Performance analysis of JPEG Pleno light field coding

Cristian Perra[a], Pekka Astola[b], Eduardo A. B. da Silva[c], Hesam Khanmohammad[d,e], Carla Pagliari[f], Peter Schelkens[d,e], and Ioan Tabus[b]

[a]DIEE, UdR CNIT, University of Cagliari, Italy
[b]Tampere University (TAU), Finland
[c]PEE/COPPE/DEL/POLI/UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[d]Vrije Universiteit Brussels, Belgium
[e]imec, Leuven, Belgium
[f]PGEE/PGED/IME, Instituto Militar de Engenharia, Rio de Janeiro, Brazil

ABSTRACT

Light fields can nowadays be acquired by several methods and devices in the form of light field images, which are at the core of new forms of media technologies. Many research challenges are still open in light field imaging, such as data representation formats, data compression tools, communication protocols, subjective and objective quality of experience measurement metrics and methods. This paper presents a brief overview of the current architecture of the JPEG Pleno light field coding standard under development within the JPEG committee (ISO/IEC JTC1/SC29/WG1). Thereafter, a comparative analysis between the performance of the JPEG Pleno Light Field codec under various modes and configurations and the performance of the considered anchor codecs is reported and discussed.

Keywords: JPEG Pleno. ligth field, compression

1. INTRODUCTION

The amount of light radiated in every direction for every wavelength and every time instance can be represented as a 7D function named plenoptic function. In practice only a sampled representation of the plenoptic function can be acquired using digital imaging devices. A common sampling of the plenoptic function is named light field and is a vector function representing the radiance of a discretized set of light rays.

The problem of light field coding has received considerable interest in the last years as confirmed by academic and industrial research papers and several light field coding challenges that took place in signal processing related conferences such as the IEEE International Conference on Multimedia and Expo (ICMEW, 2016), and the IEEE International Conference on Image Processing (ICIP, 2017).

Several methods proposed in literature for tackling the problem of light field coding apply standard image coding tools, such as JPEG standards from ISO/IEC JTC1/SC29/WG1, or video coding tools based on MPEG standards from ISO/IEC JTC1/SC29/WG11. A preprocessing step is commonly applied to the light field for improving the compression performance of a given standard coding tools. Some example of preprocessing are: representing the light field as a single arrangement of all the subaperture images (i.e. as a multiview image); or

Further author information: (Send correspondence to C.Perra)
C. Perra: E-mail: cperra@ieee.org
P. Astola: E-mail: pekka.astola@tuni.fi
E.A.B. da Silva: E-mail: eduardo@smt.ufrj.br
H. Khanmohammad: E-mail: khesam@etrovub.be
C. Pagliari: E-mail: carla@ime.eb.br
P. Schelkens: E-mail: peter.schelkens@vub.be
I. Tabus: E-mail: ioan.tabus@tuni.fi

representing the light field as a pseudo-video sequence arrangement of the subaperture images ordered in several ways (e.g. raster scan order, spiral order).[1–4]

A lossy to lossless lenslet image compression method has been designed defining two sets of subaperture image: reference view and dependent views. The former are encoded by a standard lossy or lossless compressor, and the latter are reconstructed by sparse prediction from the reference set using the geometrical information from the depth map.[5] This method has been improved with additional elements providing a unitary treatment of plenoptic camera data and high density camera array data and the introduction of hierarchical encoding of the references images, resulting in a more flexible architecture.[6]

Other approaches to light field coding are based on the exploitation of the 4D redundancy in the light field data. The light field is divided into 4D blocks, a 4D Discrete Cosine Transform of each block is computed, the transform coefficients of the 4D block are grouped using hexadeca-trees on a bitplane-by-bitplane basis, and the generated stream is encoded using an adaptive arithmetic coder. This methods has been proved very effective for light fields acquired by lenslet based camera devices.[7]

In 2015, the ISO/IEC SC29 WG1 JPEG committee has started a new standardization activity, named JPEG Pleno,[8] aiming at defining coding tools for novel image modalities such as point cloud, light field, and holography. In particular, the standardization activities of the last years have been devoted to the development of a generic file format and light field image coding tools.

The outline of the paper is as follows. In Section 2 the JPEG Pleno light field architecture is briefly summarized. The evaluation of the performance of the JPEG Pleno light field encoder is reported in Section 3 presenting the test conditions, the experimental evaluation and the discussion of the obtained results. Section 4 concludes the paper.

2. JPEG PLENO LIGHT FIELD ARCHITECTURE

The JPEG Pleno light field coding architecture supports two coding modes. One mode exploits the redundancy in 4D light field data by utilizing a 4D transform technique, the other mode is based on 4D prediction. These modes are briefly summarized in this paper, for a detailed description we refer to.[9]

2.1 4D PREDICTIVE MODE CODEC

In the 4D predictive mode the depth based warping and merging of reference views provide the main prediction stage. Initially, a set of views are selected as reference views with rest of the views labeled as intermediate views, and the texture and depth of the reference views are encoded using a suitable codec selected from the JPEG family of image coding standards. In the 4D prediction scheme,[6] the pixel correspondence information between the reference views and an intermediate view is obtained from the depth maps and camera parameters, and the pixels of each reference view are warped to the intermediate view location followed by the optimal prediction stage where the multiple warped views are merged into a complete view. The design of the optimal prediction is based on the varying degrees of occlusions in the warped reference views. Optionally, an additional prediction stage using sparse filtering is used for final adjustment prior to the (optional) view prediction residual coding. Similarly to the encoding of texture and depth components of the initial reference views, image codecs from the JPEG family are used for coding of the view prediction residual. The encoder works in an hierarchical order where the views are partitioned into disjoint sets with the initial set of reference views corresponding to the lowest hierarchical level. Intermediate views on the higher hierarchical levels are predicted from the views on the lower hierarchical levels. Due to the hierarchical depth-based prediction, the 4D predictive codec is able to efficiently encode light fields obtained with a variety of light field imaging technologies such as plenoptic cameras and high density camera arrays.

