JPEG PLENO: STANDARDIZING A CODING FRAMEWORK AND TOOLS FOR PLENOPTIC IMAGING MODALITIES

Pekka Astola[1], Luis A. da Silva Cruz[2,3], Eduardo A. B. da Silva[4], Touradj Ebrahimi[5], Pedro Garcia Freitas[6], Antonin Gilles[7], Kwan-Jung Oh[8], Carla Pagliari[9], Fernando Pereira[10,3], Cristian Perra[11], Stuart Perry[12], Antonio M. G. Pinheiro[13,3], Peter Schelkens[14,15], Ismael Seidel[16], Ioan Tabus[17]

[1]Tampere University, Finland, [2]Universidade de Coimbra, Portugal, [3]Instituto Superior Tecnico - Instituto de Telecomunicaç es, Portugal, [4]Universidade Federal do Rio de Janeiro, Brazil, [5]Ecole Polytechnique Fdrale de Lausanne (EPFL), Switzerland, [6]Samsung, Brazil, [7]Institute of Research & Technology b-com, Rennes, France, [8]Electronics and Telecommunications Research Institute (ETRI), Korea, [9]Instituto Militar de Engenharia, Brazil, [10]Instituto Superior Técnico, Universidade de Lisboa, Portugal, [11]University of Cagliari, Italy, [12]University of Technology Sydney, Australia, [13]Universidade da Beira Interior & Instituto de Telecomunicacoes, Portugal, [14]Vrije Universiteit Brussel, Belgium, [15]imec, Leuven, Belgium, [16]Samsung, Brazil, [17]Tampere University, Finland

Abstract – *JPEG Pleno is an upcoming standard from the ISO/IEC JTC 1/SC 29/WG 1 (JPEG) Committee. It aims to provide a standard framework for coding new imaging modalities derived from representations inspired by the plenoptic function. The image modalities addressed by the current standardization activities are light field, holography, and point clouds, where these image modalities describe different sampled representations of the plenoptic function. The applications that may benefit from these emerging image modalities range from supporting varying capture platforms, interactive content viewing, cultural environments exploration and medical imaging to more immersive browsing with novel special effects and more realistic images. These use cases come with a set of requirements addressed by the JPEG Pleno standard. Main requirements envision high compression efficiency, random access, scalability, error-resilience, low complexity, and metadata support. This paper presents a synopsis of the status of the standardization process and provides technical insights as well as the latest performance evaluation results.*

Keywords – Holography, JPEG Pleno, light field, plenoptic, point cloud, standardization.

1. INTRODUCTION

Modern imaging devices can capture a visual scene not only as texture information, as in conventional images, but also as geometric scene information. These devices capture a sampling of the so-called plenoptic function which represents the amount of light in time and in space, theoretically obtained by positioning a pinhole camera at every viewpoint in 3D spatial coordinates, every viewing angle, and every wavelength.

In practice, and depending on the capturing device, a sampling of the plenoptic function can be described by different image modalities. JPEG Pleno particularly supports point cloud, light field, and holography modalities. Some of the current consumer and industrial applications exploiting these image modalities support use cases such as: medical and industrial applications adopting holography for visual inspections or object metrology; gaming and augmented/virtual reality applications that use point clouds as a main modality for scene description; and photography and video creation industries authoring applications that use light fields for computational post-production targeting scene refocusing, segmentation, and immersive experience. Depending on the application, conversion from one image modality to another is essential (Fig. 1).

Moreover, due to the rich description of scenes, often large amounts of data are required for storing, processing and transmission. Current users of plenoptic devices predominantly use proprietary solutions for data description and file structure and for the selection of data compression tools before storage. Therefore, it is evident that there is a lack of common and recognized tools to represent and compress the different plenoptic functions. At the same time there are several already existing applications, as well as novel applications with great opportunities in potential near future markets, that can benefit form these tools.

