Comparison and Evaluation of Light Field Image Coding Approaches

Irene Viola, Martin Řeřábek, and Touradj Ebrahimi, *Member, IEEE*

***Abstract*—The recent advances in light field imaging, supported among others by the introduction of commercially available cameras, e.g., Lytro or Raytrix, are changing the ways in which visual content is captured and processed. Efficient storage and delivery systems for light field images must rely on compression algorithms. Several methods to compress light field images have been proposed recently. However, in-depth evaluations of compression algorithms have rarely been reported. This paper aims at the evaluation of perceived visual quality of light field images and at comparing the performance of a few state-of-the-art algorithms for light field image compression. First, a processing chain for light field image compression and decompression is defined for two typical use cases, professional and consumer. Then, five light field compression algorithms are compared by means of a set of objective and subjective quality assessments. An interactive methodology recently introduced by authors, as well as a passive methodology is used to perform these evaluations. The results provide a useful benchmark for future development of compression solutions for light field images.**

***Index Terms*—Image coding, image compression, light field, objective quality evaluation, subjective quality evaluation.**

I. Introduction

THE idea that light, just as electromagnetism, can be interpreted as a field was first proposed by Michael Faraday in 1846. The concept was subsequently formalized by Andreī Gershun, who coined the term *light field* (LF) in his book on radiometric properties of light in 3D space [1].

One way to represent the LF is to describe the radiance along the light rays in a 3D space with constant illumination. This can be achieved by using the *plenoptic function*, which was first introduced in 1991 by Adelson and Bergen [2]. More particularly, the plenoptic function $\mathcal{L}$ describes the intensity of the light rays passing through every possible point in space (V_x, V_y, V_z) at every possible angle (θ, ϕ), wavelength λ, and time t, represented as follows:

$$\mathcal{L} = \mathcal{L}(\theta, \phi, \lambda, t, V_x, V_y, V_z). \tag{1}$$

Assuming a 3D region free of occlusions at a single time instance and considering the fact that radiance along rays remains constant in a free space, the above 7D plenoptic function can be further simplified into a *4D light field* function [3]. Such 4D function, representing a set of light rays, can be parametrized as an intersection of rays with two planes: uv describing the rays position in aperture (object) plane and xy describing the rays position in image plane.

$$\mathcal{L} = \mathcal{L}(u, v, x, y). \tag{2}$$

The parametrized LF function, further referred to as 4D LF, can be considered as a collection of perspective images of the xy plane, each observed from a position on the uv plane.

In the past, LF images have been mainly used to navigate through 3D scenes. Recently, their applications have expanded thanks to the creation of commercially available devices, such as Raytrix or Lytro cameras, as well as development of new visualization devices capable of properly coping with LF images. However, due to the enhanced features that LF imaging offers, a vast amount of data is created during the acquisition step. Therefore, it is necessary to find an efficient way to compress LF images for transmission and storage.

Currently available techniques to capture and visualize LF images determine two general approaches for LF image compression. A general diagram of workflow for LF image acquisition and visualization is depicted in Fig. 1. The first coding approach assumes that the raw sensor data obtained during the acquisition step is compressed directly with minimal signal preprocessing such as demosaicing or devignetting (point A in Fig. 1). The actual format of raw data strongly depends on the exact acquisition device, e.g., a lenslet based hand-held camera, a multi-camera array, or a multi-view plus depth acquisition device. Often, extensive post-processing of the decompressed LF image is necessary prior to its visualization. Furthermore, additional metadata about the captured scene and acquisition device, e.g., camera and color calibration data, is needed to properly process and visualize the LF image. The second coding approach considers creation of a 4D LF representation of LF image prior to compression (point B in Fig. 1). As mentioned above, the 4D LF represents a collection of perspective images, which can be visualized without a need for acquisition related metadata or post-processing. Process of creation of 4D

Manuscript received December 1, 2016; revised April 11, 2017 and June 19, 2017; accepted August 3, 2017. Date of publication August 15, 2017; date of current version October 23, 2017. This work was conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021_159575) project Light field Image and Video coding and Evaluation (LIVE) and also in the framework of ImmersiaTV under the European Unions Horizon 2020 research and innovation programme (Grant 688619) funded by Swiss State Secretariat for Education, Research and Innovation SERI. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Yebin Liu. *(Corresponding author: Irene Viola.)*

The authors are with the Multimedia Signal Processing Group, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne 1015, Switzerland (e-mail: irene.viola@epfl.ch; touradj.ebrahimi@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2017.2740167

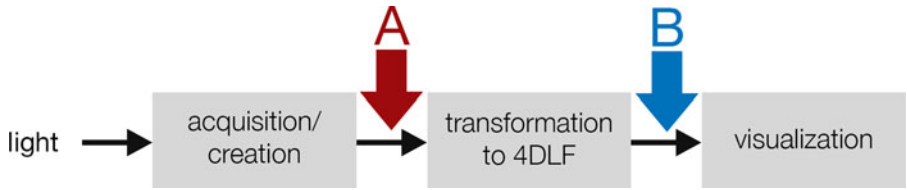

Fig. 1. General acquisition and display pipeline for LF images.

LF from the raw sensor data also strongly depends on the exact acquisition device.

In the context of general LF image manipulation, one can think of two specific use cases related to either of the two coding approaches defined above. On the one hand, professional photographers, operators, and artists may benefit from LF image acquisition technologies, since they allow for greater flexibility in terms of optimal parameters selection after capture. For example, an erroneous selection of focal plane in a scene may lead to several retakes and thus to greater costs. Other features, such as change of point of view or zoom, may dramatically impact the way scenes are captured. In this case, it is of paramount importance that key factors in the acquisition, such as white balance, color, and exposition, are not altered in the compression step, and that acquisition metadata is stored to be used during post-processing.

Consumers, on the other hand, may turn to LF imaging when looking for an enhanced experience to capture a special moment. Ability to change zoom, perspective, and focus in a simple and intuitive way without the need for expensive post-processing software, is in line with the interactivity already seen in applications like Instagram, in which users can modify the appearance of the captured scenes with predefined filters. In this case, the fidelity to the acquisition parameters is less important. However, the resulting image should not be too large and ready to be visualized and shared in devices with limited resources.

In this paper, we compare two main coding approaches to compress LF images through a set of objective and subjective quality evaluations. More specifically, five different compression algorithms suitable to either of predefined use cases are described, investigated, and evaluated. A standard PSNR metric, adjusted to the properties of LF images, is used as objective quality metric.

In the past, the authors have used single viewpoint image visualization to assess the visual quality of LF contents [4]. However, this type of assessment is suboptimal for two main reasons. The first concerns the number of stimuli needed to perform the tests, which increases significantly once multiple viewpoints and refocusing points are selected (in case of the test conducted in this paper, 720 stimuli were tested). The increased number of stimuli leads to strain and fatigue for the subjects and is generally not recommended. Secondly, analysis of variance performed on the results obtained from different viewpoints showed that no significant difference can be found between scores, meaning that subjects would rate different viewpoints from the same content in a similar way. Thus, evaluating different viewpoints leads to a lengthy, more costly and less efficient test, with no visible gain. For the reasons mentioned above, two methodologies have been selected to subjectively assess the visual quality of LF image contents. The first methodology was recently introduced by the authors and allows for interaction with the displayed content [5]. The second methodology passively shows the different viewpoints composing the LF content, in order to ensure that all users see and rate the exact same content. The two methodologies have been analysed by the authors and were found to be highly correlated [6]. The two methodologies are described in details and applied to the problem of assessing the visual quality of the five compression algorithms.

The remainder of the paper is organized as follows. Section II presents the state of the art in related areas of LF imaging, specifically LF image acquisition, compression, and quality evaluation. Section III discusses in details the two approaches examined in the paper. Section IV presents our experiments in details, whereas Section V exposes and comments the results. Section VI concludes the paper.

II. State of the Art

This section briefly reviews the state of the art in LF imaging. First, a quick overview of LF image acquisition techniques is presented. Then, existing algorithms to compress LF images are discussed in relation to the requirements of previously defined use cases. Finally, currently available objective and subjective evaluation methodologies to assess the quality of LF images are presented.

A. Light Field Image Acquisition

A digital 4D LF, which is a collection of perspective images, can be obtained by sampling the 4D LF function defined in (2). The density of the sampling depends on the acquisition technology used to capture the LF image. In general, depending on the requirements for baseline, different acquisition techniques can be used to capture LF images. More specifically, for baseline in a range of meters, one way of acquiring LF images is by means of a moving camera. The idea behind is that by moving a single camera throughout the scene, we can acquire the different perspective images that form our 4D LF. In this case, the sampling in xy plane depends on the camera resolution, and sampling in uv plane depends on the position of the capturing device and its shutter speed. Examples of such acquisition devices are the Stanford Spherical Gantry, a motorized gantry with four degrees of freedom that can be used to capture 4D LF [7] and Apple's setup to construct 360-degree cylindrical panoramic images [8]. 4D LF can also be acquired by using hand-held cameras, as long as their position on the uv plane can be precisely estimated [9].

Another approach is to construct an array of cameras with synchronized shutter speed capturing the 4D LF at once. In this case, the uv sampling depends on baseline parameter of the camera array grid. Using a camera array, a full 4D LF is formed and new views corresponding to narrower baseline parameter must be further synthesized if needed. An example of such acquisition technology is the Stanford Multi-Camera Array [10].

LF images can also be acquired from multi-view plus depth data [11]. In this case, the baseline can be wide or narrow, depending on how the data was created [12].