2.2 4D TRANSFORM MODE CODEC

The 4D Transform encoder exploits the 4D redundancy of a light field by first partitioning it into variable-size 4D blocks, and transforming each of them using a 4D Discrete Cosine Transform (DCT). The bit-planes of the generated 4D array of transform coefficients are partitioned and encoded using hexadeca-trees. The leaves of an hexadeca-tree correspond to either a 4D coefficient or a 4D block of zero-valued coefficients. The

resulting bitstream is further encoded using an adaptive arithmetic encoder. The partition of the light fields into 4D blocks, as well as the clustering of the bit-planes of the 4D-DCT coefficients using the hexadeca-trees, are jointly determined though Rate-Distortion Lagrangian optimization. A detailed explanation of the 4D Transform encoder can be found in.[9]

The 4D Transform mode encoder does not need depth data, exploiting the light field's 4D redundancy as a whole. The absence of depth data in the encoding pipeline has its pros and cons. If the light field presents sufficient inter-view redundancy, such as the ones from the Lenslet datasets, its performance is very good, as reported in Section 3. In contrast, for sparsely sampled light fields, where there is small inter-view redundancy, 4D transform mode has a significant loss in Rate-Distortion performance. For these cases, the best solution is the one provided by the 4D Prediction mode described in Section 2.1.

A desirable characteristic of the 4D transform mode is its random access functionality. Prior to encoding (4D variable size block partitioning, transform and hexadeca-tree encoding), the light field is divided into fixed-size 4D blocks that are independently encoded. This way, the 4D transform mode provides straightforward random access to these fixed-size 4D blocks, which can be an important feature in a number applications. Details of the random access capabilities of the 4D transform mode are reported in.[10]

3. PERFORMANCE EVALUATION

3.1 Test conditions

In order to fairly evaluate experimental results of the different proposed codecs, the JPEG committee provided a Common Test Conditions (CTC).[11] The light field images selected for this purpose are chosen such to cover the diversity of light field images in terms of acquisition technology, bit depth, spatial resolution, number of views, texture, and scene geometry. The summary of all 9 selected images is represented in Table 1.

Moreover, to compare different codecs, different metrics are used that are elaborated upon in this section. Before computing any metric, RGB images (both original images and reconstructed ones) must be converted to YCbCr color space using ITU-R BT.709-6 recommendations.[12] The first metric used is Peak Signal to Noise Ratio (PSNR). The PSNR for any image is computed by:

$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right), \tag{1}$$

where n is the bit depth and MSE is Mean Square Error computed. MSE for a component of two images ($\mathrm{I_{orig}}$, $\mathrm{I_{rec}}$) is given by:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(I_{\mathrm{orig}}(i,j) - I_{\mathrm{rec}}(i,j) \right)^2. \tag{2}$$

In this equation, $M \times N$ is the spatial resolution of the images. Once $\mathrm{PSNR_Y}$, $\mathrm{PSNR_{Cb}}$, and $\mathrm{PSNR_{Cr}}$ are obtained, $\mathrm{PSNR_{YC_bC_r}}$ can be calculated as follows:

$$PSNR_{\mathrm{YCbCr}} = \left(\frac{6PSNR_{\mathrm{Y}} + PSNR_{\mathrm{Cb}} + PSNR_{\mathrm{Cr}}}{8} \right). \tag{3}$$

The total PSNR of a light field image is the average of the PSNRs of all the views. The CTC procedures ask to report the total $\mathrm{PSNR_{YCbCr}}$ as well as the total $\mathrm{PSNR_Y}$.

The second metric is the Structural Similarity (SSIM) index. The Matlab program for calculating SSIM is provided in CTC document.[11] Similar to total PSNR, the total SSIM is the average of the SSIMs of all the views. However, only the SSIM for the luminance component (Y) is needed.

Finally, the last metric is the Bjøntegaard metric. This metric provides numerical averages for rate-distortion (RD) curves. The Bjøntegaard metric reports how much one RD-curve is superior to the other one in terms of image quality (which is represented by BD-PSNR(dB)) and bitrate (which is given as BD-rate(%)).

Source	Type	Image name	Number of views	spatial resolution (pixels)	Bit depth (bit)
EPFL	Lenslet	Bikes	13×13	625×434	10
		Danger de Mort			
		Stone Pillars Outside			
		Fountain & Vincent 2			
Fraunhofer IIS	HDCA	Set 2 2K Sub	33×11	1920×1080	10
Poznan University of Technology	HDCA	Laboratory 1	31×31	1936×1288	8
New Stanford LF Archive	HDCA	Tarot Cards	17×17	1024×1024	8
Synthetic HCI	HDCA	Greek	9×9	512×512	8
		Sideboard			

Table 1. Summary of selected light field images.

In order to calculate the rate, the proposed metric is bit per pixel(bpp) which is given by:

$$R = \frac{N_{tot\_bits}}{N_{tot\_pixels}}, \tag{4}$$

where N_{tot_bits} is the total number of bits of the compressed codestream of the light field and N_{tot_pixels} is the total number of pixels of the whole light field image. For instance, for the Tarot Cards light field, which has 17×17 views and a spatial resolution of 1024×1024, $N_{tot\_pixels} = 17 \times 17 \times 1024 \times 1024 = 30,3038,464$ pixels. Target bitrates for each image are specified and listed in Table 2.

The CTC document also specifies the Random Access Penalty, which shows how many encoded bits are required to have access to a certain Region of Interest (RoI), which can be for example a subaperture image or a spatial region in a subaperture image, and it can be calculated using the following equation:

$$Random\ Access\ Penalty = \frac{Total\ amount\ of\ encoded\ bits\ required\ to\ access\ an\ RoI}{Total\ amount\ of\ encoded\ bits\ to\ decode\ the\ full\ lightfield}. \tag{5}$$

All the reported results are finally compared to an anchor. The CTC selected H.265/HEVC as an anchor, where the subaperture views are encoded as a pseudo-videosequence. The encoding/decoding pipeline for H.265/HEVC anchor is shown in Figure 1. The details of H.265/HEVC anchor generation are provided in the CTC document.[11]

Light field image	Target Bitrate (bpp)					
All Lenslet images		0.001	0.005	0.02	0.1	0.75
Greek and Sideboard		0.001	0.005	0.02	0.1	0.75
Tarot Cards		0.001	0.005	0.02	0.1	0.75
Set2 2K sub	0.0005	0.001	0.005	0.02	0.1	0.75
Laboratory 1	0.0005	0.001	0.005	0.02	0.1	0.75

Table 2. Target bitrates for the JPEG Pleno light field .