A common framework for facilitating the acquisition, conversion, processing, and rendering of these image modalities is hence needed for fostering the development of novel applications and for providing interoperability between different systems and technologies (Fig. 1). In this context, the JPEG Committee – formally known as ISO/IEC JTC 1/SC 29/WG 1 – is working on new standards for light field, holography and point cloud compression. This paper reviews the current status of the JPEG Pleno (ISO/IEC 21794) standardization efforts [1] and is organized as follows. The use cases for the JPEG Pleno standard are summarized in Section 2. The requirements that have been established are reported in Section 3. The four

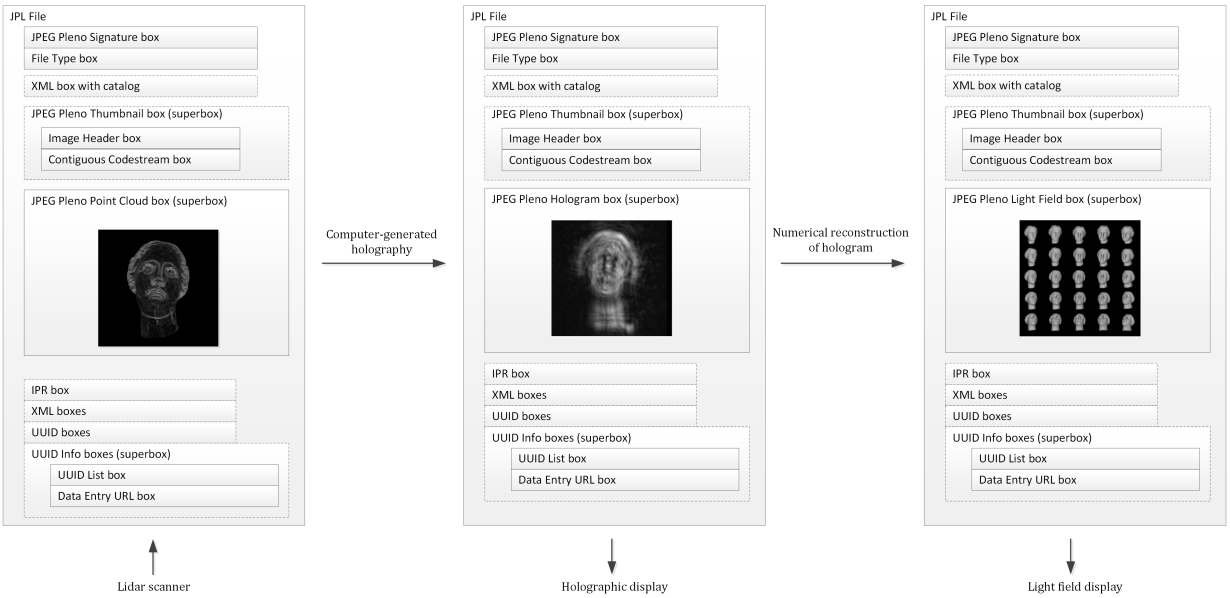

Fig. 1 – Conceptual illustration of the JPEG Pleno workflow and file format. A point cloud is encoded an encapsulated in a JPEG Pleno file. Thereafter it is converted to a hologram deploying a computer-generated holography technique with the original intention to display it on an holographic display. However, an end user does not have a holographic display at its disposal and converts it to a light field for display on a light field display. Note that one of the powers of the JPEG Pleno file format is that it allows for storing simultaneously e.g. the holographic and light field data (not illustrated).

current parts of the JPEG Pleno standard are reviewed in Section 4 and, in addition, future extensions are discussed that are currently being prepared. Finally, Section 5 concludes this paper.

2. USE CASES

The intrinsic 3D nature of plenoptic imaging naturally provides the basis to a large number of applications. In the context of JPEG Pleno standardization, multiple use cases were identified and analyzed to derive a list of appropriate requirements for the representation of the various addressed modalities. These use cases have been discussed in several modality specific "Use Cases and Requirements documents" ' [2, 3, 4]. Examples of use cases are:

- photography and video production,
- virtual, augmented and mixed reality,
- 3D content creation,
- health applications,
- manufacturing and construction,
- retail, robotics,
- surveillance with drones,
- cultural heritage.