For LF image acquisition with narrow baseline a hand-held plenoptic camera capturing so called "single lens stereo" can be achieved [13]. It exploits an array of micro lenses placed in front of the sensor plane [14]. The aperture size of the main lens limits the possible viewpoints. The spatial resolution (xy plane) depends on the number of micro lenses, whereas the angular resolution (uv plane) depends on the number of pixels behind each micro lens. Hand-held cameras implementing this model were presented in [15] and are already widely available to consumers.[1,2] The raw image obtained with this type of cameras closely resembles the honeycomb array of lenses that has been used for the acquisition, and will be from now on referred to as a lenslet image. It is possible to convert the lenslet image to a 4D LF that effectively constitutes a sampling of (2).

B. Light Field Image Compression

Acquiring or synthesizing LF images creates a vast amount of data (around 150 MB for lenslet images with 15×15 viewpoints of resolution of 635×434, around 6.8 GB for 15×15 4K images acquired with a multi-camera array). Thus, a lot of research has been recently focused on finding efficient compression algorithms to effectively store and transmit LF images. Depending on the acquisition process, several approaches have been proposed. Early work focused on compressing synthetic 4D LF using disparity compensation [16]–[18] and geometry estimation [19], and enabling random access coding [3].

More recently, the effort has been focused on compression of LF images acquired through hand-held devices. Several compression algorithms have been proposed to directly compress lenslet images through intra coding, exploiting redundancies in its structure. For example, Perra proposed a lossless compression scheme based on adaptive prediction [20]. Li *et al.* incorporated a full inter prediction scheme in HEVC intra prediction, explicitly exploiting the redundancy in lenslet images [21], as well as using the disparity compensation and impainting to efficiently code lenslet images [22]. More recently, five proposals for lenslet image compression have been collected within the ICME 2016 Grand Challenge. Three proposals relied on improving HEVC intra compression efficiency by exploiting the redundancies in the lenslet image [23]–[25], whereas the rest used pseudo-temporal sequences to code the lenslet images [26], [27].

Another approach is based on compression of 4D LF, which can be created from raw lenslet image data using specific transformations. A precursor of the approach is proposed by Olsson *et al.* [28]. They propose the creation of sub-images from integral images. Such sub-images are then encoded through a pseudo-sequence using H.264. Choudhury *et al.* proposed to adapt the method of coded snapshots to LF image compression through random codes [29]. Dai *et al.* coded sub-aperture images using different scanning methods, including line and rotating scanning [30]. Helin *et al.* proposed predictive coding for sub-aperture views to achieve lossless compression [31].

No standard approach has been agreed on to compress LF images. Nevertheless, to help finding a standard representation for LF images, the JPEG standardization committee launched in 2014 a new activity called JPEG Pleno. Its goal is to create a standard framework for efficient storage and transmission of not only LF images, but also point-cloud, holographic and other plenoptic content. In particular, JPEG Pleno aims at finding an efficient way to represent plenoptic content, while, where needed, also offering compatibility with existing solutions, such as JPEG and JPEG 2000. In such a framework, it is essential to analyse and compare different approaches for compression of LF images as one of the plenoptic content variations, especially in relation to different use cases, in order to provide a solid guideline towards the creation of a standard method. Additionally, a call for proposals for both lenslet and high density camera array for 4D LF compression, aiming at the definition of a standard for compression of plenoptic content, has been issued during the 73^{rd} JPEG Meeting [32].

C. Light Field Image Quality Evaluation

Evaluation of visual quality is essential to design and improvement of coding solutions for LF images. Several publications have been devoted to comparison and evaluation of state of the art standard solutions. Alves *et al.* assessed the performance of existing still image coding solutions, such as JPEG 2000 and AVC, on lenslet images [33]. The objective evaluation was carried out using PSNR as a full reference metric. Similarly, Vieira *et al.* compared five different HEVC compatible coding of lenslet images with different data formats [34], again using PSNR as a full reference metric. Rizkallah *et al.* reported the impact of compression of LF images on refocusing and extended focus images through objective metrics [35].

A Grand Challenge was organized at ICME 2016 under a collaboration between Qualinet and JPEG standardization committee. The goal was to collect new compression solutions for LF images, and to evaluate them using both objective and subjective quality assessment methodologies [4]. The Grand Challenge, however, had some limitations. For starters, it required to have lenslet images in YUV 420 format as input of the compression, as well as output for the decompression. Moreover, the reference was obtained by performing the transformation to 4D LF after chroma subsampling was applied on the raw lenslet image, rather than transforming directly the lenslet image in RGB 444. The subjective assessments were carried out on five discrete views from each content, which were evaluated separately. Since the assessment was conducted separately on predefined

[1]https://www.raytrix.de/
[2]https://www.lytro.com/

views, it did not address the issue of evaluating global quality of experience offered by a compressed LF image.

Recently, we proposed a new methodology to evaluate a plenoptic content in an interactive way [5] allowing users to interact with LF images, visualize different views, apply refocusing, and globally evaluate the quality of LF images. To best of our knowledge, no extensive analysis has been carried out to assess which coding approach for LF image compression yields better results in terms of visual quality.

III. Light Field Coding Strategies

This section describes in details the coding approaches investigated in the evaluation process, including a thorough description of five selected algorithms to compress LF images.

Two main coding approaches can be considered for compression of LF images. Referring to the general diagram of workflow presented in Fig. 1, we can compress the data at two different stages. Compression can be performed on the raw sensor data that has been captured with the selected acquisition technology, after minimal processing, such as demosaicing and devignetting (point A in Fig. 1). The 4D LF can be recovered from the decompressed bitstream through extensive post-processing, involving camera and color calibration metadata which needs to be sent along with the bitstream. The second coding approach performs compression on the 4D LF obtained from the raw data (point B in Fig. 1). The 4D LF is a collection of perspective views which can be visualized as they are, or combined to create new interpolated views, synthetic aperture, refocusing, and extended focus. Since the transformation of raw sensor image data to 4D LF is performed before the compression step, no metadata is required for visualization. The compression solutions used to code the raw sensor data, as well as the transformation to 4D LF from the raw sensor data, strongly depend on the acquisition technology used to capture LF images. If compression is applied at point A, the selected scheme will profoundly differ based on the acquisition technology. On the other hand, a compression scheme operating at point B can compress 4D LF image information captured with any acquisition technology.

In order to compare the two coding approaches on a common ground, we decided to focus our attention on evaluating coding strategies for lenslet-based acquisition. Lenslet-based acquisition allows to compare the two approaches on the same image content captured within the same conditions. In this case, the raw sensor data is minimally pre-processed to obtain a lenslet image. From the lenslet image, the 4D LF can be recovered through rectification, calibration and extraction of perspective images, using camera and color calibration data. The extraction of perspective images from the lenslet image generates $N \times M$ views, depending on the uv resolution. However, the most external views contain too many distortions to be properly visualized. The 4D LF coding approach can take advantage of this fact by not coding the most distorted views, which will likely not be used in the visualization process, thus further reducing the size of the bitstream.

The rest of the section is organized as follows. The first coding approach, which deals with compression of lenslet images, is described. Then, the second coding approach, which focuses on compression of 4D LF obtained from lenslet images, is presented. Finally, one hybrid approach to compress lenslet images through transformation to 4D LF, introduced in ICME 2016 Grand Challenge, is detailed. Authors are aware of practical drawbacks and flaws in this solution. However, it was decided to include it in the evaluation process because of its optimal performance within the Grand Challenge, and because it represents a transition point between the two coding approaches. A summary of the compression schemes can be found in Table I.

A. Lenslet Image Compression

The lenslet coding approach performs compression on the lenslet image, obtained from the raw sensor data after demosaicing and devignetting. Fig. 2 depicts the workflow for the coding approach. The workflow was adopted following the definition of the ICME 2016 Grand Challenge, which required to perform compression on YUV 420 lenslet images in 8-bit precision.

The raw sensor data is first demosaiced, devignetted and clipped to 8 bits to obtain a lenslet image. The lenslet image is subsequently converted to YUV 420 format, and compressed and decompressed using the selected compression scheme. The output of the decompression step is then upsampled and converted to RGB 444. Conversion to RGB 444 format is required to perform the transformation from lenslet image to 4D LF. The 4D LF is created from the decompressed lenslet image using camera metadata. Color and gamma corrections are applied separately on each view. The perspective views forming the 4D LF can subsequentially be visualized on commercially available displays, or combined to create new interpolated views, synthetic aperture or refocusing effect.

Two compression algorithms ($P01$ and $P02$) compliant with the workflow depicted in Fig. 2 were evaluated. One compression scheme was selected among the best performing submitted to ICME 2016 Grand Challenge, and it was compared to an HEVC anchor. The anchor $P01$ exploits HEVC intra profile with default settings to compress the YUV 420 lenslet image. To perform the compression, the software x265 was used.[3] The second algorithm $P02$ uses a modified version of HEVC intra profile, which integrates Locally Linear Embedding (LLE) and Self Similarity (SS) to exploit the redundancies in the lenslet structure [25]. The image is partitioned into blocks using HEVC intra prediction scheme. LLE estimates the current block by selecting the best linear combination of k nearest neighbors through a least-square optimization problem. SS predicts the current block using the best among two blocks, one given by best block matching in the search window, the other chosen by searching for best linear combination between the first selected block and another block in the search window. The codec performs both LLE and SS, and then choses the prediction method that gives the smallest rate distortion cost.