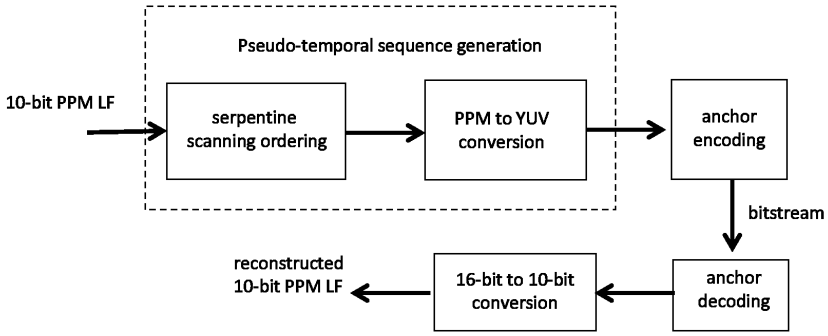


Figure 1. Encoding/decoding pipeline for the H.265/HEVC anchor.

3.2 Experimental evaluation

This section reports the experimental evaluation performed according to the test conditions. As mentioned above, the compression results obtained by compressing the light field test data with H.265/HEVC is used as anchor for assessing the performance of the JPEG Pleno light field coding modes, the 4D predictive coding mode (4D-PC), and the 4D-DCT coding mode (4D-DCT). The obtained bitrates are very close to the target bitrates established in the test conditions. The following plots report: the PSNR computed in dB on the Y channel (PSNR-Y, aka PSNR_{Y}) and in the YCbCr channels (PSNR-YUV, aka $\text{PSNR}_{\text{YCbCr}}$), and the structural similarity index (SSIM) on the Y channel.

As outlined above, the H.265/HEVC (x.265) results have been obtained by following the instructions described in the CTC document[11] and the results (bitrate and metrics) shown are the ones reported in Annex 1 - HEVC Anchor Rate-Distortion Tables and Plots.[11] The 4D Prediction mode and the 4D Transform mode results have been produced using the Verification Model 2.0 (VM 2.0) software.[13]

The HDCA datasets[11] have low inter-view redundancy. Since, as mentioned above, the 4D Transform mode codec is suitable only for very high density light fields, such as the lenslets ones,[11] results for the 4D Transform mode codec will not be shown for the HDCA datasets.

Table 3 shows the parameters used by the 4D Transform mode encoder in its configuration files.

Parameter	Parameter Value	Description
-nv	13	light field number of rows of the view array (t)
-nh	13	light field number of columns of the view array (s)
-off_h	1	initial value of the horizontal view counter
-off_v	1	initial value of the vertical view counter
-v	31	maximum transform size in the spatial vertical direction
-u	25	maximum transform size in the spatial horizontal direction
-t	13	maximum transform size in the views vertical direction
-s	13	maximum transform size in the views horizontal direction
-min_v	4	minimum transform size in the spatial vertical direction
-min_u	4	minimum transform size in the spatial horizontal direction
-min_t	13	minimum transform size in the views vertical direction
-min_s	13	minimum transform size in the views horizontal direction
-lenslet13x13	<no value>	signals to the encoder that is a lenslet dataset
-lambda		value of the RD Lagrangian multipler

Table 3. VM 2.0 4D Transform mode parameters.

The maximum transform sizes used in the spatial horizontal and vertical directions (-u and -v) provide random access capability with minimal penalties to the 4D Transform mode RD performance. In addition, since for the lenslet case the interview dimensions are small (13×13), the large inter-view redundancy makes it natural to use the largest possible inter-view dimensions of the 4D block, which should be then 13×13.

Table 4 shows the parameters used by the 4D Prediction mode encoder in its configuration files. The parameters are defined separately for each view (t, s), and the Table 4 provides the range of values used in the experiments for encoding the JPEG Pleno Dataset.[11] In the 4D Prediction codec the random access capabilities are determined by the hierarchical configuration used, and by using for example $H = 1$ for all views, the codec provides direct access to all views without the need of inter-view prediction. However, for maximal RD performance, the hierarchical configuration is usually defined to contain several hierarchical levels. The 4D Prediction mode offers the user great flexibility in determining the RD performance of the encoded light field, and the RD performance can be tuned, for example, to optimize the image quality only for certain views of interest.

Figures 2 to 5 show the RD results for the datasets Bikes, Danger de Mort, Fountain&Vincent2 and Stone Pillars Outside.[11] All datasets have dimensions $13 \times 13 \times 434 \times 625$ ($t \times s \times v \times u$) and are part of the JPEG Pleno Datasets.[14]

Parameter	Parameter Value	Description
H	$1 \leq H \leq 6$	hierarchical level
N_T	$0 \leq N_T \leq 5$	number of reference views used in texture view prediction
N_D	$0 \leq N_D \leq 5$	number of depth views used in depth view prediction
NNt	$0 \leq NNt \leq 3$	template size for the sparse post-filter
Ms	$0 \leq Ms \leq 25$	order of the sparse post-filter
R_T	$7.3 \times 10^{-5} \leq R_T \leq 5.4 \times 10^{-2}$	rate (bpp) used for reference view coding
R_R	$0.0 \leq R_R \leq 4.4 \times 10^{-2}$	rate (bpp) used for prediction residual coding
R_D	$0.0 \leq R_D \leq 6.3 \times 10^{-3}$	rate (bpp) used for depth coding
$MMODE$	$MMODE \in \{0, 1\}$	view prediction mode
YCC		vertical camera center position
XCC		horizontal camera center position

Table 4. VM 2.0 4D Prediction mode parameters for the view (t, s). The range of values used in the experiments are provided.