For example, in industry, plenoptic representations are deployed in non-destructive testing to build a better understanding of the structure of an object under test. Furthermore, lenslet-based plenoptic cameras capture light fields with a small baseline, but potentially a large number of supported viewing angles. A sufficiently large number of pixels under each microlens in both parallax dimensions results in a depth resolution that allows for accurate surface profiling. Alternatively, holographic interferometry [5] can provide nano-precision surface profiles and even measure surface reflection and polarization features. Finally, 3D profiling of larger scenes can be achieved by deploying lidar technology, producing precise scene descriptions based on point clouds.

A particular use case is the recording of archaeological sites, statues and monuments which provide a detailed register of historical data [6]. Smaller artifacts can be scanned more accurately utilizing these modalities. Even the 3D texture of paintings can be registered allowing for more in-depth analysis of e.g. the painting technique and issued materials.

Plenoptic data also improves the perception of the real world by machines, providing accurate 3D representations of the surrounding environment. Safe navigation of autonomous vehicles is enabled by 3D scanning of the surrounding environment [7].

An exploding market is 3D visualization provided by head-mounted displays or holographic/light field displays. These devices require also compact representations in order to stream the content to these displays and support interactivity. The applications here are ubiquitous. Finally 3D and holographic printing are relevant in the context of the manufacturing of prostheses, and holographic prints.

As shown in Fig. 2, the JPEG Pleno framework provides support for these use cases both at creation or acquisition, or during consumption. Furthermore, a set of tools based on JPEG Pleno associated metadata allows the user to interact at different levels in an E2E chain and maintain even the necessary information during format conversion.

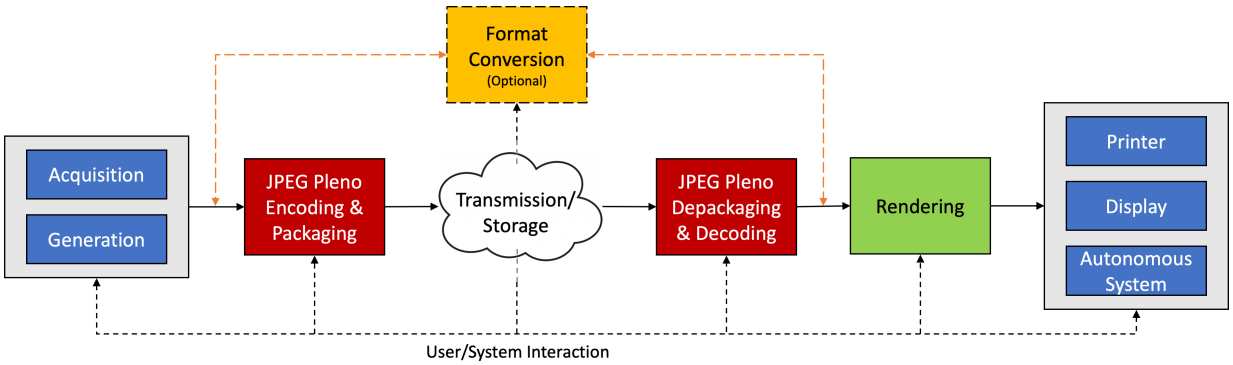

Fig. 2 – JPEG Pleno intends to support the full E2E chain of a plenoptic system. Acquired or generated plenoptic data and associated metadata is being compressed and packaged by the JPEG Pleno encoder and subsequently stored or transmitted. Thereafter, this data is depackaged, decoded and rendered to be reproduced by e.g. a 3D/holographic printer or plenoptic display. Alternatively, the data can be consumed/analyzed by an autonomous (AI) system. Note that the JPEG Pleno framework also facilitates conversion to other modalities throughout the E2E chain.