[3]https://www.videolan.org/developers/x265.html

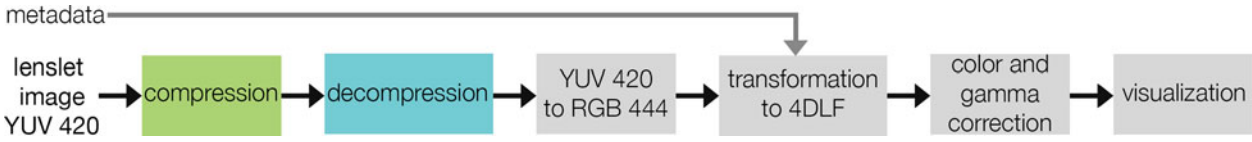

Fig. 2. Processing chain for lenslet image compression used for two compression algorithms (anchor $P01$, proponent $P02$).

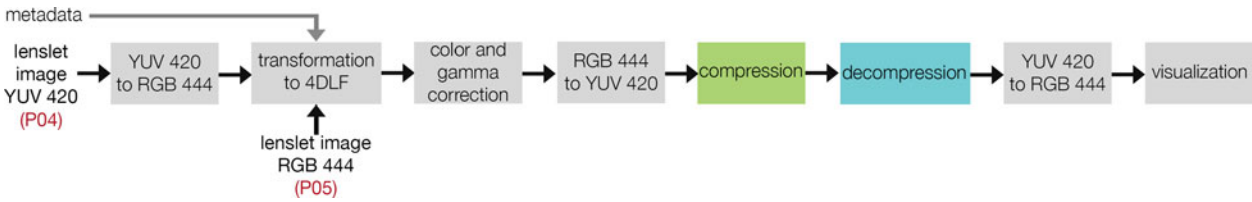

Fig. 3. Processing chain for 4D LF compression used for two compression algorithms (anchors $P04$ and $P05$).

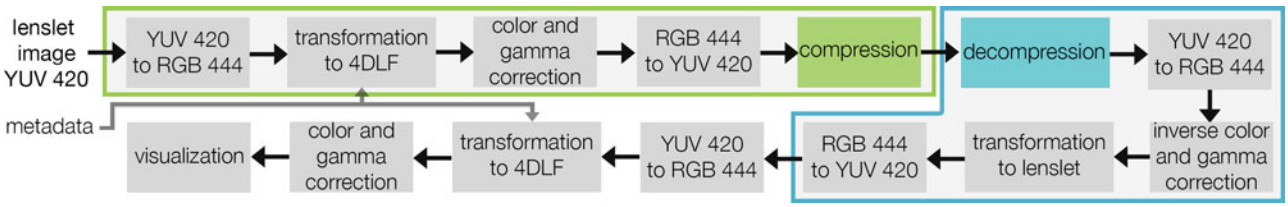

Fig. 4. Processing chain for hybrid compression of lenslet using intermediate 4D LF transformation (proponent $P03$). The green and blue blocks highlight how the compression step involves intermediate transformation to 4D LF, and the decompression step involves the inverse transformation to lenslet image.

B. 4D Light Field Compression

The 4D LF coding approach performs the compression on the 4D LF, obtained from the lenslet image, after color and gamma corrections. Fig. 3 depicts the workflow for this coding approach. Two anchors were created to assess the visual quality of this coding approach. The first anchor $P04$ assumes the same input as the compression schemes using the lenslet coding approach (YUV 420 lenslet images). The color space is then upsampled and converted again to RGB 444, to be used in the transformation process. To assess the effect of chroma subsampling of the lenslet image on the resulting quality of the final 4D LF, the second anchor $P05$ performs the compression on the 4D LF, obtained from lenslet in RGB 444 format, after color and gamma corrections (Fig. 3). In this case, the lenslet image is not transformed to YUV from RGB, and the color space is not subsampled before the transformation.

For both anchors, the 4D LF is created from the uncompressed lenslet image using camera metadata, and color and gamma corrections are applied separately on each view, prior to compression. Each view is converted from RGB 444 to YUV 420. The views are arranged in a pseudo-temporal sequence in spiral order, as depicted in Fig. 6. Due to the geometrical distortions present in the most external views, only a subset of the views is coded. Specifically, only the 13×13 internal views out of 15×15 views are encoded. The pseudo-temporal sequence is coded with HEVC software x265. In the decompression step, the views which have not been coded are replaced with copies of neighboring views, to reconstruct the 15×15 images that compose the 4D LF. After decompression, the views are upsampled and converted to RGB 444 and rearranged in the 4D LF. The perspective images composing the 4D LF can then be visualized on commercially available displays, or combined to create synthetic aperture, refocusing and new interpolated views.

C. Hybrid Compression of Lenslet Images

Among the participants to the ICME 2016 Grand Challenge, which required to have YUV 420 lenslet image as input and output of the compression and decompression step, one algorithm performed compression on lenslet images using intermediate transformation to 4D LF [26]. Fig. 4 depicts the workflow for this algorithm.

The algorithm $P03$ proposes a compression of 4D LF images based on pseudo-sequences of perspective views. Due to the constraints of the Grand Challenge, the YUV 420 lenslet image is first converted to RGB 444 color space, to be used in the transformation step. Then the lenslet is processed to obtain the perspective views that compose the 4D LF. The views are color and gamma corrected and then converted back to YUV 420. A subset of them is then rearranged in a specific coding order, that accounts for similarities between adjacent views, and coded using the JEM encoder.[4] In the decompression step, the views are rearranged in the 4D LF. Inverse color and gamma corrections are applied and the lenslet image is formed following the inverse process of the transformation to 4D LF.

The conversion from lenslet to 4D LF and back was needed to be compliant with the requirements of the grand challenge. However, it can be clearly seen that the proposed approach is

[4]https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/tags/HM-16.6-JEM-2.0rc1/

TABLE I
SUMMARY OF COMPRESSION SCHEMES

Proponents	Description
P01	Anchor: Lenslet image compressed using HEVC intra (default settings of x265 software implementation).
P02	Lenslet image compressed using HEVC intra with LLE and SS (software HM-14.0) [25].
P03	Hybrid approach: lenslet image compressed using intermediate transformation to 4D LF and HEVC (software JEM 2.0) [26].
P04	Anchor: 4D LF, obtained after chroma subsampling, compressed using HEVC (default settings of x265 software implementation).
P05	Anchor: 4D LF compressed using HEVC (default settings of x265 software implementation).

Fig. 5. Central viewpoint image from each content used in our experiment. (a) Bikes. (b) Stone_Pillars_Outside. (c) Fountain_&_Vincent_2. (d) Friends_1.

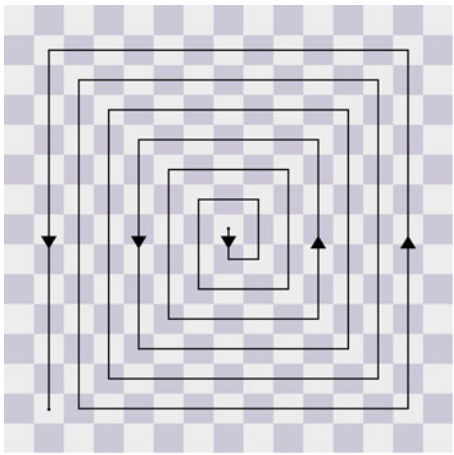

Fig. 6. Ordering of the views for coding.

hybrid, in the sense that it compresses lenslet images through transformation to 4D LF. The tranformation from lenslet images to 4D LF and back is lossless, as it is defined in [26].

IV. LIGHT FIELD QUALITY ASSESSMENT EXPERIMENT

This section describes the evaluation process in details. First, the data preparation process is presented, along with the coding conditions. Methodologies and metrics for objective and subjective evaluation are then presented in details.

A. Dataset Preparation and Coding Conditions

Four LF images, acquired by a Lytro Illum camera, were selected from the publicly available EPFL LF image dataset [36]. More specifically, *Bikes*, *Stone_Pillars_Outside*, *Fountain_&_Vincent_2* and *Friends_1* contents were selected for our experiments. The central view of each content used is depicted in Fig. 5. The images were carefully selected from those used in the ICME 2016 Grand Challenge [4] in order to provide a wide range of scenarios, containing details that would prove challenging for the compression algorithms. To obtain the 4D LF, the lenslet images were processed using the LF MATLAB toolbox [37], [38].

The compression algorithms were evaluated on four bitrates (corresponding to four compression ratios), namely $R1 = 1$ bpp $(10:1)$, $R2 = 0.5$ bpp $(20:1)$, $R3 = 0.25$ bpp $(40:1)$, $R4 = 0.1$ bpp $(100:1)$. The compression ratios are computed as ratios between the size of the uncompressed raw images in 10bit precision ($5368 \times 7728 \times 10$ bits $= 414\,839\,040$ bits) and the size of the compressed bitstreams.

The uncompressed reference was obtained by demosaicing, devignetting and clipping to 8 bits the raw sensor data, transforming it to 4D LF and applying color and gamma corrections. Unlike the reference used in ICME 2016 Grand Challenge, which used as a reference the 4D LF obtained from YUV 420 lenslet image, we obtain our reference from the lenslet image in RGB 444, without any chroma subsampling. This reference was selected to have a proper comparison with acquisition data obtained with minimal pre-processing. For this reason, chroma subsampling was not applied on the reference, since it alters the data.

A total of five compression schemes were evaluated. Each compression scheme was given a label, as stated before, for easier identification. A summary of the compression schemes can be found in Table I. It should be noted that the Quantization Parameters (QP) were selected to match the bitrates described above.

