Light Field Imaging Coding

João Garrote
Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
joao.garrote@ist.utl.pt

***Abstract*—Light fields have emerged as one of the most promising 3D representation formats, enabling a richer and more faithful representation of a visual scene. One of the most popular light field acquisition methods consists in the placement of an array of micro-lenses between the camera main lens and the photosensor allowing to capture both the radiance and the direction of the light rays. This type of representation format offers new possibilities to interact with the visual content, namely, refocusing and visualization of different perspectives of the visual scene, after acquisition is performed; however, the acquisition process results in large amounts of data which require a large storage space and transmission bandwidth. Therefore, light field data compression has a critical role in the transmission and storage of this type of format. The objective of this work is to develop an efficient lenslet light field image coding solution, exploiting the characteristics of this type of visual representation. Therefore, a wavelet-based encoder was designed and implemented to offer view, quality and spatial scalability that can meet the demand of different types of display and data transmission rates. The proposed coding solution exploits the Inter view correlation with disparity estimation and compensation tools and reuses a JPEG 2000 image codec. The results obtained show that an improvement regarding the Intra codec (only JPEG 2000) when the proposed Inter scheme is used, achieving constant performance gains for the set of light field images selected for evaluation**.

***Index Terms*—light field; microlens array; disparity estimation and compensation; image compression; wavelet transform; JPEG 2000.**

I. Introduction

In the last years, there were major developments in sensors/cameras, such as Lytro and Raytrix cameras, and displays that allow for a better acquisition and replication of the visual world. With this richer representation, the visual data can be manipulated a posteriori by the users, allowing to control the focus, scene perspective or even stereoscopy image creation.

The acquired lenslet light field image carries a huge amount of data, thus asking for efficient coding solutions. While the light field coding domain is rather recent, there are already several solutions proposed in the literature; which can be grouped into four categories:

i) Standard compliant coding solutions; naturally, these coding solutions cannot exploit all available redundancy but they allow benefiting from the standard ecosystem as standard bitstreams and decoders are used. These coding solutions can be divided in still image coding standards, namely JPEG and JPEG 2000, and video coding standards used in the Intra coding mode, namely HEVC Intra and H.264/AVC Intra [1].

ii) Standard compliant coding solutions applied after some data re-organization; the main target of this re-organization is to better exploit the redundancy in the data. Some methods consist in a simple re-organization of the data, e.g. [2], which takes the set of SA images as a sequence of video frames; others, e.g. [3], code the set of SA images using some appropriate 2D spatial prediction structure.

iii) Extended standard coding solutions; where additional tools are included to improve the standard compression performance for lenslet light field images; e.g. [4] [5],which extend the HEVC standard with additional prediction tools.

iv) Non-standard based coding solutions; some solutions perform the light field coding aided by depth information [6], while others, have as cornerstone different transforms, notably the discrete wavelet transform (DWT) [7] [8] [9], discrete cosine transform (DCT) [8] and Karhunen-Loeve transform (KLT) [10]. These coding solutions use these transforms alone or even combined [11] [12].

The proposed coding solution estimates and compensates the disparity between the SA images of the 4D light field using a lifting-based wavelet decomposition scheme. This coding solution is inspired in [13], namely on the way the disparity compensation is integrated into the lifting-based wavelet transform. The proposed coding solution is able to provide view scalability and quality, or resolution, scalability.

The remainder of this paper is organized as follows: Section II presents the proposed solution architecture with a brief description of the different modules. Section III presents a detailed description of the inter-view disparity compensated wavelet transform. Section IV presents the proposed solution performance using an appropriate assessment methodology, test conditions and metrics. Section VI presents the final conclusions and suggest future work directions.

II. Disparity Compensated Light Field Coding Architecture

The proposed coding architecture, shown in Figure 1, aims to reuse as much as possible available coding tools, now organized to build an efficient and scalable lenslet light field coding solution. The Light Field Toolbox [14] is used to pre-process the lenslet light field image data, i.e. to obtain the SA images in the RGB color space, which are then converted to the YCrCb color space. Some architectural modules are based on the JPEG 2000 codec (highlighted in red in Figure 1), such as the 2D-DWT Intra-View Transform, the Uniform Scalar Quantization and finally, the EBCOT Coding. The Disparity Compensated Inter-View DWT corresponds to the main novelty of this solution, where the redundant content between SA images is exploited while targeting offering view scalability.

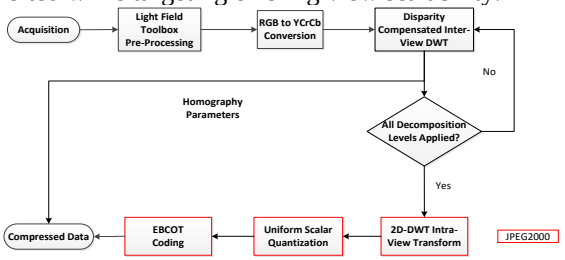


***Figure 1:** Architecture of the disparity compensated light field encoder.*

A brief description of each module is presented next:

i) **Light Field Toolbox Pre-Processing:** The objective of this module is to convert the light field obtained directly from the sensor into a more suitable representation format. First, the so-called lenslet image is created from the raw sensor data by applying demosaicing, devignetting, clipping, and some color processing. Then, the lenslet image, formed by thousands of micro-images, is converted into an array of SA images, each representing a different perspective view. This module uses the available Light Field Toolbox v0.4 software [14]. While the original light field is composed by 225 SA images, it was decided to discard both the first and last row and column of SA images, resulting into 169 SA images, to avoid using SA images without enough quality, notably some black images that are obtained in the corners due to the vignetting effect. This strategy has been also adopted by JPEG in the JPEG PLENO Call for Proposals [15].

ii) **RGB to YCrCb Conversion:** The objective of this module is to improve the compression efficiency by converting the RGB data into YCrCb data which is a more compression friendly format as different sampling density may be used for the chrominances and luminance. Thus, at this stage, the SA images are converted from the RGB to the YCrCb color space.

iii) **Disparity Compensated Inter-View DWT:** An inter-view wavelet transform is chosen to decorrelate the various SA images and compact their energy into a small number of bands. This transform was designed with a lifting structure to allow including disparity estimation and compensation techniques in the prediction and updating steps [13]. The overall objective of the designed transform is to obtain low-frequency and high-frequency bands in such a way that the low-frequency band corresponds to a smoothed representation of a view and the high-frequency band corresponds to high frequency information necessary to obtain the other view. The wavelet transform with disparity compensated lifting is applied to an array of SA images with size *N* and its frequency decomposition capabilities lead to *N*/2 low-frequency bands and *N*/2 high-frequency bands. To further exploit the correlation between the low-frequency bands, the wavelet transform can be used again in a second decomposition level, using now as input the low-frequency. With the application of 1-level decomposition transform, two scalability layers are available, the first associated to the low-frequency bands and the second associated to the high-frequency bands which correspond to details; for each decomposition level added, one more scalability layer becomes available. A simplified architecture of the forward transform is shown in Figure 2 which just includes an additional step (disparity estimation) and the predict and update steps that are designed to exploit the disparity between the SA images.