3. REQUIREMENTS

Following the use cases identified in the previous section, the JPEG Committee has identified the relevant requirements for the JPEG Pleno project. These requirements have been classified either as generic requirements, in the sense that they simultaneously apply to light field, point cloud and holography modalities, or specific requirements in the sense that they are relevant only for one of these modalities. While the mandatory requirements are stated with a 'shall' (SA), the optional requirements are stated with a 'should' (SO).

3.1 Generic requirements

The most recent JPEG Pleno generic requirements are listed in the JPEG Pleno Call for Proposals on Light Field Coding, issued on January 2017 [8]. Many of these generic requirements are common in coding standards, although a few of them are JPEG or plenoptic imaging specific; further details for both of them may be found in [8].

The common coding requirements are:

- high dynamic range, wide colour gamut, colour space, transparency and opacity representation (SA);
- highest possible compression efficiency (SA);
- near-lossless and lossy coding up to high quality (SA);
- lossless coding (SO);
- compression efficiency/functionality tuning (SO);
- random access to subsets of the complete compressed image data with fine granularity (SA);
- quality (SNR), spatial, depth and spectral resolution, number of viewing angles, viewing angle range complexity and content (object) scalability (SO);
- editing and manipulation such as change of depth of field, refocusing, relighting, viewpoint change, navigation, and enhanced analysis of objects (SA);
- error resilience for a large set of networks and storage devices (SO);
- low encoding and decoding complexity in terms of computational complexity; memory complexity and parallelization (SO);
- privacy and security guarantees (SA);
- support for parallel and distributed processing (SO);
- low latency and real-time behaviour (SO);
- support for hierarchical data processing (SO).

The more plenoptic imaging specific requirements are:

- minimal number of representation models for the various plenoptic modalities (SO);
- content description/metadata tools for efficient search, retrieval, filtering and calibration of content from the various imaging modalities for all plenoptic modalities (SA);
- signalling syntax to enable sharing of content between different displays or display elements (SO);
- JPEG ecosystem appropriate degree of forward and backward compatibility (SO).

3.2 Light field specific requirements

Regarding the specific light field requirements, the JPEG Pleno Call for Proposals on Light Field Coding [8] enumerates the following features:

- Representation format – shall support a relevant set of spatial and angular resolutions, multiple colour/spectral components and 6 degrees of freedom (6-DoF); furthermore, should support capture and display dependent/independent light field representation, and a universal light field representation;

- Calibration model in metadata – shall support the signalling of a calibration model in metadata to enable correct content transformation in the imaging pipeline;

- Synchronization of data between sensors – should support signalling syntax to enable synchronization of content captured by different sensors;

- Different (non-)linear capturing configurations – shall support the representation of content produced by various linear/nonlinear configurations and signalling the associated metadata, notably microlens arrays on sensors, sensor arrays of different regular or irregular layouts, rotating sensors/objects;

- Carriage of supplemental depth maps – should support carriage of supplemental depth maps as part of the codestream or file format.

3.3 Point cloud specific requirements

Regarding the specific point cloud requirements, the most recent JPEG Pleno Point Clouds – Use Cases and Requirements document [3] focuses on the following properties:

- Geometry scalability – shall allow for decoding the point cloud geometry using only a part of the full bitstream. This shall provide for point clouds with an increasing number of points, with increasing quality, for a specified bit depth and increasing bit depth as additional layers are decoded;

- Attributes scalability – shall allow for decoding point cloud attributes using only a part of the full bitstream. This shall provide point cloud attributes of increasing quality as additional layers are decoded;

- Random access – shall allow the selective decoding of a portion of the point cloud, allowing for selectively decoding a portion of the point cloud corresponding to a volume in 3D space or a portion of the point cloud visible from a given viewpoint.