B. Objective Evaluation

To analyze the performance of evaluated coding schemes, PSNR was selected as a full reference metric. PSNR values were computed with respect to the uncompressed reference. The computation is thus performed on the 4D LF after color and gamma corrections (point B in Fig. 1). The PSNR metric was

TABLE II
VALUES OF REFOCUSING SLOPE FOR EACH CONTENT

	Slopes										
Content	1	2	3	4	5	6	7	8	9	10	11
Bikes	−10	−8	−6	−4	−2	0	2	4	6	8	10
Stone_Pillars_Outside	−10	−8	−6	−4	−2	0	2	4	6	8	10
Fountain_&_Vincent_2	−10	−8	−6	−4	−2	0	2	4	6	8	10
Friends_1	−5	−4	−3	−2	−1	0	1	2	3	4	5

adapted to better suit properties of LF images. Therefore, the PSNR value is computed on the Y channel as follows:

$$PSNR_Y(k,l) = 10\log_{10}\frac{255^2}{MSE(k,l)}, \tag{3}$$

in which k and l are the indexes of the acquired views. The $MSE(k,l)$ for each image is computed as follows:

$$MSE(k,l) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - R(i,j)]^2, \tag{4}$$

where m and n are the dimensions of one viewpoint image (i.e., $n = 625$, $m = 434$). $I(i,j)$ is the Y value for the selected acquired view in the evaluated 4D LF, whereas $R(i,j)$ is the corresponding value in the reference 4D LF. In the same way, the PSNR for the other two channels U and V is obtained. A weighted average [39] is then computed as follows:

$$PSNR_{YUV}(k,l) = \frac{6PSNR_Y(k,l) + PSNR_U(k,l) + PSNR_V(k,l)}{8} \tag{5}$$

The mean of all viewpoint images is subsequentially computed to have an average value for PSNR for Y channel and for YUV:

$$PSNR_{X_{mean}} = \frac{1}{(K-2)(L-2)}\sum_{k=2}^{K-1}\sum_{l=2}^{L-1}PSNR_X(k,l), \tag{6}$$

in which $K = 15$ and $L = 15$ represent the number of perspective views, and $X = Y$ and $X = YUV$ for Y channel and for YUV channels, respectively.

C. Subjective Evaluation

1) Interactive Test Methodology: A recently introduced methodology for interactive evaluation of plenoptic content was selected to perform the first of the two visual assessments [5]. The methodology is based on Double Stimulus Impairment Scale (DSIS) [40].

The participants were asked to interact with the LF images and to rate the level of impairment of the test LF image with respect to the reference, on a scale from 1 (*Very annoying*) to 5 (*Imperceptible*). Each LF image was presented together with the uncompressed reference in a side-by-side fashion. The position of the reference was set to either left or right for each experiment, and the participants were informed about its location on the screen.

For each stimulus, the central viewpoint image from the 4D LF was displayed. By clicking inside the displayed image and dragging the mouse, the other viewpoints from the 4D LF were accessed and displayed. Each image was displayed in its native resolution of 625×434 pixels.

Eleven refocused images were created for each content, using a modified version of the toolbox function *LFFiltShiftSum*. The function shifts all the perspective views according to a parameter, called slope, and performs a sum of the shifted images to obtain a single image that is refocused on a specific plane, which depends on the value of the slope. The number of images to be shifted and consequently summed defines the depth of field. Summing all 15×15 images creates the smallest depth of field, in which only one specific plane in the image is in focus. On the other hand, taking just the central image, which is equivalent to summing just 1×1 images, brings all the objects in focus (largest depth of field). For our tests, it was decided to sum images from index 3 to index 13 (11×11 images) to have a larger depth of field that still shows an effects of refocusing. The values of the slopes are summarized in Table II. The refocused images were accessible through a slider shown at the bottom of each stimulus. The slopes were selected so as to assure gradual transition between refocusing on the foreground and on the background with respect to semantically relevant objects in each content.

Before the experiments, a training session was organized to allow participants to get familiar with artifacts and distorsions in the test images. Five training samples were manually selected by expert viewers. In order not to influence the results, the training samples were created by compressing other content on various bitrates. The content used for the training was chosen from the same LF image database used for the test images [36]. The training samples were presented along with the uncompressed reference, exactly as they were shown in the test.

The experiment was split into two sessions. In each session, 40 stimuli were shown side by side with the uncompressed reference, corresponding to approximately 20 minutes per session. The display order of the stimuli was randomized, and the same content was never displayed twice in a row. Each subject took part in all the sessions, thus evaluating the entire set of stimuli. A break of ten minutes was enforced between the sessions to avoid fatigue. Before the test, one dummy sample was inserted to ease the participants into the task. The resulting scores from dummy stimuli were not included in the results.

A total of 24 subjects (19 males and 5 females) participated in the experiment, for a total of 24 scores per stimulus. Subjects

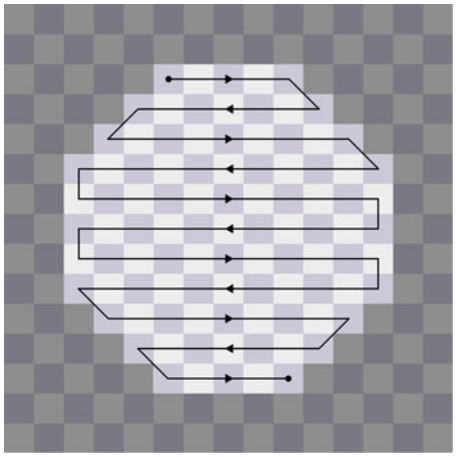

Fig. 7. Ordering of the views for animation for passive methodology.

were between 18 and 35 years old, with an average of 24.8 and a median of 25 years of age. All subjects were screened for correct visual acuity with Snellen charts, and color vision using Ishihara charts.

2) Passive Test Methodology: The second of the two visual assessments of quality was performed using a passive methodology, to ensure that all participants would visualize and rate exactly the same viewpoints and refocused views. The methodology is based on DSIS.

The participants were shown the LF content as a video sequence navigating between the viewpoints and the refocused images. Each stimulus was displayed alongside with the uncompressed reference, in a side by side fashion, and the subjects knew in advance on which side of the screen the reference was displayed.

Due to distortions caused by the lenslet structure, several viewpoints presented artefacts independent from the coding procedure, and thus had to be discarded, in order not to negatively influence the results. Therefore, only a subset of 97 out of 225 viewpoints was chosen to be displayed. Ten viewpoints per second were displayed, to ensure a smooth transition of the different viewpoints. The viewpoints were accessed from top to bottom and from left to right and right to left in alternate order (Fig. 7). At the end of the viewpoint animation, the eleven refocused images were displayed with a framerate of four refocused images per second, going from foreground to background and from background to foreground. The animation setup was chosen and validated by expert viewers in order to mimic the parallax effect, as well as to mimic the refocusing effect that occurs when trying to change the focal point. The total length of the animation for each stimulus was 14 seconds.

Test subjects were asked to rate the level of impairment of the test stimuli when compared to the uncompressed references. The rating was performed on a scale from 1 (Very annoying) to 5 (Imperceptible). Before the experiment, a training session was organized to allow participants to get familiar with artefacts and distorsions in the test images. Five training samples were manually selected by expert viewers. To help subjects localize and identify compression artefacts in the fast-paced video, the same content used in the test was selected for the training. The training samples were presented along with the uncompressed reference, exactly as they were shown in the test.

The experiment was split into two sessions. In each session, 40 stimuli were shown side by side with the uncompressed reference, corresponding to approximately 20 minutes per session. The display order of the stimuli was randomized, and the same content was never displayed twice in a row. Each subject took part in all the sessions, thus evaluating the entire set of stimuli. A break of ten minutes was enforced between the sessions to avoid fatigue.

A total of 29 subjects (24 males and 5 females) participated in the experiment, for a total of 29 scores per stimulus. Subjects were between 18 and 35 years old, with an average and median of 23 years of age.

3) Interactive Test Environment: To avoid the involuntary influence of external factors and to ensure the reproducibility of results, the laboratory for subjective video quality assessment was set up according to ITU recommendation BT.500-13 [40]. Professional Eizo ColorEdge CG301W 30-inch monitors with native resolution of 2560×1600 pixels were used for the test. The background color of the display was set to mid grey, according to requirements in ITU Recommendation ITU-R BT.2022 [41]. The monitors were calibrated using an i1Display Pro color calibration device according to the following profile: sRGB Gamut, D65 white point, 120 cd/m^2 brightness, and minimum black level of 0.2 cd/m^2. The room was equipped with a controlled lighting system that consisted of neon lamps with 6500 K color temperature, while the color of all the background walls and curtains present in the test area was mid grey. The illumination level measured on the screens was 15 lux. The distance of the subjects from the monitor was approximately equal to 7 times the height of the displayed content, conforming to requirements in ITU Recommendation ITU-R BT.2022 [41].

4) Passive Test Environment: To perform the tests, the QualityCrowd 2 framework [42] was used. Nonetheless, it should be noted that all the participants performed the tests in the same environment at the same time, with equal lighting conditions, using the same display model and the same screen resolution.

Since there is no browser video plugin capable of reliable real-time decoding and displaying for HEVC, the animations were encoded with AVC. A two-pass encoding was used and the deblocking filter was disabled to ensure transparency and to preserve the original blockiness artefacts when encoded at low bit rates. Expert viewing session conducted prior to the main subjective assessment concluded that the AVC video encoding was visually lossless. Selected settings for AVC coder are summarised in Table III.