The PSNR-Y values of the 4D Transform mode RD curve for the Bikes dataset are very close to the ones of the 4D Prediction mode (left curve of Figure 2). The latter exhibits a better RD performance at higher bitrates, while the former at the lower bitrates. Different RD results are reported when analyzing the PSNR-YUV curves (center curve of Figure 2). The 4D Transform mode shows a better performance for all bitrates, except for very low ones. In contrast, SSIM results show that the 4D Prediction mode codec outperforms the 4D Transform mode (right curve of Figure 2).

Despite being also a lenslet dataset, the Danger de Mort dataset presents different texture and depth characteristics when comparing with the Bikes dataset. The PSNR-Y values obtained by the 4D Transform mode codec outperform the ones produced by the 4D Prediction mode codec in the intermediate rates and in almost rates when comparing the PSNR-YUV values. A similar analysis is valid for the RD results pictured in Figures 4 and 5. In Figure 4, the 4D Prediction mode codec presents a slightly better PSNR-Y RD performance then the 4D Transform mode codec. This result is corroborated by the ones shown in Tables 7 and 8.

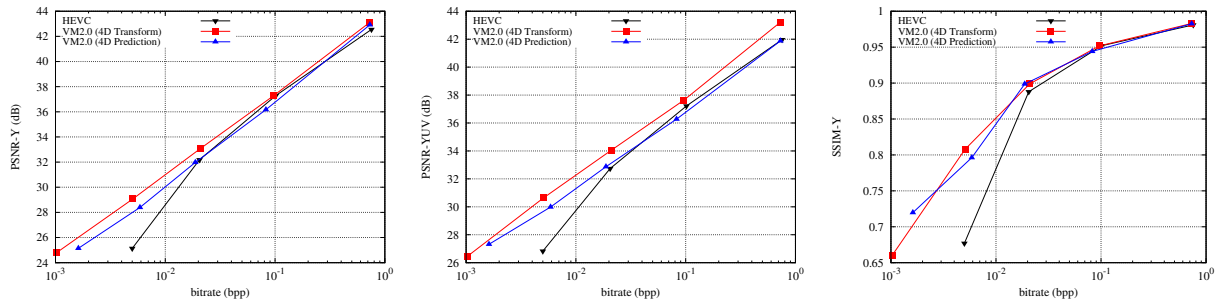


Figure 2. Bikes: Rate-distortion comparison among H.265/HEVC (x.265), JPEG Pleno 4D Predictive coding mode and JPEG Pleno 4D Transform coding mode.

Figures 6 to 10 show the performance of the 4D Prediction mode on the HDCA images.[11] In Figures 6 and 7 the RD performance is evaluated for the synthetic images Greek and Sideboard. In the RD evaluation only estimated depth maps were used. In Figure 6, for the image Greek, the 4D Predictive outperforms or matches HEVC at all rates. For image Sideboard, as seen in Figure 7, HEVC obtains better RD performance at the highest rate point. The images Tarot, Laboratory 1, and Set 2 2K Sub are non-synthetic, and illustrate the performance of the 4D Prediction mode when both camera parameters and depth maps have to be estimated. For all non-synthetic HDCA images, the 4D Prediction has better RD performance at the lower rates, while at the higher rates HEVC obtains better RD performance. Figure 9 illustrates the RD performance when the camera parameter and depth estimation provide moderately good results, while in Figures 8 and 10 the challenges in depth map estimation can be seen to decrease the RD performance of the 4D Prediction mode.

Tables 5 to 8 show the Bjøntegaard delta rate (BD-BR (%)) and the Bjøntegaard delta PSNR (BD-PSNR (dB)

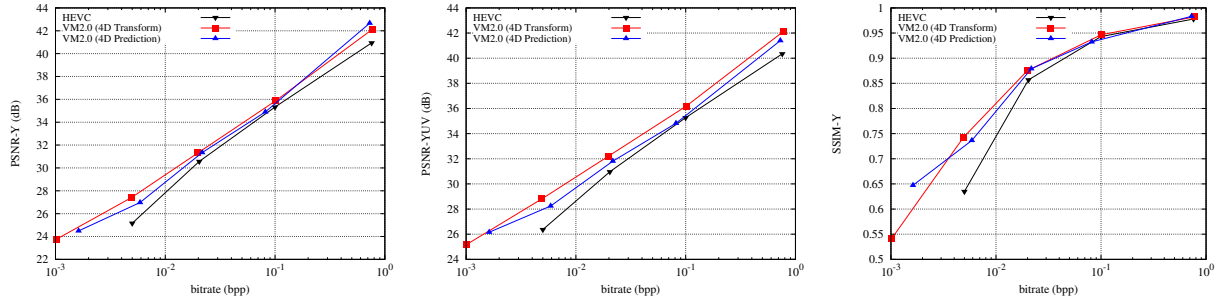


Figure 3. Danger de Mort: Rate-distortion among H.265/HEVC (x.265), JPEG Pleno 4D Predictive coding mode and JPEG Pleno 4D Transform coding mode.