3.4 Holography specific requirements

Regarding the specific holography requirements, the most recent JPEG Pleno Holography – Use Cases and Requirements [4] lists the following additional features:

- Hologram types – shall support the following hologram types: amplitude modulation hologram, phase modulation hologram, and complex modulation holograms;

- Representation format – Shall support one or more of the following representation formats for the supported hologram types: amplitude-only, phase-only, amplitude-phase, and real-imaginary;

- Field of view – shall support a large field of view;

- Bit depth – shall support holographic data with bit depth ranging from bi-level to 16-bit integer precision.

In summary, the requirements above are those defining the Call for Proposals and guiding the whole JPEG Pleno specification process.

4. JPEG PLENO

The JPEG Pleno activities are currently at the draft international standard (DIS) phase for both Part 1 of the standard, which addresses the JPEG Pleno framework, and for Part 2, which addresses the light field coding tools. The deadline for Part 1 and 2 to be submitted for the approval stage of the Final Draft International Standard (final draft international standard (FDIS)) is spring 2020. Part 3 and Part 4 activities started in October 2019 and the expected submission deadline for the FDIS approval is October 2020. Fig. 3 summarizes the current JPEG Pleno parts.

4.1 JPEG Pleno Part 1: Framework

This part defines the file format, named JPL (Jpeg PLeno), providing a signalling syntax for storing more detailed information and contextualization for plenoptic content of which codestreams are encapsulated in the file. The committee issued an FDIS ballot text for this part as document wg1n86056 in February 2020 [9]. The JPL format is a box-based media file format, which relates back to the Apple Quicktime format and which belongs to the same family as the ISO/IEC 14496-12 ISO Base Media File Format. The actual specification is based on the same syntax as the box-based file format of JPEG 2000 in JPEG 2000 Part 1 [10] or Part 2 [11]. The box-based file format acts as a binary container composed out of boxes, where each box – named as a superbox – can contain other boxes.

A JPL file will start with a JPEG Pleno Signature box and a File Type box, respectively signalling it concerns a JPEG Pleno family file and versioning/compatibility information. Next, an optional XML box with a catalogue can be included that provides a directory of the included plenoptic content. One of the nice features of a JPL file is that it can contain multiple plenoptic data sets represented in different modalities, i.e. point cloud, light field or holographic data. This allows us to maintain different representations of the same content throughout an application chain and/or to fragment a complex data set into smaller chunks that can be independently encoded. The latter could potentially be interesting in the context of large point clouds, which are hard to handle as a whole due to memory constraints, or which should preferably be handled in smaller units because individual parts have a different semantic interpretation. A particular problem of plenoptic data is that a single codestream can be extremely large compared to a regular image or video content. Identifying the location of a particular codestream in a JPL file might hence result in an excessive file parsing effort. Therefore, including an XML catalog box is a more efficient mechanism that allows for providing pointers that can direct the file parser immediately to the right location in the file to read a particular codestream. Besides the XML catalog box, the JPL file might contain other XML boxes, UUID (info) boxes and IPR (Intellectual Property Rights) boxes, signalling additional information

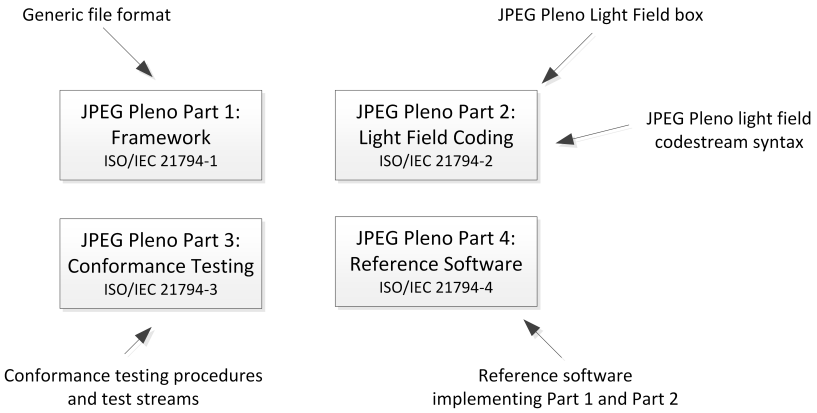

Fig. 3 – JPEG Pleno is currently composed of four parts covering the file format, coding of light fields, conformance testing and reference software [9]

about the plenoptic content, eventually vendor-specific information.