5) Data Analysis: Outlier detection was performed according to the guidelines defined in ITU recommendation BT.500-13 [40]. One outlier was detected in both interactive and passive tests, and the relative scores were discarded, thus leading to 23 and 28 scores per stimulus, respectively. The Mean Opinion Score (MOS) was computed, separately for each methodology, for each coding condition j (i.e., each content, codec and compression ratio) as follows:

$$MOS_j = \frac{1}{N} \sum_{i=1}^{N} m_{ij}, \tag{7}$$

TABLE III
SELECTED SETTINGS FOR AVC CODER FOR PASSIVE METHODOLOGY

-r 30 -s <*size*> -f rawvideo -pix_fmt yuv420p -i <*input*> -c:v libx264 -profile:v high -x264opts no-scenecut:no-deblock:pass=1 -b:v 8M tmp.mp4
-r 30 -s <*size*> -f rawvideo -pix_fmt yuv420p -i <*input*> -c:v libx264 -profile:v high -x264opts no-scenecut:no-deblock:pass=2 -b:v 8M <*output*>

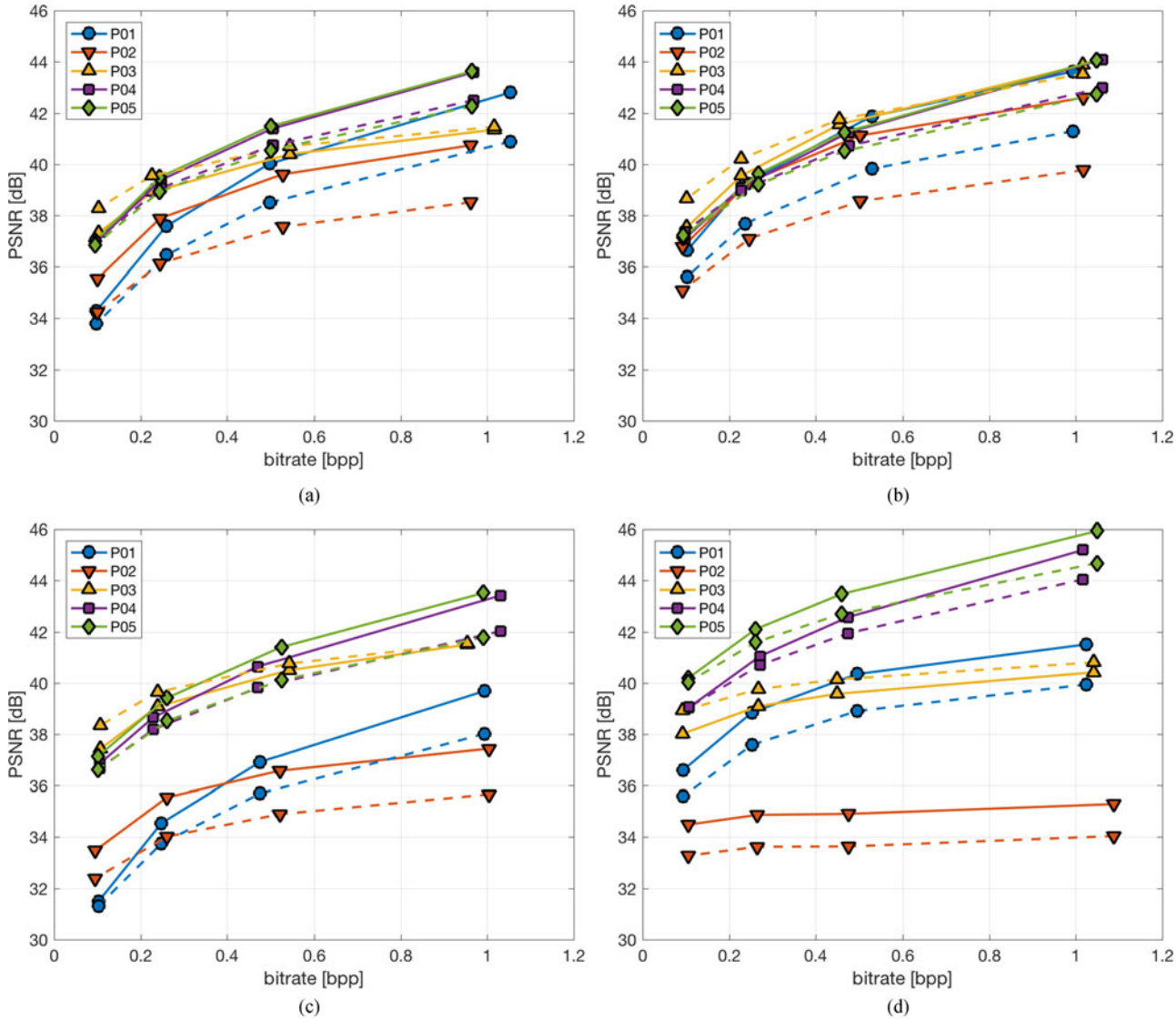

Fig. 8. Rate distortion plots for Y channel (solid line) and for YUV channels (dashed line). PSNR was computed on the 4D LF after color and gamma correction. (a) Bikes. (b) Stone_Pillars_Outside. (c) Fountain_&_Vincent_2. (d) Friends_1.

where N is the number of participants and m_{ij} is the score for stimulus j by participant i. The corresponding 95% confidence intervals were computed. To determine whether the results yield statistical significance, a one-sided Welch's test at 5% significance level was performed on the scores, with the following hypotheses:

$$H0 : MOS_A \leq MOS_B$$
$$H1 : MOS_A > MOS_B \,,$$

in which A and B are the codecs that are being compared. The test was performed for each compression ratio and for each content, separately for each methodology. If the null hypothesis were to be rejected, then it could be concluded that codec A performed better than codec B for the given content and compression ratio at a 5% significance level.

V. RESULTS AND DISCUSSION

In this section, results of the objective and subjective quality assessments are presented. Results on the coding approaches presented in Section III will be discussed separately. First, the lenslet image compression is analyzed. Then, the 4D LF compression is discussed. The hybrid approach is compared to the

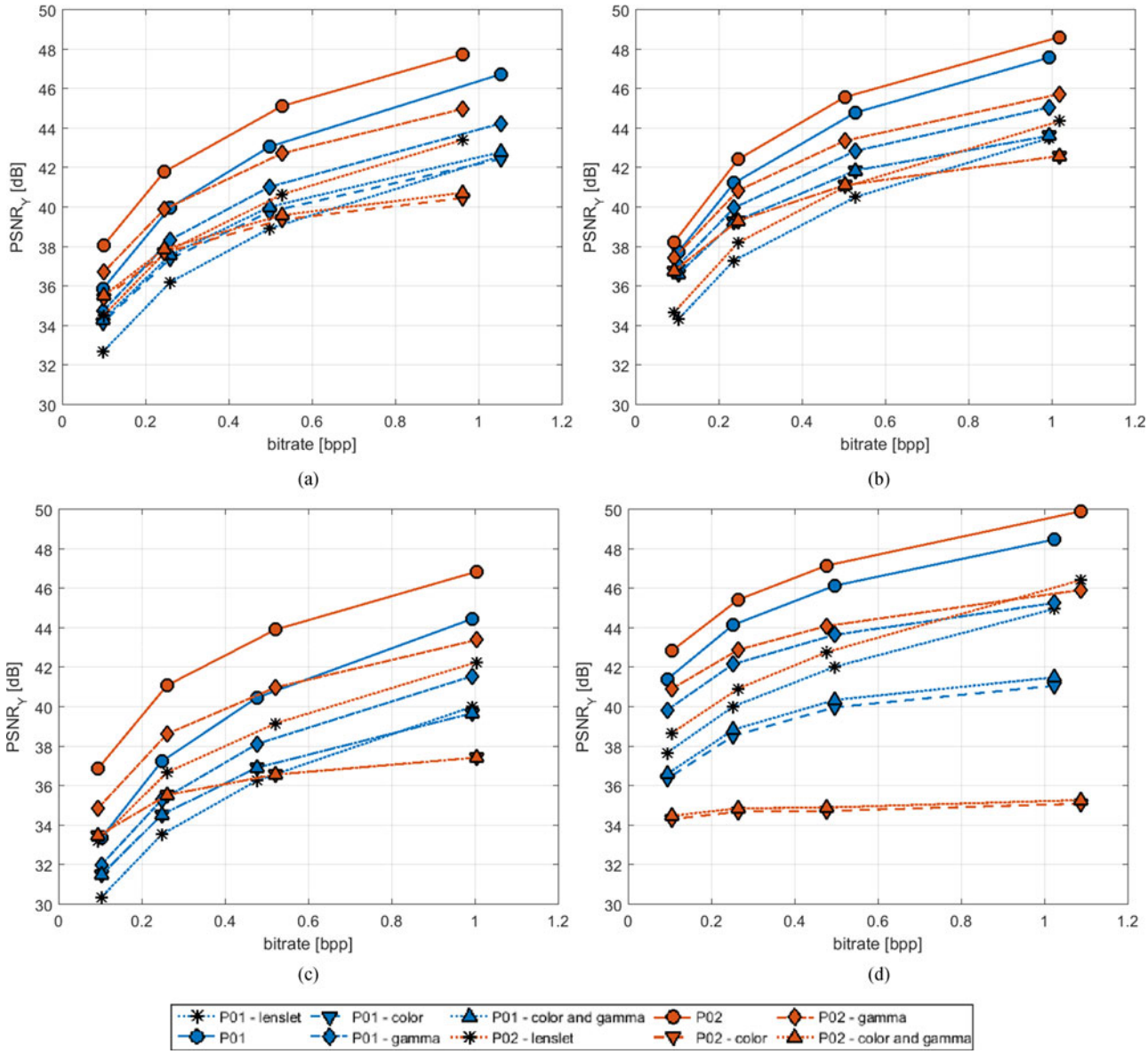

Fig. 9. Rate distorsion plots for Y channel. PSNR was computed at various stage of the pipeline (See Fig. 2). (a) Bikes. (b) Stone_Pillars_Outside. (c) Fountain_&_Vincent_2. (d) Friends_1.

other approaches. Finally, a comprehensive review of all the codecs is performed.