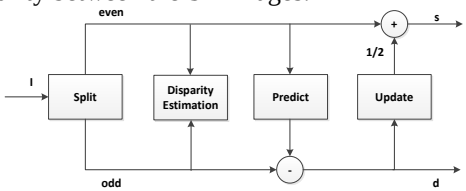


Figure 2: *Architecture for the 1-level Haar wavelet transform.*

The forward transform is applied to a row or column of SA images (in the first decomposition level) in the following way: i) the input, *I*, is a sequence of *N* SA images that are split into: a) even SA images and b) odd SA images (**Split**); ii) Compute a global view perspective geometric transformation matrix that describes the disparity from the even to the odd SA images and referred as w_{01} (**Disparity Estimation**); iii) From the even SA image, the odd SA image is predicted, thus implementing a warping operation that is referred as $w_{01}(even)$. At this stage, the high-frequency band can be computed as $d = odd - w_{01}(even)$ (**Predict**); iv) A warped version of the high-frequency band, d, referred as $w_{10}(d)$ is now used to update the even SA image. Thus, after the update step, the low-frequency band is computed as $s = even + \frac{1}{2} w_{10}(d)$ (**Update**).

iv) **Intra-View 2D-DWT**: The objective here is to exploit the spatial redundancy within each SA image or high-frequency/low-frequency band. The 2D-DWT transform with six decomposition levels, as proposed in the OPENJPEG software [16], has been adopted for application to all the frequency bands resulting from the inter-view transform. This process consists basically in applying a 1D-DWT along the X-axis (spatially horizontally) and, after, again along the Y-axis (spatially vertically) to each image/band. The result of a 1-level 2D wavelet decomposition is four filtered and subsampled images, also known as bands. The application of the DWT JPEG 2000 transform will lead to the decomposition shown in Figure 3 up). The 2D-DWT enables resolution scalability as the SA images can be decoded at full resolution or only at a fraction resolution of it [13]. For example, just decoding the LL3 band allows obtaining an image with a rather small resolution.

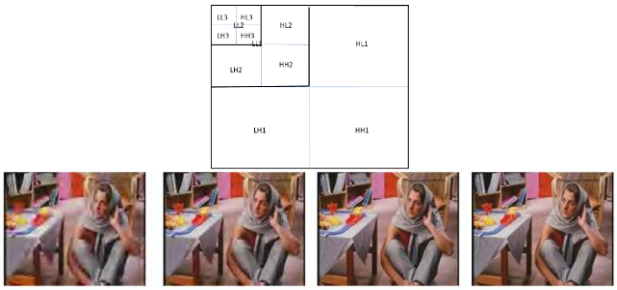


Figure 3: *up) 2D-DWT decomposition details* [17] *; down) Example of quality scalability, the more bits received the better is the quality* [17].

v) **Quantization**: The objective is here to reduce the accuracy of the DWT coefficients to obtain higher compression. The quantization is performed using uniform scalar quantization with a dead-zone, which is one of the available JPEG 2000 quantization methods [18]. In JPEG 2000 Part 1, the deadzone size is twice larger than the quantizer step size while in JPEG 2000 Part 2, the deadzone size can be adjusted for each band. This quantization method allows also to progressively transmit the coefficients (quality or SNR scalability) by progressively sending the most significant bitplanes (MSB) and then advancing to the least significant bit (LSB) bitplanes; Figure 3 down) shows an example demonstrating the quality improvements when receiving more bits. All the bands obtained after applying the Intra-view 2D-DWT are quantized using this method [18].

vi) **EBCOT**: The objective is to exploit the statistical redundancy (entropy coding) of the band coefficients. First, each band is divided into small rectangular blocks, referred to as codeblocks, and each codeblock is independently encoded with EBCOT. All codeblocks from low-frequency to high-frequency are scanned together from top to bottom and left to right. Note that each band is independently coded from the other bands. EBCOT performs multiple-pass coding of the

codeblock bitplanes obtained in the previous step. Three passes are used, notably significance propagation, magnitude refinement and cleanup; more details about each pass are available in [18]. For JPEG 2000 to be compression efficient, a context-based adaptive binary arithmetic coding method is used which exploits the correlation among bitplanes.

III. Inter-View Disparity Compensated Wavelet Transform

The Haar DWT with 1-level of decomposition show in Figure 2, which performs disparity-compensation using a perspective geometric transform, is displayed in Figure 4 with more detail, highlighting the main steps.

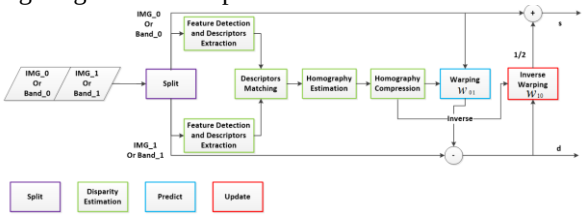


Figure 4: *Inter-view DWT applied to two SA images (IMG_0, IMG_1) or low-frequency/high-frequency bands (Band_0, Band_1); highlighting the relationship with the modules from Figure 2.*

In more detail, the Inter-View Disparity-Compensated DWT consists in the following steps, see Figure 4:

1. **Split**: Divides the input set of SA images (or bands) into two different, complementary sets where the even SA images are grouped into one set and the odd SA images (or bands) into another set.
2. **Feature Detection and Descriptor Extraction**: The objective of this module is to detect distinctive features in the images associated to keypoints or blobs, and extract after descriptors for those positions which represent the features in some space that is invariant to common deformations such as translation, scaling, rotation, perspective changes and partially invariant to illumination changes. In this case, it is proposed to use the popular SIFT (Scale Invariant Feature Transform) descriptor [19]. The main steps of this module are described below [19]:

   i) SIFT Feature Detection – The feature detection can be divided in the following steps [20]: i) Scale-space extrema detection: first, the Gaussian scale-space representation of the image is obtained, which corresponds to a pyramid of images at different resolutions obtained by progressively smoothing and downsampling the input image. Then, adjacent images in the pyramid are subtracted to obtain the Difference of Gaussians (DoG) from which maxima and minima can be found. To identify potential interesting points invariant to scale and orientation, the DoG pyramid is searched over all scales and image locations; ii) Keypoint localization: In the previous step, potential keypoints are found, which need to be refined to get more accurate results. Based on measures of stability, undesirable feature points are removed, e.g. keypoints with low contrast and unstable [21]; iii) Orientation assignment: The last step is to compute the best orientation for each detected keypoint, to achieve invariance to image rotation, based on a histogram of local gradient directions at the specified scale.

   ii) SIFT Descriptor Extraction – At this stage, each keypoint is characterized with location, scale and orientation and, therefore, it is possible to compute a descriptor that characterizes the image region (patches) surrounding it. The descriptor extraction step can be divided into [20]: i) Normalization where the patch is appropriately rotated and scaled; ii) Keypoint description where the local image gradients are measured at the selected patch scale with the objective of computing the orientation histogram, as shown in Figure 5. Both the gradient magnitude and orientation at each position of the patch are considered and the gradient magnitudes are weighted by a Gaussian window.

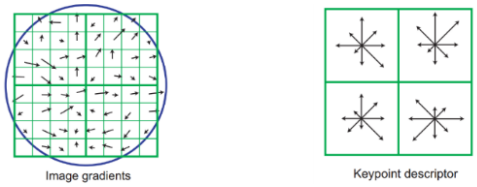


Figure 5: *left) Gradient magnitude and orientation at each image sample in a region around the keypoint location; these are weighted by a Gaussian window, represented by the overlaid circle* [20]*; right) Orientation histograms summarizing the contents over 4x4 sub-regions* [20] .