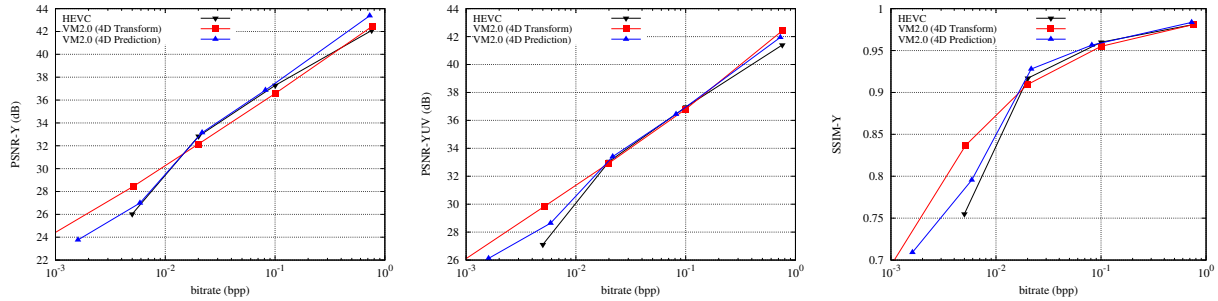
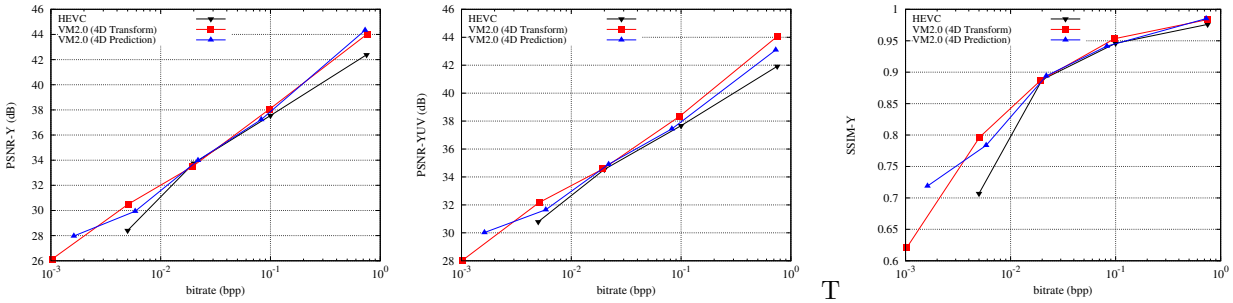


Figure 4. Fountain&Vincent2: Rate-distortion among H.265/HEVC (x.265), JPEG Pleno 4D Predictive coding mode and JPEG Pleno 4D Transform coding mode.



T

Figure 5. Stone Pillars Outside: Rate-distortion among H.265/HEVC (x.265), JPEG Pleno 4D Predictive coding mode and JPEG Pleno 4D Transform coding mode.

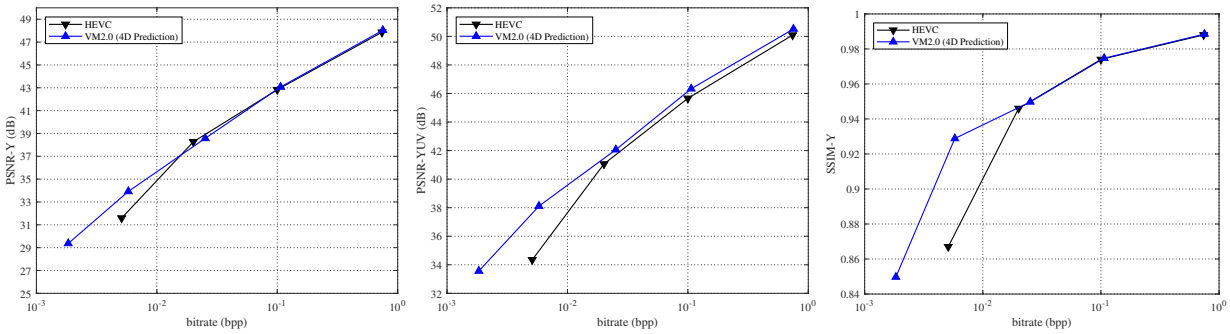


Figure 6. Greek: Rate-distortion among H.265/HEVC (x.265) and JPEG Pleno 4D Predictive coding mode.

regarding the PSNR-Y and PSNR-YUV results obtained for the four lenslets datasets: Bikes, Danger de Mort, Fountain&Vincent2 and Stone Pillars Outside.[11] From the results presented in Table 5, both 4D Prediction and Transform modes outperform the H.265/HEVC (x.265) codec. The 4D Transform mode presents bitrate savings for the same PSNR-YUV values when comparing with the 4D Prediction mode. The former also outperforms the

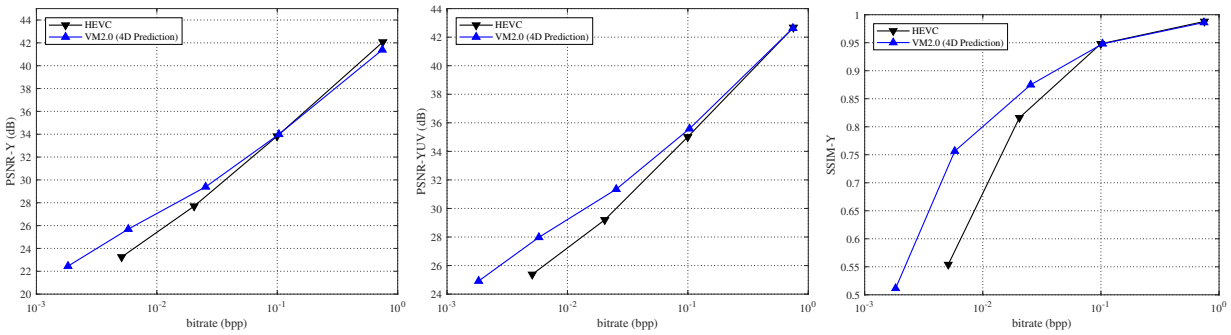


Figure 7. Sideboard: Rate-distortion among H.265/HEVC (x.265) and JPEG Pleno 4D Predictive coding mode.

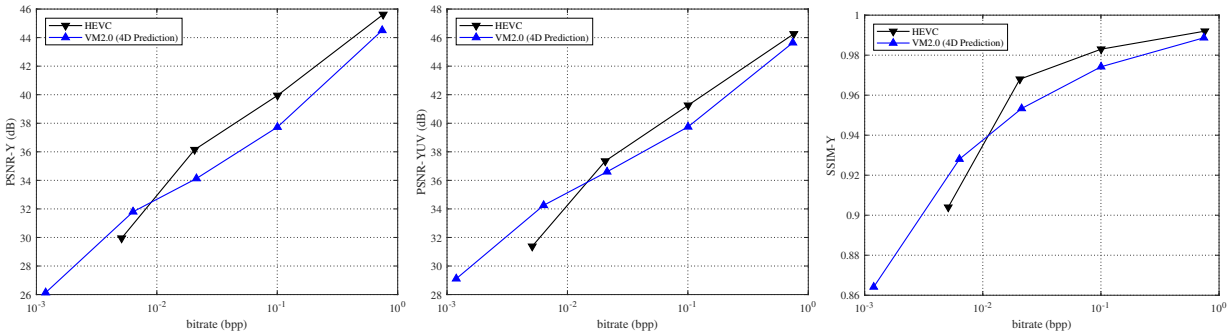


Figure 8. Tarot: Rate-distortion among H.265/HEVC (x.265) and JPEG Pleno 4D Predictive coding mode.