A. Compression of Lenslet Images

For PSNR computed on Y channel (Fig. 8, solid lines), the performance of the two codecs examined here ($P01$ and $P02$) strongly depends on the content, as it is common when computing PSNR. In general, $P01$ outperforms codec $P02$ for high bitrates. For low bitrates, $P02$ outperforms $P01$ for contents *Bikes* and *Fountain_&_Vincent_2*, and is outperformed in the remaining cases. PSNR computed on YUV channels (Fig. 8, dashed lines) shows similar trends.

Codec $P02$ has a particularly poor performance with content *Friends_1*, and in general performs worse than codec $P01$ for high bitrates. Results are particularly surprising since the codec proposed in $P02$ is supposed to improve the performance of HEVC Intra (anchor $P01$) with new prediction schemes. To better investigate the reasons behind this behaviour, we computed PSNR at different stages of the pipeline. Results from PSNR computation are shown in Fig. 9. In particular, we computed PSNR on the 4D LF without any color or gamma correction, on the color-corrected 4D LF, on the gamma-corrected 4D LF and when both corrections were applied on the 4D LF. Additionally, we compute PSNR on the lenslet image prior to the transformation, to better assess the performance of the two codecs on 2D images. The PSNR was computed with respect to the uncompressed reference at the same stage of the pipeline.

Results show that, on the lenslet image and on the 4D LF without any correction, $P02$ always outperforms $P01$. On the

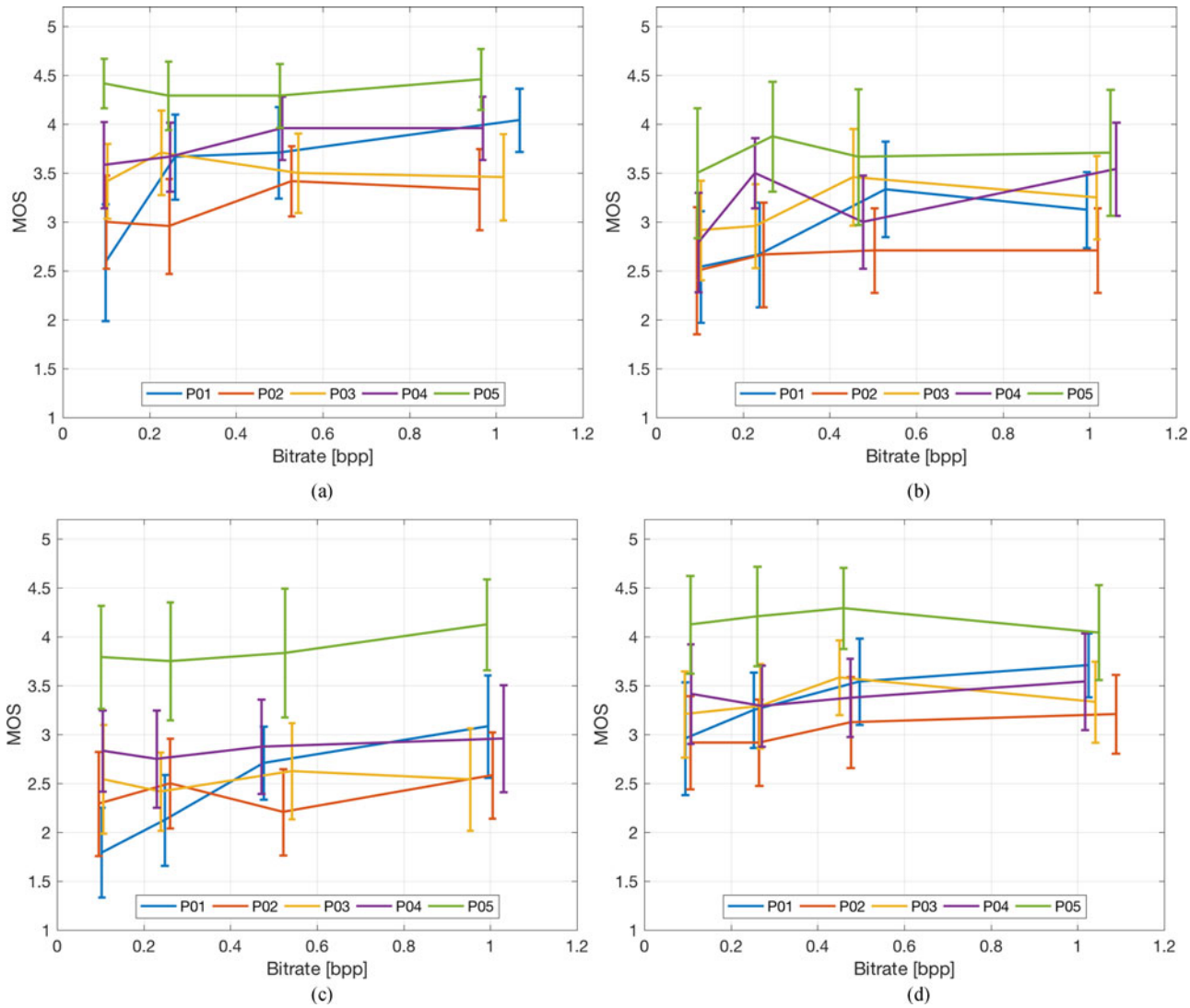

Fig. 10. Results of interactive subjective tests. MOS vs bitrate for all contents, with respective confidence intervals. (a) Bikes. (b) Stone_Pillars_Outside. (c) Fountain_&_Vincent_2. (d) Friends_1.

gamma-corrected 4D LF, $P02$ performs better than $P01$ on half of the contents. When color correction is applied on the 4D LF, however, we see a degradation in performance, with $P01$ outperforming $P02$ for high bitrates. This suggests that the prediction method, while working efficiently on compression of regular images, as proven by the results obtained on the lenslet image prior to transformation, adapts rather poorly to the peculiarities of LF images, and is more susceptible to errors after color correction is applied. Results from PSNR computed on YUV channels follow the same trend.

Results from both interactive and passive subjective evaluations show that $P01$ is performing significantly better than $P02$ for the highest bitrate. In particular, in the interactive test $P01$ is significantly better than $P02$ for all contents, whereas in the passive test it is significantly better for 3 out of 4 contents. For bitrate $= 0.5$ bpp, interactive tests show that $P01$ performs better than $P02$ for only 1 out of 4 contents, whereas passive tests indicate that it outperforms $P02$ on half of the contents. For lower bitrates (0.2 and 0.1 bpp) the difference between the two codecs is negligible.

B. Compression of 4D Light Field

As discussed in Section III, we want to analyse the effect of downsampling the lenslet image prior to transformation to 4D LF. For this reason, we compare the performance of $P04$, which uses a chroma subsampled version of the lenslet image, with $P05$, which creates the 4D LF from the lenslet image which has not been subsampled (see Fig. 3).

For PSNR computed on Y channel (Fig. 8, solid lines), the two codecs have similar performance for all bitrates for contents *Bikes* and *Stones_Pillars_Outside*, whereas for contents *Fountain_&_Vincent_2*, $P05$ performs better than $P04$ for all bitrates. Although downsampling of chroma channels should not affect the Y channel, color correction is applied on RGB channels of the single views, which are then converted to YUV