3. **Descriptors Matching**: The objective here is to match individually a set of descriptors from one SA image with the descriptors from another SA image, obtaining a set of one-to-one correspondences. A simple approach has been followed consisting in taking each descriptor in the first set and matching it with all the descriptors in the second set, using some distance metric, e.g. the Euclidean distance. Then, a ratio test was applied. This test compares the ratio of distances between the two top matches for a given keypoint. If the ratio is above the threshold of 0.7, the match is rejected. The objective of this test is to increase the reliability of the matching procedure, thus avoiding some wrong matches between keypoints [20].
4. **Homography Estimation**: In this context, a homography is a geometric transformation which establishes a relationship between corresponding positions of two different, but somehow related, images; these correspondences were obtained in the previous step. Thus, at this stage, the objective of this module is to estimate the transformation between one SA image (or a low-frequency band) and the other. Several formulations for this transformation are possible, such as [22]: i) Affine Transform; ii) Perspective Transform; iii) Bilinear Transform; iv) Polynomial Transform. Considering the lenslet light field data characteristics, the most adequate transform for modelling the data seems to be the perspective transform since it is able to model complex geometry relationships between different perspectives of the objects in the visual scene. The perspective transform is defined by a 9-parameter matrix which is able to describe the displacements that the objects in the visual scene suffer when the perspective changes. When applied to the case of SA images, this transform should describe well the disparity between the SA images (mainly determined by the characteristics of the microlens array). To estimate the perspective transform parameters, RANSAC, an iterative method to estimate parameters that is robust even when there are some wrong matches (outliers), is used. To avoid that outliers reduce the accuracy of the estimated transformation matrix, the method attempts to identify inliers, which is data that can be explained by a set of model parameters (typically estimated with a standard regression method) and, therefore, not considering outliers (erroneous correspondences) in the estimation.
5. **Homography Parameters Compression**: The parameters of the perspective transform (matrix values) are expressed

with 8 bytes (64 bits), assuming double floating point precision. Since this precision may require a significant amount of rate, as these parameters have to be transmitted to the decoder, it is important to propose a quantization technique to compress this type of data. Note also that, each time the inter-view wavelet transform is applied, a different perspective transform matrix is used and, therefore, since there are many pairs of SA images (bands) for which a transform is applied, the number of matrices may be rather high. The technique proposed is implemented as described here:

1. **Average matrix computation**: the first step is to compute an average matrix, H_{avg}; this is done using as input all perspective matrices for a given decomposition level, for each matrix parameter all values are summed and divided by the number of matrices.
2. **Min and Max matrix computation**: after the minimum and maximum value of each matrix parameter is computed again using all matrices, thus obtaining H_{min} and H_{max}. These last two values represent the dynamic range of the perspective parameters and, therefore, may be used to adjust the quantization step efficiently.
3. **Residual matrix computation**: subtract the original matrix H_{org} to be coded by the average matrix H_{avg}, as in Equation (1); the purpose is to obtain residual parameter values close to zero.

$$H_{res} = H_{org} - H_{avg} \quad (1)$$

4. **Normalization**: the H_{res} residual parameters are now normalized to the interval [0;1] using Equation (2) and considering the auxiliary matrices computed in step 2; the purpose is to have a fixed dynamic range and, therefore, a more compression friendly representation of the perspective matrix, the normalized matrix H_{norm}.

$$H_{norm} = \frac{H_{res} - H_{min}}{H_{max} - H_{min}} \quad (2)$$

5. **Quantization**: to reduce the parameters precision, a quantization process is applied thus obtaining the quantized matrix H_{quant} to be transmitted to the decoder. To use *N* bits (24 bits) for each parameter, it is necessary to apply Equation (3) and Equation (4). The value of *N* was obtained by testing different precisions values and selecting the one which resulted in a better RD performance.

$$H_{quant} = \text{round}(H_{norm} * 2^N) \quad (3)$$

$$round(x) = \begin{cases} 1 \; if \; x \geq 0{,}5 \\ 0 \; if \; x < 0{,}5 \end{cases} \quad (4)$$

The mentioned steps must be applied to each matrix obtained for a given decomposition level; the auxiliary matrices H_{avg}, H_{min} and H_{max} are also subjected to the same process of quantization, i.e. the step 5 mentioned above. The H_{quant} matrix is transmitted to the decoder without any entropy coding, just using a fixed number of bits per parameter. Using this approach, where 24 bits are used to represent each homography parameter, it is possible to reduce the number of bits from the original 64 to 24. The decoding process is straightforward after the auxiliary matrixes H_{avg}, H_{min} and H_{max} are inverse quantized. The remaining matrixes are obtained by first, inverting the quantization, using Equation (5), then de-normalizing the parameters using Equation (6), and finally, the decoded matrix is computed using Equation (7).

$$H_{quant} = \frac{H_{norm}}{2^N} \quad (5)$$

$$H_{de-norm} = H_{quant} \times (H_{max} - H_{min}) + H_{min} \quad (6)$$

$$H_{dec} = H_{de-norm} + H_{avg} \quad (7)$$

6. **Warping or Disparity Compensation (w_{01})**: The objective is to transform/warp an input even SA image in such a way that it becomes similar to the odd SA image, which in this case corresponds to a slightly different perspective, in practice performing disparity compensation. This warping process is performed by using the decoded transformation matrix, H_{dec}. Considering $[x,y]$ the coordinates of a sample in the warped image and $[u,v]$ the coordinates of the corresponding sample in the input image, the SA image prediction is computed by multiplying each sample position in the input image by the transformation matrix Computing the difference between an odd view and the warped even view results in the high-frequency band.
7. **Inverse Warping (w_{10}):** As the scene disparity involved in this kind of data mostly corresponds to translations, because it is mostly due to the spatial separation between the microlenses and only a little due to optical defects in the microlens, the transformation matrix from a reference view into another view can be inverted, and thus an inverse transformation matrix can be obtained. This is also a requirement from the disparity compensated wavelet transform which can only be applied when the warpings w_{01} and w_{10} are symmetric as otherwise the process may end up adding a residual to the odd view that is not aligned (prediction step), thus creating ghost artifacts. By computing a weighted sum between the even view and the warped high-frequency band, a low-frequency band (smoothed) SA image is obtained.

The lifting scheme for the inverse disparity compensated wavelet transform to be performed at the decoder follows the scheme presented in Figure 6. As the homography matrix parameters computed in the homography estimation step are transmitted to the decoder, the inverse transform only needs to perform the predict and update steps in the reverse order flipping the signal in arithmetic operations, thus resulting in the original signal.

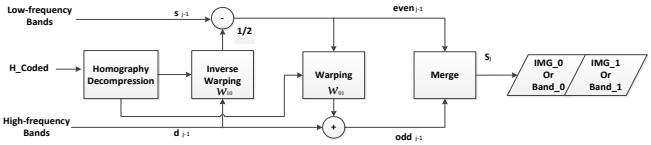


Figure 6: *Inverse inter-view DWT scheme.*

IV. Performance Assessment

The objective here is to assess the rate-distortion (RD) performance of the proposed coding solution, denominated as Disparity Compensated Light Field Coding (DCLFC). The performance will be compared against some relevant alternative coding solutions, notably still image coding standards, such as JPEG and JPEG 2000, and the three coding solutions which were provided to JPEG as responses to the Call for Proposals on Light Field Coding in July 2017 [23] [24] [25].

A. Test Material, Benchmarks and Coding Conditions

i) Test Material

To evaluate the RD performance of the DCLFC solution, five lenslet light field images have been selected from the MMSPG-EPFL Light Field Dataset [26]; this dataset has also been selected as the test set for the Light Field Compression Grand Challenge organized at ICME 2016 [27] and for the JPEG Pleno

Call for Proposals on Light Field Coding [28]. The set of selected images is: i) *Bikes*; ii) *Danger_de_Mort*; iii) *Stone_Pillars_Outside*; iv) *Fountain_&_Vincent_2*; v) *Friends_1*. The images were chosen by their content, aiming to have a diversified dataset, with both high and low frequency content and objects at different depths. To simplify the text, from now on, the names will just be *Bikes*, *Danger*, *Stone*, *Fountain* and *Friends*. Figure 7 shows thumbnails of the *Friends* and *Danger* light fields.

