-Based Light Field Image Coding with Bi-Predicted Self-Similarity Compensation," IEEE Int. Conf. Multimed. Expo - ICME, Seattle, USA (2016).

[21] Li, Y., Sjöström, M., Olsson, R., Jennehag, U., "Coding of Focused Plenoptic Contents by Displacement Intra Prediction," IEEE Trans. Circuits Syst. Video Technol. 26(7), 1308–1319 (2016).
[22] Lucas, L. F. R., Conti, C., Nunes, P., Soares, L. D., Rodrigues, N. M. M., Pagliari, C. L., da Silva, E. A. B., de Faria, S. M. M., "Locally Linear Embedding-Based Prediction for 3D Holoscopic Image Coding using HEVC," 2014 Proc. 22nd Eur. Process. Conf., 11–15 (2014).
[23] Dick, J., Almeida, H., Soares, L. D., Nunes, P., "3D Holoscopic Video Coding using MVC," 2011 IEEE EUROCON - Int. Conf. Comput. as a Tool, 1–4, IEEE, Lisbon (2011).
[24] Shi, S., Gioia, P., Madec, G., "Efficient Compression Method for Integral Images using Multi-View Video Coding," 2011 18th IEEE Int. Conf. Image Process., 137–140, IEEE, Brussels (2011).
[25] Olsson, R., Sjostrom, M., Xu, Y., "A Combined Pre-Processing and H.264-Compression Scheme for 3D Integral Images," 513–516 (2006).
[26] Vieira, A., Duarte, H., Perra, C., Tavora, L., Assuncao, P., "Data Formats for High Efficiency Coding of Lytro-Illum Light Fields," 2015 Int. Conf. Image Process. Theory, Tools Appl., 494–497, IEEE (2015).
[27] Choudhury, C., Chaudhuri, S., "Disparity Based Compression Technique for focused Plenoptic Images," Proc. 2014 Indian Conf. Comput. Vis. Graph. Image Process. - ICVGIP '14, 1–6, ACM Press, New York, New York, USA (2014).
[28] Graziosi, D. B., Alpaslan, Z. Y., El-Ghoroury, H. S., "Depth Assisted Compression of Full Parallax Light Fields," Proc. SPIE 9391, Stereosc. Displays Appl. XXVI, 93910Y – 93910Y – 15, San Francisco, CA (2015).
[29] Li, Y., Sjöström, M., Olsson, R., Jennehag, U., "Scalable Coding of Plenoptic Images by Using a Sparse Set and Disparities.," IEEE Trans. Image Process. 25(1), 80–91 (2016).
[30] Bishop, T. E., Favaro, P., "Plenoptic Depth Estimation from Multiple Aliased Views," 2009 IEEE 12th Int. Conf. Comput. Vis. Work. ICCV Work., 1622–1629, IEEE (2009).
[31] Conti, C., Nunes, P., Ducla Soares, L., "Using self-similarity compensation for improving inter-layer prediction in scalable 3D holoscopic video coding," Proc. SPIE 8856 Appl. Digit. Image Process. XXXVI 8856, 88561K – 1–88561K – 13, San Diego, USA (2013).
[32] Vetro, A., Wiegand, T., Sullivan, G. J., "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," Proc. IEEE 99(4), 626–642 (2011).
[33] Tech, G., Chen, Y., Muller, K., Ohm, J.-R., Vetro, A., Wang, Y.-K., "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding," IEEE Trans. Circuits Syst. Video Technol. 26(1), 35–49 (2016).
[34] Ng, R., "Fourier Slice Photography," 2005, 735–744, ACM.
[35] Yu, Z., Yu, J., Lumsdaine, A., Georgiev, T., "An analysis of color demosaicing in plenoptic cameras," June 2012, 901–908.
[36] Dansereau, D. G. D. G., Pizarro, O., Williams, S. B. S. B., "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras," 2013 IEEE Conf. Comput. Vis. Pattern Recognit., 1027–1034, IEEE (2013).
[37] Sullivan, G. J., Ohm, J.-R., Han, W.-J., Wiegand, T., "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. Circuits Syst. Video Technol. 22(12), 1649–1668 (2012).
[38] Tan, T., Boon, C., Suzuki, Y., "Intra Prediction by Template Matching," 2006 Int. Conf. Image Process., 1693–1696, IEEE (2006).
[39] Geogiev, T., "Todor Georgiev Gallery of Light Field Data," July 2014, <http://www.tgeorgiev.net/Gallery/> (7 July 2014 ).
[40] "3D Holoscopic Sequences (download link).", 2013, <http://3dholoscopicsequences.4shared.com/> (30 November 2015 ).
[41] "MV-HEVC Reference Software HTM-12.0.", <https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-12.0/> (22 December 2014 ).
[42] Bossen, F., "Common HM Test Conditions and Software Reference Configurations," JCTVC-L1100, Geneva (2013).
[43] Bjøntegaard, G., "Calculation of Average PSNR Differences between RD Curves," VCEG-M33, Austin, TX, USA (2001).