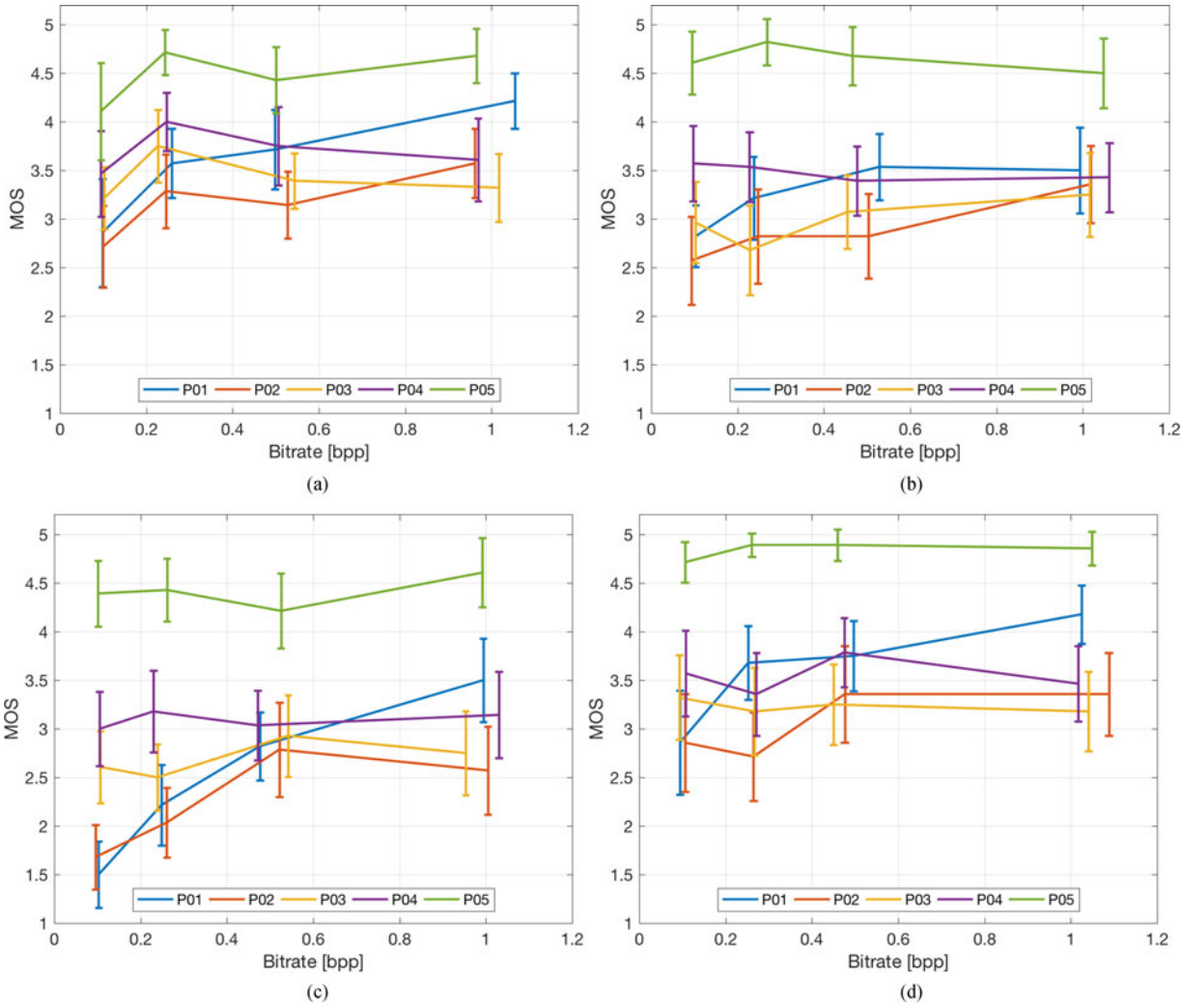

Fig. 11. Results of passive subjective tests. MOS vs bitrate for all contents, with respective confidence intervals. (a) Bikes. (b) Stone_Pillars_Outside. (c) Fountain_&_Vincent_2. (d) Friends_1.

to compute the PSNR. The downsampling thus affects the Y channel as well.

Results are similar for PSNR computed on YUV channels (Fig. 8, dashed lines), although for content *Fountain_&_Vincent_2*, the difference between $P04$ and $P05$ is now negligible.

Results from subjective evaluations and pairwise comparison (Figs. 10–13) show a stronger preference for codec $P05$ when compared to codec $P04$. In particular, results from the interactive tests show that, for the highest bitrate, $P05$ performs significantly better than $P04$ for two out of four contents. For bitrate $= 0.25$ bpp, $P05$ performs better on three out of four contents, whereas for the remaining two bitrates (0.5 bpp, 0.1 bpp) it always performs significantly better than $P04$.On the other hand, results from the passive tests show that $P05$ always performs significantly better than $P04$, for all bitrates.

C. Hybrid Compression of Lenslet Images

As seen in Section III, the third compression scheme $P03$ is compressing 4D LF and then converting them back to lenslet images. It is thus worthy of note to compare the performances of $P01$, $P02$ and $P03$, since they have the same input and output in the compression and decompression steps, although they use different approaches.

From the objective metric point of view, $P01$ and $P03$ outperform codec $P02$ for high bitrates. For low bitrates, $P03$ always outperforms $P01$ and $P02$, although in case of content *Stone_Pillars_Outside*, the difference between the codecs is negligible. For PSNR computed on YUV channels (Fig. 8, dashed lines), codec $P03$ outperforms the others for all contents. Interestingly enough, for codec $P03$ PSNR computed on YUV channels always has higher values than PSNR computed on the Y channel, while for all the other codecs the opposite is true. One possible explanation for this peculiar behavior is that the inverse color and gamma transformation applied before transforming the 4D LF back to lenslet has an effect on the final color performance, leading to better results in the YUV channels.

The subjective evaluation results do not show the same trends as the objective results (Figs. 10–13). In particular, results from the interactive tests show that for the lowest bitrate (0.1 bpp) $P03$ outperforms $P01$ on two out of four contents and never

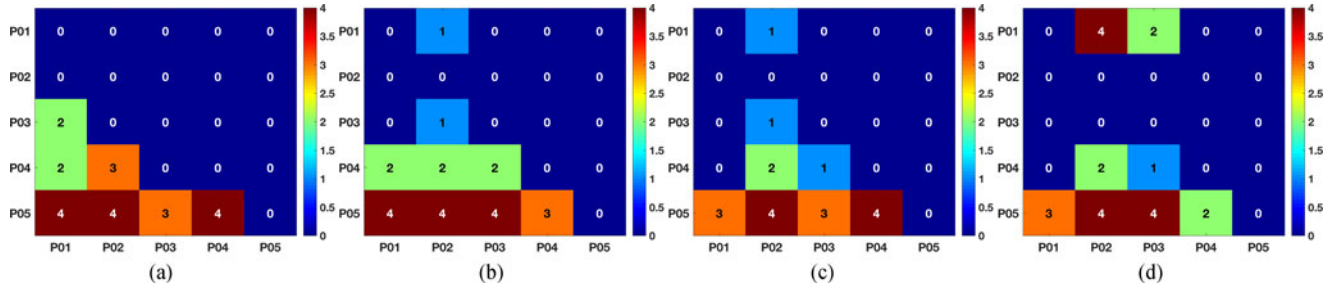

Fig. 12. Pairwise comparison results for interactive subjective tests. Each cell contains the number of contents for which the null hypothesis was rejected, for each compression ratio. The null hypothesis is defined as $MOS_i \leq MOS_j$, in which i indicates the row and j the column of the matrix. (a) R4 (b) R3 (c) R2 (d) R1.

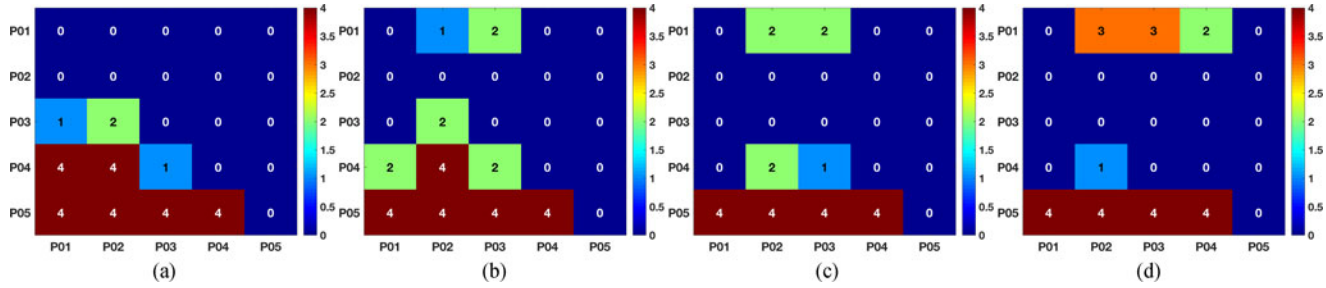

Fig. 13. Pairwise comparison results for passive subjective tests. Each cell contains the number of contents for which the null hypothesis was rejected, for each compression ratio. The null hypothesis is defined as $MOS_i \leq MOS_j$, in which i indicates the row and j the column of the matrix. (a) R4 (b) R3 (c) R2 (d) R1.

outperforms $P02$, whereas the passive tests show that $P03$ performs better than $P01$ on only 1 out of 4 contents, and performs better than $P02$ on 2 out of 4 (Fig. 13(a)). For intermediate bitrates (0.5 bpp and 0.25 bpp), interactive tests show that $P01$ and $P03$ both perform significantly better than $P02$ on one out of four contents, whereas passive tests additionally show that $P01$ performs significantly better than $P03$ on half of the contents for both bitrates. For the highest bitrate, $P01$ performs significantly better than $P03$ on at least half of the contents (3 out of 4 in case of passive tests, 2 out of 4 in case of interactive tests), and outperforms $P02$ in the majority of cases (3 out of 4 in case of passive tests, 4 out of 4 in case of interactive tests).

For the objective evaluation, the hybrid scheme $P03$ performs better than the other lenslet compression schemes. However, results from the subjective evaluation suggest that the difference in performance with respect to $P01$ (simple HEVC Intra) is negligible for low bitrates, and leads to poorer results for the highest bitrates.

Since $P03$ compresses lenslet images through transformation to 4D LF, it is useful to compare its performance to the performance of $P04$. For PSNR computed on Y channel (Fig. 8, solid lines), the performance of codecs $P03$ and $P04$ strongly depends on the content, as expected. For high bitrates, $P04$ performs better than $P03$, with the notable exception of content *Stones_Pillars_Outside*, in which codec $P03$ performs slightly better for all bitrates. For low bitrates, however, $P04$ performs slightly worse than $P03$ for all contents except *Friends_1*, for which $P04$ performs better than $P03$. PSNR computed on YUV channels (Fig. 8, dashed lines) show similar trends.

Results from subjective evaluation and pairwise comparison (Figs. 10–13), however, show that codec $P03$ never performs significantly better than codec $P04$. In particular, for the lowest bitrate, results from the interactive tests indicate that no codec performs significantly better than the other, whereas results from passive tests suggest that $P04$ performs better than $P03$ on 1 out of 4 contents. For bitrate $= 0.25$ bpp and 0.5 both interactive and passive tests agree that $P04$ outperforms $P03$ for 2 out of 4 contents and 1 out of 4 contents, respectively. For the highest bitrate, interactive tests indicate that $P04$ performs significantly better than $P03$ on 1 out of 4 contents, whereas for the passive tests they are statistically equivalent for all contents.