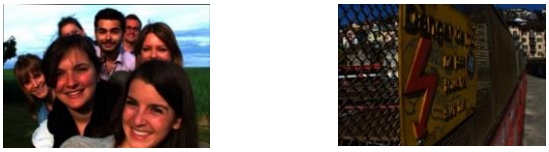


Figure 7: *Thumbnails for light field: left) Friends; right)Danger.*

All the test images were acquired using a Lytro Illum camera and, thus, are available in the LFR (light field raw) format. The software used to extract the light field data from the LFR file for further processing was the Light Field Toolbox, made available by D. Dansereau [29]; with the light field data obtained as RGB images with an 8-bit unsigned integer representation. The light field output structure is a matrix of 225 SA images; however, for compression purposes, only 169 SA images, each with a spatial resolution of 625×434 pixels. Because this is usually enough, the performance assessment will be made only for the luminance (Y) component of the SA images.

ii) Benchmarks

All benchmarks consider the same set of 169 SA images, each with a resolution of 625×434 pixels. When coding with JPEG and JPEG 2000, the input SA images will be organized in two ways: i) SA images individually coded, each SA image is individually encoded and, thus, the RD points are defined by the total rate spent with the 169 SA images and the PSNR computed as the average of the PSNR for all these SA images; ii) SA images coded as a single "super image", where instead of encoding the 169 SA images one-by-one, the SA images are first arranged in a single "super image" which is then encoded all at once. The RD points are defined by the rate spent in the "super image" and the PSNR is computed as the average PSNR of all SA images extracted from the decoded "super image" (as for the previous case, the PSNR compares the decoded SA images with the original ones).

For the three solutions proposed to JPEG as responses to the Call for Proposals, follows bellow a brief description:

- **JPEG Pleno Proposal by Univ. of Science and Technology of China** (labelled Pleno1) [23] – The Pleno1 proposed codec is a lenslet light field image coding solution based on the division of the SA images into two sets, one coded with the HEVC standard, called set S_A, and the other using linear view synthesis from the first set, called set S_B. In summary, the coded bitstream consists on a HEVC stream for the set S_A, and the transformation coefficients for the set S_B. This solution may deliver three layers of scalability: i) the first layer corresponds to the coded set S_A, which consists on a light field image sparsely sampled in the angular dimension; ii) the second layer correspond to adding the synthetized set S_B which offers angular resolution scalability also called view scalability; iii) the third layer (not used in the results) corresponds to adding the set S_B residues which offers quality scalability for these images.
- **JPEG Pleno Proposal by Univ. Tampere, Finland** (labelled Pleno2) [24] – The Pleno2 proposed codec performs the lossy compression of lenslet images by combining sparse predictive coding and JPEG 2000 to the stack of SA images. The first step is to code the entire lenslet image with JPEG 2000 at a fraction of the available bitrate. To improve the reconstruction, the geometry of the scene is computed using a depth estimation algorithm; after, the displacements of the regions in the various SA images regarding the central view are estimated. The pixels of one arbitrary region in one view are predicted using a sparse predictor, having as regressors the pixels from the corresponding pixels in nine neighbor views. The bitstream consists on the full lenslet image coded with JPEG 2000, the depth map, the displacements for each region in each dependent view and, finally, the sparse predictors for each region and each view. Random access can be configured by selecting a set of views to be reference views, i.e. to be coded in a backward compatible way, e.g. using JPEG 2000.
- **JPEG Pleno Proposal by Ostendo, USA** (labelled Pleno3) [25] – The Pleno3 proposed codec makes use of both camera and scene metadata, along with the light field images, to compute the minimum number of reference images required to faithfully reconstruct the full light field. Then, for each reference image, disparity maps are obtained from the computed depth. If a higher quality reconstruction is needed, more reference images and disparity maps are selected for coding. Both the reference images and the reference disparity maps are compressed with HEVC to allow view synthesis at decoder. This solution provides scalability in terms of quality, computational complexity and views.

iii) Coding Conditions

Two quantization parameters may be used for the compression with JPEG 2000 using the adopted software. While the parameter "r", designated in the following as compression-driven quantization parameter (*CDQP*), controls the size of the output compressed image, the second parameter "*q*", designated in the following as quality-driven quantization parameter (*QDQP*), controls the quality of the output image. The quantization parameters values used for each coding solution are:

i) **JPEG** – The software used for JPEG encoding was the OpenCV software, version 2.4.13 [30]; each RD point was obtained using a so-called quality parameter which is in the interval [0; 100] (the higher, the better the quality) with the following values: 1, 15, 30, 50, 70 and 90.

ii) **JPEG 2000** – The software used for JPEG 2000 encoding was the OPENJPEG software version 2.1.2 [31]; each RD point was obtained by setting *QDQP* with the following values: 32, 36, 40 and 45.

iii) **DCLFC** – In the context of this word, the software for the inter-view disparity compensated wavelet transform based on a lifting scheme was developed; moreover, the OPENJPEG software, version 2.1.2 was integrated in the coding solution to implement the JPEG 2000 codec, and also some image processing tools were integrated using the OpenCV software, version 2.4.13; each RD point was obtained by controlling the JPEG 2000 component of the designed DCLFC coding solution by setting the *CDQP*, with the following values: 65, 24, 16, 8, 6 and 4.

iv) **JPEG Pleno proposals** – The JPEG Pleno proposals used the rate points defined in the JPEG Pleno Call for Proposals, notably 0.75, 0.1, 0.02 and 0.005 bpp [14].

During the performance assessment, it was observed that due to differences in terms of content between the low-frequency and high-frequency bands, the *CDQP*, which controls the size of the compressed stream, needed to be controlled by using different values for different bands. In this context, the procedures to determine the appropriate compression-driven quantization parameters are briefly described:

i) **JPEG 2000 coded SA images** – Some SA images are directly encoded with JPEG 2000, which means that the inter-transform scheme is not applied to them. For these images, the compression-driven quantization parameter (*CDQP_SA*) is given by Equation (8), and the *CDQP* takes the values 65, 24, 16, 8, 6 and 4 has already stated.

$$CDQP\_SA = \frac{CDQP}{2} \tag{8}$$

ii) Bands – While the low-frequency bands are an averaged representation of the two input SA images, the high-frequency bands consist mostly on a rather low energy (black) image as their content mainly represents the differences between neighboring SA images or bands. As the *CDQP* controls the output compressed size, it is appropriate to differentiate this parameter for the low-frequency and high-frequency bands as increasing this parameter for the high-frequency bands allows spending less rate without degrading significantly the final reconstruction quality, thus resulting in a better RD performance. In summary, the *CDQP* value for each band is selected in the following way:

a. **Low-frequency bands:** The rate spent (in Bytes) is computed using Equation (9) and the compression-driven quantization parameter for the low-frequency band, *CDQP_LB*, is given by Equation (10). The *Input_Image_Size* corresponds to the size of the input SA images or low-frequency/high-frequency bands, i.e. $625 \times 434 \times 1$ ($Bytes$) for the used test material and for a SA image, *CDPQ* corresponds to one of the initially defined *CDQP* values for the DCLFC solution and, lastly, *CDQP_LB/HB_Ratio* corresponds to the ratio between the *CDQP* for the low-frequency bands and the high-frequency bands; the following values have been tested for *CDQP_LB/HB_Ratio*: 1, 2, 4, 6, 8, 10,12, 14, 16, 18, 20.

$$Bitrate\_LB = \frac{\left(\frac{2 \times Input\_Image\_Size}{CDQP}\right)}{\left(1 + \frac{1}{CDQP\_LB/HB\_Ratio}\right)} \tag{9}$$

$$CDQP\_LB = \frac{Input\_Image\_Size}{Bitrate\_LB} \tag{10}$$

b. **High-frequency bands:** The compression-driven quantization parameter for the high-frequency bands, *CDQP_HB* is given by Equation (11). For each value of *CDQP*, all the corresponding *CDQP_LB/HB_Ratio* values were tested with the objective of finding the best (*CDQP*, *CDQP_LB/HB_Ratio*) pair in terms of RD performance. These tests have shown that the *CDQP_LB/HB_Ratio* parameter should not be constant along the rate, as maintaining the same *CDQP_LB/HB_Ratio* value for rates higher than 0.4 bpp would reduce the reconstruction quality, thus degrading the overall RD performance. In summary, the first three RD points are obtained using a given *CDQP_LB/HB_Ratio_1* and the last three RD points using another *CDQP_LB/HB_Ratio_2*, with the first being greater than the second.

$$CDQP\_HB = CDQP\_LB \times CDQP\_LB/HB\_Ratio \tag{11}$$

Another interesting characteristic to note is that the content of the high-frequency bands significantly changes for a number of wavelet decomposition levels higher than two, i.e. the high-frequency bands exhibit only more and more fine details; thus, it is proposed to use a lower *CDQP_HB* value for the decompositions with a higher number of levels.