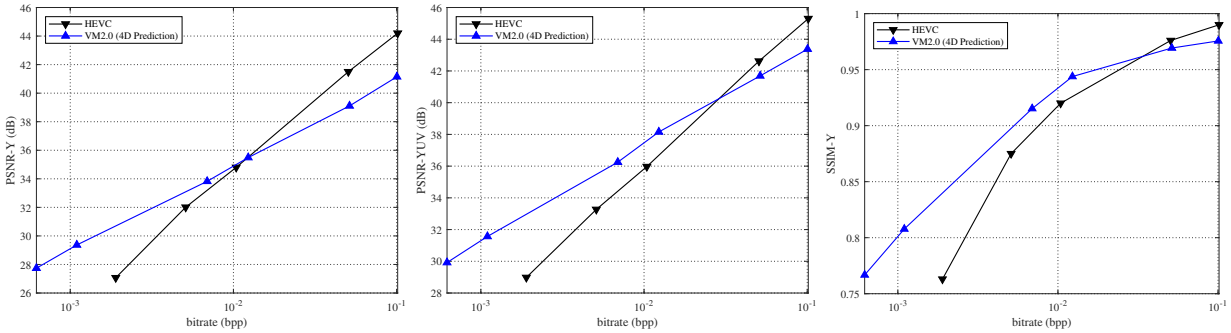


Figure 9. Set 2 2K Sub: Rate-distortion among H.265/HEVC (x.265) and JPEG Pleno 4D Predictive coding mode.

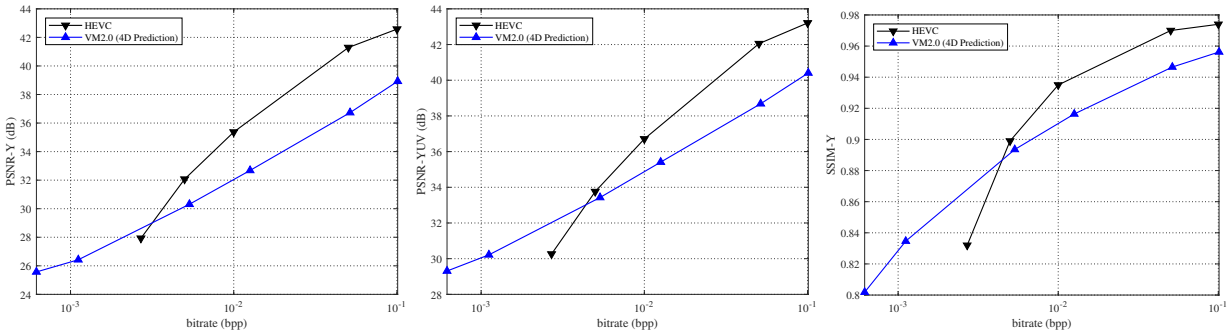


Figure 10. Laboratory 1: Rate-distortion among H.265/HEVC (x.265) and JPEG Pleno 4D Predictive coding mode.

latter when assessing the BD-PSNR(dB) values for the same average PSNR-YUV difference in dB for the same bitrate, as reported in Table 6. When comparing only the luminance results (PSNR-Y), the 4D Transform mode is outperformed by the 4D Prediction mode in both BD-rate and BD-PSNR values for the Fountain&Vincent2 dataset (Table 8).

	Bikes	Danger	Fountain	Pillars
4D Transform mode vs **H.265/HEVC**	-32.15	-36.37	-6.22	-30.24
4D Prediction mode vs **H.265/HEVC**	-0.32	-13.65	-5.86	-13.07
4D Transform mode vs **4D Prediction mode**	-31.47	-21.97	-11.15	-10.51

Table 5. PSNR-YUV Bjøntegaard delta rate (BD-BR (%)): Bitrate savings comparison (reference in **boldface**).

	Bikes	Danger	Fountain	Pillars
4D Transform mode vs **H.265/HEVC**	1.41	1.46	0.54	0.96
4D Prediction mode vs **H.265/HEVC**	0.31	0.62	0.33	0.43
4D Transform mode vs **4D Prediction mode**	0.95	0.69	0.26	0.31

Table 6. PSNR-YUV Bjøntegaard delta PSNR (BD-PSNR (dB)): Average PSNR-YUV difference in dB for the same bitrate (reference in **boldface**).

	Bikes	Danger	Fountain	Pillars
4D Transform mode vs **H.265/HEVC**	-19.02	-24.18	13.04	-20.23
4D Prediction mode vs **H.265/HEVC**	0.80	-14.87	-8.59	-11.39
4D Transform mode vs **4D Prediction mode**	-22.37	-7.06	3.08	-0.83

Table 7. PSNR-Y Bjøntegaard delta rate (BD-BR (%)): Bitrate savings comparison (reference in **boldface**).

Tables 9 to 12 show the Bjøntegaard delta rate (BD-BR (%)) and the Bjøntegaard delta PSNR (BD-PSNR (dB) regarding the PSNR-Y and PSNR-YUV results obtained for the five HDCA images: Greek, Sideboard, Tarot, Set 2 2K Sub, and Laboratory 1.[11] In Tables 9 and 10 the 4D Predictive mode can be observed to outperform HEVC on images Greek, Sideboard, and Set 2 2K Sub, while HEVC outperforms on the images Tarot and Laboratory 1. Both Tarot and Laboratory 1 are challenging images with respect to scene depth estimation causing the quality of the view synthesis results to suffer. In Tables 11 and 12, when assessing the luminance only, the 4D Predictive mode outperforms HEVC for the same three images as with the YUV, however the gain with respect to HEVC being slightly reduced compared to the YUV.