D. General Discussion

In general, both objective and subjective results show that coding 4D LF (point B in Fig. 1) leads to better performance when compared to coding lenslet images directly. In particular, pseudo-temporal ordering of 4D LF, obtained from RGB 444 lenslet image, performs significantly better than the other proposals for at least half of the contents for all bitrates examined in the subjective assessment of quality, showing that chroma subsampling of lenslet images can lead to a considerable reduction in visual quality. It is worth noting, however, that results from passive tests show that $P01$ performs statistically better than $P04$ on 2 out of 4 contents for the highest bitrate (Fig. 13(d)).

Comparison of different lenslet image compression algorithms shows that improvements in performance for 2D image coding do not necessarily result in better visual quality of LF image. In particular, whereas HEVC intra with LLE and SS has

better performance in objective evaluation carried out on lenslet images and 4D LF without color correction, it performs significantly worse when color correction is applied. Further work on lenslet image compression should address the effect of color correction on the final 4D LF, and propose new strategies to appropriately cope with this issue.

Coding 4D LF has the benefit of not requiring any metadata to be correctly displayed. Moreover, it can be used to code 4D LFs acquired with different acquisition technologies. Since the most distorted views in the 4D LF can be discarded in the compression process, it also allows for bitrate saving. As we previously mentioned, the transformation to 4D LF is an additional step that would be not suitable for low-memory devices. Thus, if consumers' market is the desired target, a solution that does not require any transformation would be preferrable. In this case, coding 4D LFs seems the most suitable choice.

The additional step of converting to 4D LF is not an issue if the target is the professional market. However, fidelity to acquisition parameters is of paramount importance. As seen before, chroma subsampling leads to poorer performances, especially after color correction has been applied. On the other hand, coding 4D LF leads to discarding metadata, which could be used in post-processing softwares, as well as potentially discarding some heavily distorted views. In this case, both approaches presented in this paper do not seem suitable. A new approach should be designed, aimed at high fidelity to acquisition parameters.

VI. Conclusion

In this paper, two different coding approaches for light field image compression were defined, described, and evaluated. Objective and subjective quality assessments of five different compression algorithms, following the aforementioned coding approaches, were conducted. Experimental results provide some insights on the impact of compression algorithms within, as well as across predefined use cases, on the perceived quality. This reveals the necessity of further investigations and improvements of compression algorithms especially in terms of processing of the metadata related to light field image data rendering.

More specifically, subjective quality evaluations show that one coding approach, namely, compressing 4D LF, yields significantly better results in terms of visual quality for all bitrates when compared to compressing lenslet images. The 4D LF coding approach is particularly suitable for general consumers' use case, since it does not involve additional computations at the decoder side to be properly rendered. Moreover, the coding approach does not require metadata to be successfully decoded and displayed, thus reducing the bitrate. Finding a successfull approach for the professional market, however, is still an open issue. A new method for compressing lenslet images while taking into account color fidelity must be designed for this type of market. Further research should focus on how to modify the proposed compression algorithm for light field images to further improve the performance and to meet the needs of all use cases.

References

[1] A. Gershun, "The light field," *Stud. Appl. Math.*, vol. 18, no. 1–4, pp. 51–151, 1939.

[2] E. H. Adelson and J. R. Bergen, *The Plenoptic Function and the Elements of Early Vision*. Cambridge, MA, USA: Vision and Modeling Group, Media Laboratory, Massachusetts Inst. Technol., 1991.

[3] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interactive Tech.*, 1996, pp. 31–42.

[4] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Proc. 32nd Picture Coding Symp.*, 2016, pp. 1–5.

[5] I. Viola, M. Řeřábek, and T. Ebrahimi, "A new approach to subjectively assess quality of plenoptic content," in *Proc. SPIE Opt. Eng.+ Appl*. 2016, pp. 99 710X–99 710X.

[6] I. Viola, M. Rerabek, and T. Ebrahimi, "Impact of interactivity on the assessment of quality of experience for light field content," in *Proc. 9th Int. Conf. Quality Multimedia Exp.*, 2017, pp. 1–6.

[7] "The (new) stanford light field archive," Jul. 2016. [Online]. Available: http://lightfield.stanford.edu/papers.html

[8] S. E. Chen, "Quicktime VR: An image-based approach to virtual environment navigation," in *Proc. 22nd Annu. Conf. Comput. Graph. Interactive Tech.*, 1995, pp. 29–38.

[9] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," in *Proc. 28th Annu. Conf. Comput. Graph. Interactive Techn*., 2001, pp. 425–432.

[10] B. Wilburn *et al.*, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.

[11] A. Ouazan, P. T. Kovacs, T. Balogh, and A. Barsi, "Rendering multiview plus depth data on light-field displays," in *Proc. 3DTV Conf.: True Vision-Capture, Transmiss. Display 3D Video*, 2011, pp. 1–4.

[12] F. Zilly, C. Riechert, M. Müller, and P. Kauff, "Generation of multi-view video plus depth content using mixed narrow and wide baseline setup," in *Proc. 3DTV-Conf.: True Vision-Capture, Transmiss. Display 3D Video*, 2012, pp. 1–4.

[13] E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[14] M. Levoy, "Light fields and computational imaging," *IEEE Comput.*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[15] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Report*, vol. 2, no. 11, pp. 1–11, 2005.

[16] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 338–343, Apr. 2000.

[17] A. Jagmohan, A. Sehgal, and N. Ahuja, "Compression of lightfield rendered images using coset codes," in *Proc. Signals, Syst. Comput. Conf. Rec. 37th Asilomar Conf.*, 2003, vol. 1, pp. 830–834.

[18] B. Girod, C.-L. Chang, P. Ramanathan, and X. Zhu, "Light field compression using disparity-compensated lifting," in *Proc. Int. Conf. Multimedia Expo, ICME'03*, 2003, vol. 1, pp. I–373.

[19] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in *Proc. IEEE Workshop Stat. Signal Processing*, 2003, pp. 30–33.

[20] C. Perra, "Lossless plenoptic image compression using adaptive block differential prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 1231–1234.

[21] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient intra prediction scheme for light field image compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 539–543.

[22] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 80–91, Jan. 2016.

[23] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2016, pp. 1–4.

[24] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2016, pp. 1–4.

[25] R. Monteiro *et al.*, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2016, pp. 1–4.

[26] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2016, pp. 1–4.

[27] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2016, pp. 1–4.
[28] R. Olsson, M. Sjostrom, and Y. Xu, "A combined pre-processing and H.264-compression scheme for 3D integral images," in *Proc. IEEE Int. Conf. Image Process*, 2006, pp. 513–516.
[29] C. Choudhury, Y. Tarun, A. Rajwade, and S. Chaudhuri, "Low bit-rate compression of video and light-field data using coded snapshots and learned dictionaries," in *Proc. IEEE 17th Int. Workshop Multimedia Signal Process.*, 2015, pp. 1–6.
[30] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 4733–4737.
[31] P. Helin, P. Astola, B. Rao, and I. Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," in *Proc. 3DTV-Conf.: True Vision-Capture, Transmiss. Display 3D Video*, 2016, pp. 1–4.
[32] ISO/IEC JTC 1/SC29/WG1 JPEG, "Jpeg pleno call for proposals on light field coding," Doc. N73013, Chengdu, China, Oct. 2016.
[33] G. Alves, F. Pereira, and E. A. da Silva, "Light field imaging coding: Performance assessment methodology and standards benchmarking," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2016, pp. 1–6.
[34] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," in *Proc. Int. Conf. Image Process. Theory, Tools Appl.*, 2015, pp. 494–497.
[35] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on refocused and extended focus images," in *Proc. 24th Eur. Signal Process. Conf.*, 2016, pp. 898–902.
[36] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proc. 8th Int. Conf. Quality Multimedia Exp.*, 2016, pp. 1–2.
[37] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1027–1034, Jun. 2013.
[38] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 1–20, Feb. 2015.
[39] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—Including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, 2012.
[40] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R BT.500-13, Jan. 2012.
[41] *General Viewing Conditions for Subjective Assessment of Quality of SDTV and HDTV Television Pictures on Flat Panel Displays*, ITU-R BT.2022, Aug. 2012.
[42] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd—A framework for crowd-based quality evaluation," in *Picture Coding Symp.*, 2012, pp. 245–248.

Irene Viola received the B.Sc. degree in cinema and media engineering and the M.Sc. degree in computer engineering from the Polytechnic University of Turin, Turin, Italy, in 2013 and 2015, respectively. She is currently working toward the Ph.D. degree in the Multimedia Signal Processing Group (MMSPG), under the supervision of Professor Touradj Ebrahimi, at the Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.

Her research interest includes image and video processing, compression, and evaluation, with a focus on light field coding.

Martin Řeřábek received the Ph.D. degree in electrical engineering from the Faculty of Electrical Engineering (FEE), Czech Technical University in Prague, Prague, Czech Republic, in 2013. He is a Postdoctoral Researcher at the Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. He is the author or the coauthor of more than 70 research publications. His main research interest is focused on image and video processing and compression, evaluation and improvement of quality of experience in immersive multimedia technologies, biomedical, astronomical, and aerial image processing.

Touradj Ebrahimi (M'95) is currently a Professor at Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, heading its Multimedia Signal Processing Group. He is also the Convener of JPEG standardization Committee. He was also adjunct Professor with the Center of Quantifiable Quality of Service at Norwegian University of Science and Technology from 2008 to 2012. His research interests include still, moving, and 3-D image processing and coding, visual information security (rights protection, watermarking, authentication, data integrity, steganography), new media, and human computer interfaces (smart vision, brain computer interface). He is the author or the coauthor of more than 200 research publications, and holds 14 patents.