B. Performance Assessment Methodology

This section describes the methodology adopted to evaluate the RD performance of the proposed coding solution; the assessment framework is displayed in Figure 8.

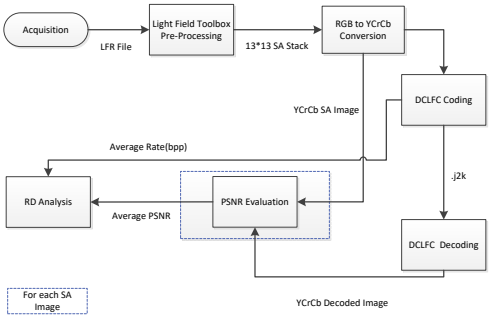


Figure 8: *Processing flow for RD performance assessment.*

The quality is evaluated by computing the PSNR between the original and the corresponding decoded SA images, following Equations (12) and (13), where MAX_I is the maximum valid value that a sample can take, in this case 255 as 8-bit samples are used, *i* and *j* are the dimensions along the *x*- and *y*-axis, and I_1, I_2 are the input and output SA images. The rate is measured in bit-per-pixel (bpp) as shown in Equation (14), where $Size_i$ is the number of bits spent in each compressed image and Homography_rate is the total rate spent in transmitting the homography parameters. To obtain the bit-per-pixel rate, it is necessary to divide the total rate by the number of coded SA images (169) and by the resolution of each one (625×434). Neither JPEG nor JPEG 2000 need homography parameters, thus, the Homography_rate in Equation (14) is equal to zero and the $\sum_{i=0}^{168} Size_i$ is equal to the total number of bits spent to code the SA images, either individually or as "super image".

$$MSE = \frac{1}{i \times j}\sum (I_1 - I_2)^2 \tag{12}$$

$$PSNR\_SA = 10 \times log_{10} \times \left(\frac{MAX_I^2}{MSE}\right) \tag{13}$$

$$rate\_bpp = \frac{\sum_{i=0}^{168} Size_i + Homography\_rate}{169 \times 625 \times 434} \tag{14}$$

When two coding solutions have a similar RD performance, it becomes harder to identify clearly which one is the best. In such cases, the Bjøntegaard Delta metrics [32] are very useful as they are able to express the relative gain between two coding solutions by measuring the average difference between two RD curves. This measurement may be done in terms of rate savings between the two coding solutions (BD-Rate) or in terms of quality gains (BD-PSNR). A negative BD-Rate means that there are rate savings for the studied coding solution regarding an alternative coding solution and vice-versa; in the same context, a positive BD-PSNR implies an increase in quality between the two solutions. For computing the BD-PSNR and BD-Rate values, the following CDQP values were adopted: 65, 16, 8, 6.

C. DCLFC Performance

The proposed light field coding solution may be applied in several different configurations, each offering different trade-offs in terms of compression performance, random access and view scalability. This section intends to study the performance impact of the most relevant DCLFC configurations, notably in terms of RD performance. After, the benchmarking with the alternative solutions and, lastly, the RD performance when compressing each image or band with several quality layers leading to a scalable stream will be studied, already using a stable, best performing coding solution in terms of number of dimensions and decomposition levels.

i) DCLFC over The Horizontal Dimension

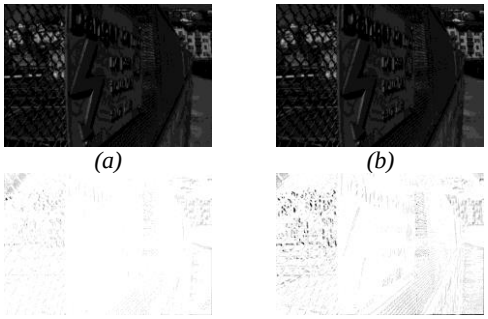
The first case concerns the application of the proposed coding solution to horizontally neighboring SA images; three configurations are possible namely: i) DCLFC H1, DCLFC encoding using only 1-level inter-transform; ii) DCLFC H2, as DCLFC H1 but now using 2-level inter-transform; iii) DCLFC H3, as DCLFC H1 and DCLFC H2 but now using 3-level inter-transform. Table 1 shows the BD-Rate and BD-PSNR for DCLFC H2 and DCLFC H3 in comparison to the JPEG 2000 standard, in the situation where the SA images are coded as a single "super image". With JPEG 2000 there is no inter-prediction between views; the redundancy is exploited mostly within each SA image and thus no relevant inter-view redundancy exploitation is performed. Regarding the RD performance, it is possible to conclude that *Friends* is the light field exhibiting higher gains, while *Danger* is the one with the worst RD performance gains. This is understandable as *Friends* exhibits a more homogenous background while *Danger* includes letters and much more details, thus reducing the proposed solution compression efficiency as there is less redundancy across the views. The conclusion is that DCLFC H2 increases the RD performance, compared to JPEG 2000, moreover it is also possible to conclude that the RD performance is enhanced for the lower rates. Regarding DCLFC H3, the RD performance also increases compared to JPEG 2000, but it is noticeable the diminishing gains; Like DCLFC H2, DCLFC H3 also displays larger RD performance gains for the lower rates; as the high-frequency bands are mostly 'black', it is possible to greatly reduce their size, when compared to the original SA images, and at the same time preserve the average reconstruction quality of the decoded light field. Results for DCLFC H1 solution are not included here due to space limitations.

Table 1: *Bjøntegaard delta results regarding JPEG 2000 with: left) DCLFC H2; right) DCLFC H3.*

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	4.24	-61.44
Fountain	3.02	-46.75
Stone	3.48	-52.95
Danger	3.00	-49.69
Bikes	3.13	-47.56
Average	**3.38**	**-51.68**

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	4.49	-64.24
Fountain	3.10	-48.75
Stone	3.70	-56.76
Danger	3.02	-50.75
Bikes	3.21	-49.36
Average	**3.51**	**-53.97**

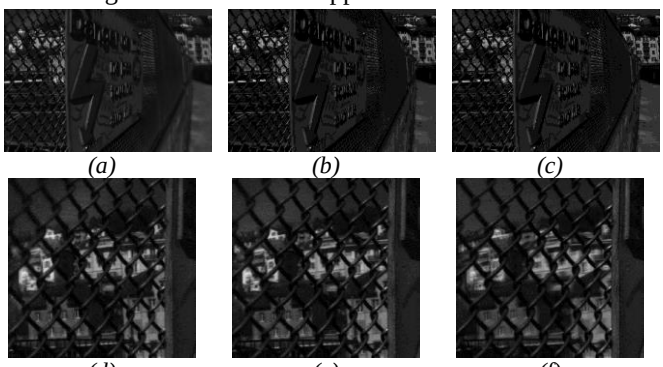
Regarding the number of decomposition levels impact it is possible to conclude that increasing the number of decomposition levels of the inter-transform allows increasing the RD performance although with a reducing gain for any additional level. These results and conclusions can be better understood with the help of the examples in Figure 9. By analyzing the high-frequency bands resulting from the inter-view transform, it is visible that the higher the number of decomposition layers, the smaller is the similarity between the low-frequency bands, which justifies the small RD performance increase for 3 decomposition levels.