A notable characteristic of the 4D Transform mode codec is the low variation of the quality of the reconstructed views. Figures 11 to 14 depict the values of average, maximum, minimum, and confidence interval of one standard deviation of PSNR-YUV, PSNR-Y and SSIM metrics for the datasets Bikes, Danger de Mort, Fountain&Vincent2 and Stone Pillars Outside, respectively. These plots show the codec's capacity to provide an evenly distributed reconstruction quality of the views.

It is important to note that, due to the lenslet light field camera acquisition process, in all lenslet datasets the four views at the corners are much darker than the rest.[11] This is dealt with in the 4D transform mode by multiplying them by 4 before encoding and dividing them by 4 after decoding. This decreases their coding error by 4, increasing their PSNR and SSIM values. Therefore, in order to show fairer results, these corner views have been removed from the PSNR-Y, PSNR-YUV and SSIM Max computation.

	Bikes	Danger	Fountain	Pillars
4D Transform mode vs **H.265/HEVC**	1.06	1.11	0.11	0.73
4D Prediction mode vs **H.265/HEVC**	0.29	0.77	0.52	0.49
4D Transform mode vs **4D Prediction mode**	0.72	0.24	-0.18	0.07

Table 8. PSNR-Y Bjøntegaard delta PSNR (BD-PSNR (dB)): Average PSNR-Y difference in dB for the same bitrate (reference in **boldface**).

	Greek	Sideboard	Tarot	Set 2 2K Sub	Laboratory 1
4D Prediction mode vs **H.265/HEVC**	-33.47	-29.59	22.37	-32.34	87.02

Table 9. PSNR-YUV Bjøntegaard delta rate (BD-BR (%)): Bitrate savings comparison (reference in **boldface**).

	Greek	Sideboard	Tarot	Set 2 2K Sub	Laboratory 1
4D Prediction mode vs **H.265/HEVC**	0.80	0.91	-0.71	1.41	-0.96

Table 10. PSNR-YUV Bjøntegaard delta PSNR (BD-PSNR (dB)): Average PSNR-YUV difference in dB for the same bitrate (reference in **boldface**).

	Greek	Sideboard	Tarot	Set 2 2K Sub	Laboratory 1
4D Prediction mode vs **H.265/HEVC**	-5.12	-17.06	68.23	-1.28	129.18

Table 11. PSNR-Y Bjøntegaard delta rate (BD-BR (%)): Bitrate savings comparison (reference in **boldface**).

	Greek	Sideboard	Tarot	Set 2 2K Sub	Laboratory 1
4D Prediction mode vs **H.265/HEVC**	0.17	0.46	-1.56	0.32	-2.17

Table 12. PSNR-Y Bjøntegaard delta PSNR (BD-PSNR (dB)): Average PSNR-Y difference in dB for the same bitrate (reference in **boldface**).

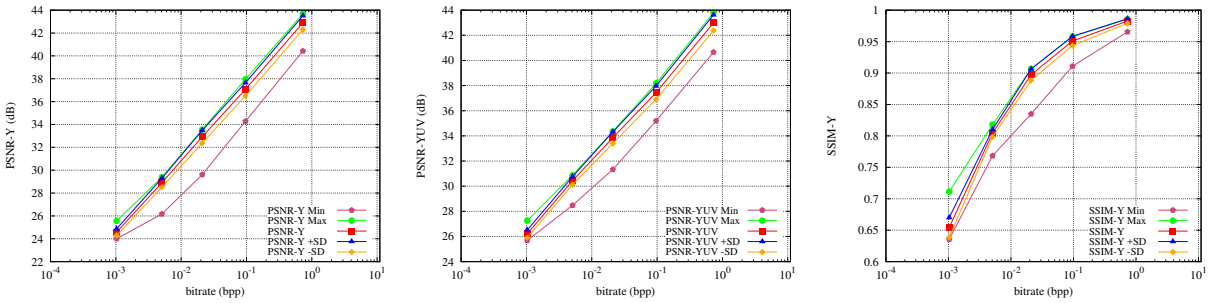


Figure 11. JPEG Pleno 4D Transform coding mode - Bikes: PSNR-Y, PSNR-YUV and SSIM average values, maximum values, minimum values, positive and negative standard deviation values from the respective average values.

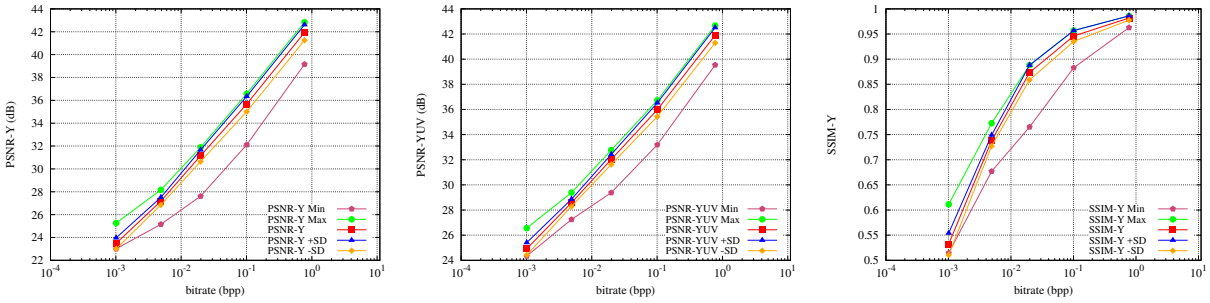


Figure 12. JPEG Pleno 4D Transform coding mode - Danger de Mort: PSNR-Y, PSNR-YUV and SSIM average values, maximum values, minimum values, positive and negative standard deviation values from the respective average values.