Thereafter, an optional JPEG Pleno Thumbnail box can be included that contains a snapshot image of the complete plenoptic scene covered described by the JPL file. This will allow the end user to get a quick visual understanding of the contained content without having to decode the complete file. This snapshot can be quite advanced and, for example, represent a multispectral image in case of multispectral plenoptic content.

Next, the actual plenoptic content boxes are included. Separate boxes are defined for light field, point cloud or holographic content, respectively the JPEG Pleno Light Field, Point Cloud, and Hologram boxes. Currently, only the formatting of the first has been defined in JPEG Pleno Part 2. The committee has issued an FDIS ballot text as document wg1n87033 in May 2020 [12]. The boxes related to other modalities are intended to be defined in the future.

What is interesting to be mentioned at this point is that JPEG Pleno Part 1 [9] contains also the specification of a global and local reference grid system. Individual plenoptic data sets can be positioned within the global reference grid system at a particular position and with a particular rotation. The local reference grid system is utilized by the codec engine operating on that elementary plenoptic unit. In the earlier example of a point cloud segmented in N individual point clouds, the global coordinates and rotation of each sub-point cloud would be signalled, but the point cloud codecs would operate in different local coordinate systems.

The file format extensions defined in JPEG Pleno Part 2 allow for efficient signalling of the codestream elements of the encoded light field (see Section 4.2), but also of calibration information for the light field. Intrinsically, Part 2 assumes a sub-aperture view based light field representation where each sub-aperture view corresponds with a particular position of a pinhole camera. The codecs assume that these (virtual) cameras are positioned in a planar array in the local coordinate system. Fig. 4 shows the sub-aperture views from the *Bikes* lenslet-based light field. View prediction and synthesis algorithms associated with the codec enable accounting for small offsets of the cameras within the camera array plane. The camera calibration box signals this information in conjunction with an additional translational offset perpendicular to the camera array plane positioning as well as rotational offsets and intrinsic camera parameters – enabling as such more complex camera arrays, where cameras might even be positioned on a spherical surface. However, note that the prediction tools in the codec are only based on the assumption of a planar configuration, albeit some small in-plane, translational calibration offsets can be accounted for.

In summary, the JPL format represents a flexible box-based file format that is extensible such that future supported modalities and functionalities can be efficiently integrated.

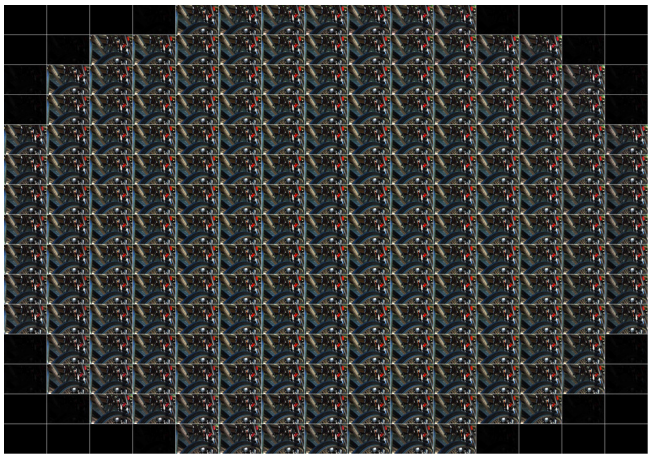

Fig. 4 – Sub-aperture views from *Bikes* lenslet-based light field.