(a) *(b)* *(c)* *(d)*

Figure 9: *top) Low-frequency bands from the Danger light field corresponding to inter-transform with: a) 1-level; b) 3-levels; bottom) High-frequency bands for the same situations.*

Figure 9 shows examples of low-frequency and high-frequency bands for the Danger light field for two cases in terms of number of decomposition levels. Note that, as the computed high-frequency bands are almost completely black, and thus its content very hard to perceive, it was necessary to multiply the high-frequency bands sample values by four (if greater than 255, it becomes 255) and then inverting the image (y=255-x) in order a clearer visual analysis of the high-frequency bands content is possible. It is notorious that the bands resultant from higher levels of decomposition exhibit more detail. Despite DCLFC H3 presenting a slightly improvement on the average BD-Rate and average BD-PSNR, the additional complexity and the quality decrease in the low-frequency bands makes DCLFC H2 a better choice for future comparisons. In Figure 10, it is possible to see the decrease in quality from the original SA image to the low-frequency band for DCLFC H3; for example, the buildings behind the fence appear to be blurred.



(a) *(b)* *(c)* *(d)* *(e)* *(f)*

Figure 10: *a) Original SA image; b) Low-frequency band for DCLFC H2; c) Low-frequency band for DCLFC H3; d) Zoom of (a); e) Zoom of (b); f) Zoom of (c).*

These artifacts may result from the update step, when the low-frequency bands, taken as input to the inter-view transform, represent more distant views, which means they are an averaged version of those more distant views, as so some artifacts may be introduced.

The proposed solution, using the same configurations as the ones presented here, was also applied over the vertical dimension, yielding similar, but diminished, RD performance gains; as so, it was not considered important to include the results in this work.

ii) DCLFC Performance over Both Light Field Dimensions

This section presents the performance results and their analysis for the situation where the proposed DCLFC solution is first applied to horizontally neighboring SA images and after applied to vertically neighboring SA images. This was the selected order because overall the application of the DCLFC solution first to horizontally neighboring SA images or low-frequency bands resulted in a better RD performance than the alternative vertical order. Considering the large number of DCLFC configurations that must be considered and to better organize their comparison, the RD results are presented considering a growing total number of levels of decomposition, independently of their direction, i.e. the number of decomposition levels applied horizontally plus the decomposition levels applied vertically.

a) 2-Levels DCLFC Performance

The first case studied concerns the use of 2-levels of decomposition and compares the DCLFC H2 solution with DCLFC H1_V1. The Bjøntegaard delta results for the five light fields are shown in Table 2 and allow concluding: i) the DCLFC H1_V1 solution improves the RD performance compared to the DCLFC H2 solution; implying that it is better to exploit the correlation, between SA images or low-frequency bands, over two light field dimensions opposed to exploit the correlation over just one light field dimension; ii) the behavior of the DCLFC H1_V1 solution against the DCLFC H2 solution is constant across all light fields (naturally, the gains are not the same for all light fields as its content impacts the RD performance).

Table 2: *Bjøntegaard delta results for DCLFC H1_V1 regarding DCLFC H2.*

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	0.18	-3.64
Fountain	0.20	-4.34
Stone	0.15	-5.61
Danger	0.41	-9.14
Bikes	0.30	-6.30
Average	**0.25**	**-5.81**

b) 3-Levels DCLFC Performance

By increasing the number of decomposition levels to three, two different DCLFC solutions become relevant, namely DCLFC H1_V2 and DCLFC H2_V1. While both exploit the correlation between neighboring SA images, horizontally and vertically, the first has a higher number of decomposition levels applied vertically contrarily to the second case. The Bjøntegaard delta results in Table 3 allow concluding that the gains in BD-PSNR and BD-Rate remain similar for all light fields and that both configurations result in similar RD performance improvements.

Table 3: *Bjøntegaard delta results using as reference DCLFC H1_V1: left) DCLFC H2_V1; right) DCLFC H1_V2.*

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	0.83	-18.56
Fountain	0.47	-10.57
Stone	0.77	-16.73
Danger	1.98	-37.78
Bikes	0.49	-10.66
Average	**0.91**	**-18.86**

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	0.75	-16.32
Fountain	0.47	-10.19
Stone	0.66	-14.19
Danger	2.13	-39.46
Bikes	0.50	-10.62
Average	**0.90**	**-18.16**

c) 4-Levels DCLFC Performance

With 4-levels of overall decomposition, it is relevant to study the configurations DCLFC H1_V3, DCLFC H3_V1 and DCLFC H2_V2. Both the first and second solutions were implemented in the usual way, this means first the levels of horizontal decomposition are applied and after the levels of vertical decomposition are applied. However, for the DCLFC H2_V2 solution each level of decomposition of the inter-transform was applied alternately between the horizontal and vertical dimensions. Overall, the studied DCLFC configurations with 4-levels of decomposition presented very different RD performances compared with DCLFC H2_V1. While DCLFC H3_V1 shows a reduced RD performance, DCLFC H1_V3 shows almost no performance differences and DCLFC H2_V2 is the only configuration able to increase the DCLFC H2_V1 performance. It is a bit strange to have such large performance differences between the DCLFC H1_V3 and DCLFC H3_V1 RD performances as until now both the horizontal and vertical approaches resulted in rather similar RD performances, as long as the number of decomposition levels remained the same. Because of this unexpected behavior, all precautions were taken to ensure the correct implementation of the inter-transform scheme and JPEG 2000 compression parameters. For all light fields, there are larger RD gains for the higher rates, opposed to what used to happen for previous configurations, and the gains are rather uniform for the set of selected light fields.

Table 4: *Bjøntegaard delta results regarding DCLFC H2_V1 for: left) DCLFC H3_V1; mid) DCLFC H1_V3 ; right) DCLFC H2_V2.*

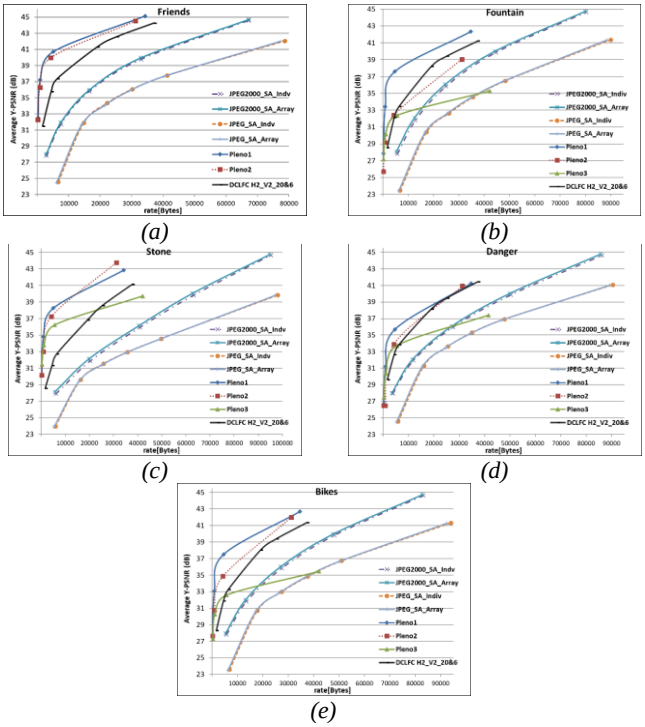
Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	-0.38	10.34
Fountain	-0.36	8.75
Stone	-0.37	11.06
Danger	-0.52	14.01
Bikes	-0.44	10.62
Average	**-0.41**	**10.95**