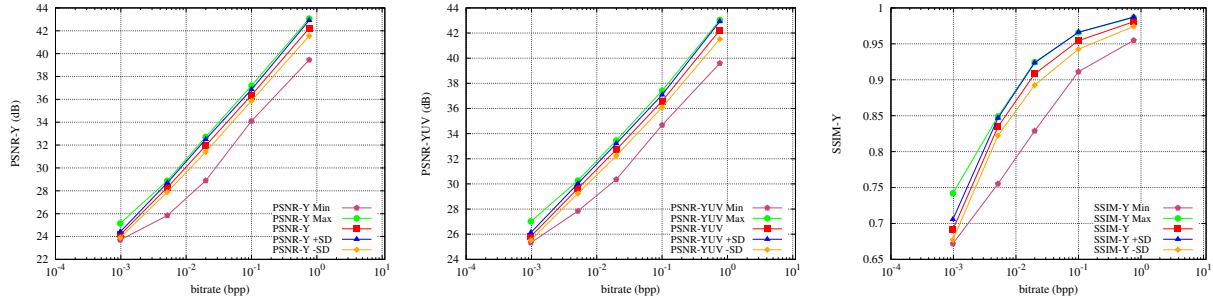


Figure 13. JPEG Pleno 4D Transform coding mode - Fountain&Vincent2: PSNR-Y, PSNR-YUV and SSIM average values, maximum values, minimum values, positive and negative standard deviation values from the respective average values.

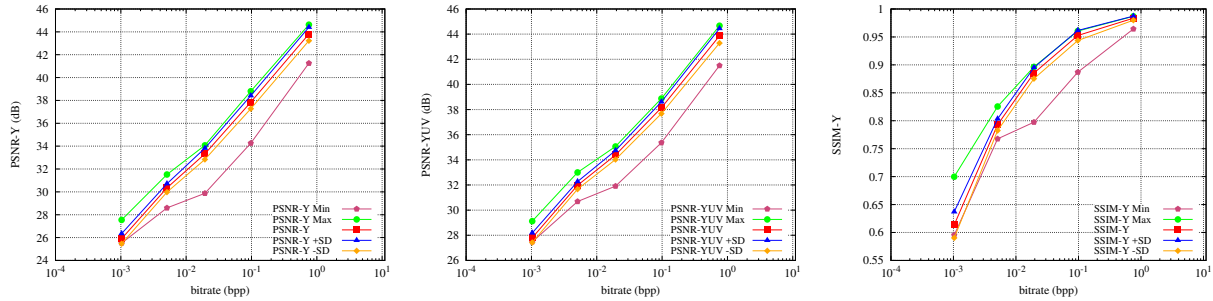


Figure 14. JPEG Pleno 4D Transform coding mode - Stone Pillars Outside PSNR-Y, PSNR-YUV and SSIM average values, maximum values, minimum values, positive and negative standard deviation values from the respective average values.

4. CONCLUSIONS

The JPEG working group is developing novel tools for the compression of light field data. These tools will be part of a new standard named JPEG Pleno Part 2. This paper presented a performance analysis of the current verification model which implements two image coding modes: one based on 4D predictive coding and the other based on 4D transform coding. The analysis was conducted according to common test condition specifications and validated by ISO/IEC SC29 WG1 experts. The obtained results show that both modes outperform state of the art coding tools.

ACKNOWLEDGMENTS

This research activity has been partially funded within the Cagliari2020 project (MIUR, PON04a2 00381) and the DigitArch Cluster Top-Down project (POR FESR, 2014-2020). The authors also would like to thank the Samsung Research Brazil (SRBR).

REFERENCES

[1] Perra, C., "Assessing the quality of experience in viewing rendered decompressed light fields," *Multimedia Tools and Applications* **77**(16), 21771–21790 (2018).

[2] Perra, C. and Assuncao, P., "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in [*2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*], 1–4 (July 2016).

[3] Vieira, A., Duarte, H., Perra, C., Tavora, L., and Assuncao, P., "Data formats for high efficiency coding of lytro-illum light fields," in [*2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*], 494–497 (Nov 2015).

[4] Perra, C. and Giusto, D., "An analysis of HEVC compression for light field image refocusing applications," in [*2018 IEEE Seventh International Conference on Communications and Electronics (ICCE)*], 273–277, IEEE (2018).

[5] Tabus, I., Helin, P., and Astola, P., "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and jpeg 2000," in [*2017 IEEE International Conference on Image Processing (ICIP)*], 4567–4571, IEEE (2017).

[6] Astola, P. and Tabus, I., "Wasp: Hierarchical warping, merging, and sparse prediction for light field image compression," in [*2018 7th European Workshop on Visual Information Processing (EUVIP)*], 1–6, IEEE (2018).

[7] de Carvalho, M. B., Pereira, M. P., Alves, G., B. da Silva, E. A., Pagliari, C. L., Pereira, F., and Testoni, V., "A 4D DCT-based lenslet light field codec," in [*2018 25th IEEE International Conference on Image Processing (ICIP)*], 435–439 (Oct 2018).

[8] Ebrahimi, T., Foessel, S., Pereira, F., and Schelkens, P., "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE MultiMedia* **23**, 14–20 (Oct 2016).

[9] Schelkens, P., Astola, P., da Silva, E. A. B., Pagliari, C., Perra, C., Tabus, I., and Watanabe, O., "JPEG Pleno light field coding technologies," in [*SPIE Optics + Photonics 2019, Applications of Digital Image Processing XLII*], *Proc. SPIE* **11137**, SPIE (2019).

[10] da Silva, E. A. B., de Carvalho, M. B., Pagliari, C. L., Pereira, F., Pereira, M. P., de Oliveira e Alves, G., Testoni, V., and Garcia, P., "ISO/IEC JTC 1/SC29/WG1M80057: Exploration Studies 1.4 for JPEG Pleno, Study on random access extensions to architecture (4D-DCT) 80th JPEG Meeting, Berlin, Germany," (2018).

[11] Pereira, F., Pagliari, C., da Silva, E. A. B., Tabus, I., Amirpour, H., Bernardo, M., and Pinheiro, A., "JPEG Pleno Light Field Coding Common Test Conditions V3.2." Doc. ISO/IEC JTC 1/SC 29/WG1 N83029, 83th JPEG Meeting, Geneva, Switzerland (2019).

[12] ITU-R, "Recommendation ITU-R BT.709-6: Parameter values for the HDTV standards for production and international programme exchange," (2015).

[13] "JPEG Pleno Light Field Verification Model 2.0." https://gitlab.com/wg1/jpeg-pleno-vm.

[14] "JPEG Pleno Light Field Datasets according to common test conditions." https://jpeg.org/plenodb/lf/pleno_lf/.