Light Field	BD-PSNR[dB]	BD-Rate [%]
Fiends	-0.02	1.34
Fountain	0.01	-0.53
Stone	0.02	-1.06
Danger	0.10	-2.90
Bikes	-0.02	0.11
Average	**-0.01**	**0.18**

Light Field	BD-PSNR[dB]	BD-Rate [%]
Friends	0.32	-7.48
Fountain	0.22	-5.36
Stone	0.25	-6.73
Danger	0.18	-5.04
Bikes	0.23	-5.34
Average	**0.24**	**-5.99**

d) Final Benchmarking

Naturally, the most important results regard the direct comparison of the proposed solution with alternative solutions already available. Figure 11 shows the RD performance results comparing the best configuration of the developed DCLFC solution, this mean DCLFC H2_V2, against the most relevant/popular standards available, JPEG 2000, JPEG and the three coding solutions proposed in the context of the JPEG Pleno Call for Proposals on Light Field Coding. JPEG was chosen as it is still the most widely used image codec and JPEG 2000 because it is a standard and here an obvious candidate for comparison. The comparison with JPEG 2000 allows concluding how much influence the inter-view scheme has on the proposed codec RD performance. Concerning the two light field data organizations used both for JPEG and JPEG 2000 encoding, the conclusions were as expected. For JPEG, encoding the SA images individually or as "super image" provides basically the same RD performance. As for JPEG 2000, the performance increases slightly when using the "super image" as input; this is not unexpected as the 2D wavelet transform is performed along the entire image, thus exploiting a bit more of the correlation than the individually coded SAs.



(a) *(b)* *(c)* *(d)* *(e)*

***Figure 11**: Benchmarking RD results for: a) Friends; b) Fountain; c) Stone; d) Danger; e) Bikes*

Concerning the overall RD performance, the proposed DCLFC H2_V2 solution is able to outperform both the JPEG and JPEG 2000 standards, which is understandable as none of these coding solutions provides decorrelation capabilities between neighboring SA images. Table 5 shows the BD-PSNR gains and BD-Rate savings for the DCLFC H2_V2 solution regarding the benchmarks. The Friends light field proved to be the one with best RD performance while Danger proved to be the light field with the worst RD performance. Overall, on average, the DCLFC H2_V2 solution offers BD-PSNR gains of 4.392 dB and BD-Rate savings of 62.846% compared with JPEG 2000. Against JPEG, the gains are even higher which is understandable as JPEG is an older codec, not equipped with such powerful tools as JPEG 2000.

***Table 5**: Bjøntegaard delta results using DCLFC H2_V2 as reference for: left) JPEG2000_SA_Array; right) JPEG_SA_Array.*

Light Field	BD-PSNR[dB]	BD-Rate[%]
Friends	5.61	-72.67
Fountain	3.91	-57.51
Stone	4.67	-65.65
Danger	3.92	-60.97
Bikes	3.84	-57.44
Average	**4.39**	**-62.85**

Light Field	BD-PSNR[dB]	BD-Rate[%]
Friends	9.30	-86.66
Fountain	7.62	-79.81
Stone	5.42	-70.56
Danger	6.64	-78.18
Bikes	7.43	-78.79
Average	**7.28**	**-78.80**

From the JPEG Pleno proposals, two proposals bring significant BD-Rate gains regarding the proposed DCLFC H2_V2 solution, namely Pleno1 and Pleno2. Pleno2 is based on JPEG 2000 and can achieve large BD-PSNR gains and BD-Rate savings, as shown in Table 6; this solution also enables random access and offers the types of scalability naturally provided by the JPEG 2000 codec. However, the Pleno1 proposal based on HEVC is the coding solution achieving the best RD performance, also due to the superior performance of HEVC regarding JPEG 2000. Lastly, the Pleno3 proposal is the only one with similar RD performance when compared to the proposed DCLFC H2_V2 solution; for the *Friends* light field there are no results provided because there was a bug degrading the Pleno3 solution RD performance.

***Table 6:** Bjøntegaard delta results using DCLFC H2_V2 as reference: left) Pleno1; mid) Pleno2; right) Pleno3.*

Light Field	BD-PSNR[dB]	BD-Rate[%]
Friends	3.94	-81.47
Fountain	4.32	-79.01
Stone	5.58	-87.79
Danger	2.23	-52.63
Bikes	4.57	-78.77
Average	**4.13**	**-75.93**

Light Field	BD-PSNR[dB]	BD-Rate[%]
Friends	3.60	-77.55
Fountain	-0.06	-1.52
Stone	5.58	-83.28
Danger	1.21	-31.41
Bikes	2.76	-58.63
Average	**2.62**	**-50.48**

Light Field	BD-PSNR[dB]	BD-Rate[%]
Friends	-	-
Fountain	-1.70	-4.14
Stone	2.99	-73.48
Danger	-0.51	-5.95
Bikes	-1.38	-17.91
Average	**-0.15**	**-25,37**

To conclude this section, Figure 12 shows some low-frequency bands obtained with the DCLFC H2_V2 solution for the Friends and Danger light fields; they show that despite the 4-levels of decomposition applied, the low-frequency bands still look good.

***Figure 12:** Examples of low-frequency bands for DCLFC H2_V2 for: left) Friends; right) Danger.*

e) Quality Scalable Stream Study

The added value of using a quality scalable stream is the possibility to easily decode from the same bitstream a number of different output qualities for the SA images. For the non-scalable case, the *DCLFC* solution was applied using just one quality layer, defined by the value of the *CDQP* applied; this means the JPEG 2000 is configured to provide a single scalability layer. The scalable stream case is achieved by controlling the JPEG 2000 codec asking for three quality layers; in this case, the *CDQP* value defines the first quality layer and the remaining two layers are defined by smaller *CDQP* values. While the non-scalable case allows decoding images with a single quality, the scalable case allows decoding images with three different qualities.

The results in Figure 13 show a slightly decrease in the RD performance when more scalability layers are requested, especially for the higher rates. The decrease in quality for the higher rates may be due to the difference in the compression parameter values used for each SA image or band; this implies that, for a given quality layer, some of the *CDQP* values may significantly change, thus degrading the light field quality reconstruction.

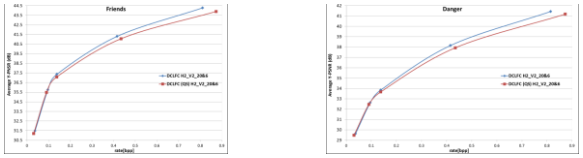


***Figure 13:** RD results with or not quality scalability for: left) Friends; right) Danger.*

Lastly, Figure 14 shows an example of quality scalability; while Figure 14 left) shows the lowest layer decoded image, thus with the lowest quality, Figure 14 right) shows the last layer, using

the totality of the available stream, thus yielding the best quality.

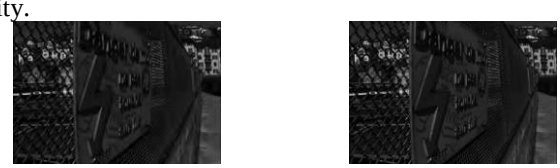


Figure 14: *Danger light field: left) SA image decoded with layer 1 stream; right) SA image decoded with layer 3 stream.*

V. CONCLUSIONS AND FUTURE WORK

From the performance assessment, it can be concluded that proposed solution improves the RD performance of the standard JPEG and JPEG 2000. In summary, the designed disparity compensated lifting based inter-view wavelet transform is able to improve the RD performance compared with only Intra coding; with the best configuration using 4-levels of overall decomposition; Moreover, the RD performance when quality scalability is provided decreases slightly when compared to the situation where only view scalability is provided.

Considering the objectives that the DCLFC solution aims to offer such as view scalability while still maintaining a high compression efficiency, the following research directions could be pursued:

- **Wavelet Transform** – As the solution seems to have reached saturation in terms of RD performance with 4-levels of overall decomposition, to improve the RD performance while still requiring scalability, a more complex wavelet transform as in [13] could be adopted. The biorthogonal Cohen-Daubechies-Feauveau 5/3 (CDF 5/3) wavelet provides a better RD performance compared to the Haar wavelet, in some cases and therefore it is expected to perform well for lenslet light field compression;
- **Homography Parameters** – To improve the RD performance, several geometric (perspective) transformations can be used in different regions of the light field SA images and thus a better prediction can be obtained. This means that the residual that is coded (high-frequency band) will have a lower energy and therefore some bitrate savings can be achieved for the same target decoded quality. The estimation of how many perspective transformations and their respective parameters should be performed with a rate-distortion criterion.

REFERENCES

[1] G. Alves, F. Pereira and E. Silva, "Light Field Imaging Coding: Performance Assessment Methodology and Standards Benchmarking," in *IEEE International Conf. on Multimedia & Expo Workshops*, Seattle, WA, USA, Jul. 2016.

[2] A. Vieira, H. Duarte, C. Perra, L. Tavora and P. Assunção, "Data Formats for High Efficiency Coding of Lytro-Illum Light Fields," in *International Conf. on Image Processing Theory, Tools and Applications*, Orléans, France, Nov. 2015.

[3] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu and W. Zeng, "Pseudo-Sequence-Based Light Field Image Compression," in *IEEE International Conf. on Multimedia & Expo Workshops*, Seattle, WA, USA, Jul. 2016.

[4] C. Conti, P. Nunes and L. Soares, "HEVC-Based Light Field Image Coding with Bi-Predicted Self-Similarity Compensation," in *IEEE International Conf. on Multimedia & Expo Workshops*, Seattle, WA, USA, Jul. 2016.

[5] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. Silva and L. Soares, "Light Field HEVC-Based Image Coding Using Locally Linear Embedding And Self-Similarity Compensated Prediction," in *IEEE International Conf. on Multimedia & Expo Workshops* , Seattle, WA, USA, Jul. 2016.

[6] C. Choudhury and S. Chaudhuri, "Disparity Based Compression Technique for Focused Plenoptic Images," in *Indian Conf. on Computer Vision Graphics and Image Processing*, Bangalore, KA, India, Dec. 2014.

[7] H. Zayed, S. Kishk and H. Ahmed, "3D Wavelets with SPIHT Coding for Integral Imaging Compression," *International Journal of Computer Science and Network Security*, vol. 12, no. 1, pp. 126 - 133, jan. 2012.

[8] A. Aggoun and M. Mazri, "Wavelet-based Compression Algorithm for Still Omnidirectional 3D Integral Images," *Signal, Image and Video Processing*, vol. 2, no. 2, p. 141–153, Jun. 2008.

[9] A. Aggoun, "Compression of 3D Integral Images Using 3D Wavelet Transform," *Journal of Display Technology*, vol. 7, no. 11, pp. 586 - 592, Sep. 2011.

[10] H. Kang, D. Shin and E. Kim, "Compression Scheme of Sub-images Using Karhunen-Loeve Transform in Three-dimensional Integral Imaging," *Optics Communications*, vol. 281, no. 14, p. 3640–3647, Jul. 2008.

[11] E. Elharar, A. Stern, O. Hadar and B. Javidi, "A Hybrid Compression Method for Integral Images Using Discrete Wavelet Transform and Discrete Cosine Transform," *Journal of Display Technology*, vol. 3, no. 3, pp. 321 - 325, Aug. 2007.

[12] S. Kishk, H. Ahmed and H. Helmy, "Integral Images Compression Using Discrete Wavelets and PCA," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 2, pp. 65 - 77, Jun. 2011.

[13] C. Chang, X. Zhu and P. Ramanathan, "Light Field Compression Using Disparity-Compensated Lifting and Shape Adaptation," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 793 - 806, Mar. 2006.

[14] D. G. Dansereau, "Light Field Toolbox v0.4," Feb. 2015. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4. [Accessed 10 01 2017].

[15] JBIG and JPEG, "JPEG Pleno Call for Proposals on Light Field Coding," Doc. number 73013, Changde, China, Oct. 2016.

[16] M. Darbois, "DocJ2KCodec," OPENJPEG, [Online]. Available: https://github.com/uclouvain/openjpeg/wiki/DocJ2KCodec. [Accessed 15 08 2017].

[17] M. Ghobadi, "Wavelet-Based Coding and its Application in JPEG000," SlidePlayer, [Online]. Available: http://slideplayer.com/slide/7978793/. [Accessed 15 08 2017].

[18] M. Marcellin , M. Lepley, A. Bilgin, T. Flohr, T. Chinen and J. Kasner, "An Overview of Quantization in JPEG 2000," *Signal Processing: Image Communication*, vol. 17, no. 1, p. 73–84, Jan. 2002.

[19] "Introduction to SIFT (Scale-Invariant Feature Transform)," OpenCV, [Online]. Available: http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html#keypoint-matching. [Accessed 27 01 2017].

[20] D. Lowe, "Distinctive Image Features from Scale-Invariant Key," *International Journal on Computer Vision*, vol. 60, no. 2, pp. 91-110, Jan. 2004.

[21] G. Zhao and Q. Song, "SIFT Image Stitching Technology Based on Music Score Scanning Recognition System," in *International Symposium on Computational Intelligence and Design* , Hangzhou, China, Dec. 2016.

[22] P. Monteiro, "Distributed Video Coding with Geometric Transforms," Ph.D. Thesis, Instituto Superior Técnico, Lisboa, Portugal, Mar. 2013.

[23] S. Zhao and C. Zhibo, "Lenslet Light Field Image Coding with Linear View Synthesis," ISO/IEC JTC 1/SC29/WG1 Coding of Still Pictures, EEIS Dept., University of science and technology of China, 2017.

[24] I. Tabus, P. Helin and P. Astola, "Lossy Compression of Lenslet Images From Plenoptic Cameras Combinning Sparse Predictive Coding and JPEG 2000," in *75th JPEG Meeting* , Sydney, AUS, Jul. 2017.

[25] Z. Alpaslan, S. Cen, W. Liu, R. Matsubara, H. El-Ghoroury and D. McNeil, "Ostendo Full Parallax Light Field Compression Codec for JPEG Pleno," Ostendo Technologies, Inc., Carlsbad, CA, Jun. 2017.

[26] M. S. P. Group, "Light-Field Image Dataset," [Online]. Available: http://mmspg.epfl.ch/EPFL-light-field-image-dataset. [Accessed 12 06 2017].

[27] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira and P. Schelkens, "ICME 2016 Grand Challenge: Light-Field Image Compression," in IEEE International Conf. on Multimedia & Expo, Seattle, USA, Jul. 2016.

[28] JPEG, ISO/IEC JTC 1/SC29/WG1, "JPEG Call for Proposals on Light Field Coding," Doc. N74014, Geneva, CHE, Jan. 2017.

[29] D. G. Dansereau, "Light Field Toolbox for Matlab," Feb. 2015. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4. [Accessed 04 12 2016].

[30] "OpenCV," OpenCV, [Online]. Available: http://opencv.org/. [Accessed 11 08 2017].

[31] (UCL), Université de Louvain, "OpenJPEG," OpenJPEG, [Online]. Available: http://www.openjpeg.org/. [Accessed 11 08 2017].

[32] B. G., "Calculation of Average PSNR Differences Between RD-curves," VCEG Contribution VCEG-M33, Austin, Texas, Apr. 2011.