4.2 JPEG Pleno Part 2: Light field coding

An effective plenoptic modality is given by the light fields, that define light rays in space by their (constant) intensity and their intersection with two planes. This is equivalent to representing the plenoptic function as a 2D array of 2D views. In JPEG Pleno light field coding two coding modes are defined: one exploiting the redundancy using a 4D prediction process, the other exploiting the redundancy in 4D light field data by utilizing a 4D transform technique [13]. It is important to note that the two coding modes are independent. The light fields are input to the JPEG Pleno codec as 2D arrays of RGB 2D views [14].

The 4D Transform Mode 4D Transform Mode (4DTM) exploits the 4D redundancy of a light field by first partitioning it into variable-size 4D blocks. Then each block is transformed using a 4D-DCT. The bit planes of the generated 4D array of transform coefficients are first partitioned and encoded using hexadeca-trees followed by an adaptive arithmetic encoder. The partition process and the bit planes clustering can be jointly determined by a Rate-Distortion (R-D) Lagrangian optimization procedure although this an encoder issue and hence not prescribed by the standard. The 4DTM mode also provides random access capabilities.

In the 4D Prediction Mode (4DPM) a subset of views is selected as reference views while the rest of the views are referred to as intermediate views. The texture and depth of the reference views are encoded using the JPEG-2000 standard. The pixel correspondence information between the reference views and an intermediate view is obtained from the depth maps and camera parameters. The pixels of each reference view are warped to the intermediate view location followed by the prediction stage where the multiple warped views are merged into a complete view using least-squares sense optimal predictors over a set of occlusion-based regions. Being depth-based, the 4DPM can efficiently encode light fields obtained with a variety of light field imaging technologies such as those obtained with micro-lens based plenoptic cameras and camera arrays.

The 4DPM can encode light fields very efficiently when reliable depth information is available. On the other hand, the 4DTM does not need depth information for encoding, but is efficient only for encoding light fields with very high angular view density, such as the ones acquired by plenoptic cameras. More details can be found in sections 4.2.2 and 4.2.1 of this paper.

4.2.1 4D-Transform Mode (4DTM)

The parameterization $L(s,t,u,v)$ is a 4D simplification of the plenoptic function that considers the intensity of each light ray constant along its path. Using the two-plane parameterization of light fields [15], a sample (pixel) of the light field is referenced in a 4D coordinate system along the t,s,v and u-axes, where t and s represent the coordinates of the addressed view, and v and u represent the sample (spatial) coordinates within the images (views).

The encoder block diagram of the 4D-Transform Mode (4DTM), introduced in Section 4.2 is pictured in Fig. 5.

The partitioning of the 4D blocks into sub-blocks is signalled with a binary tree using ternary flags indicating whether a block is transformed as is, is split into 4 blocks in the s,t (view) dimensions or is split into 4 blocks in the u,v (spatial) dimensions. Next, a separable 4D-DCT is applied.

The optimized partitioning for each transformed block may be calculated by obtaining, for example, the Lagrangian encoding cost J_N, defined as $J_N = D + \lambda R$, where D is the distortion incurred when representing the original block by its quantized version and R is the necessary rate to encode it. The other possible R-D costs are calculated whenever a 4D block is partitioned in its spatial or view dimensions. For example, the left-hand side of Fig. 6 pictures a $t_k \times s_k \times v_k \times u_k$ 4D-block subdivided into four sub-blocks of sizes $t_k \times s_k \times \lfloor \frac{v_k}{2} \rfloor \times \lfloor \frac{u_k}{2} \rfloor$, $t_k \times s_k \times \lfloor \frac{v_k}{2} \rfloor \times (\lfloor u_k - \frac{u_k}{2} \rfloor)$, $t_k \times s_k \times (\lfloor v_k - \frac{v_k}{2} \rfloor) \times \lfloor \frac{u_k}{2} \rfloor$ and $t_k \times s_k \times (\lfloor v_k - \frac{v_k}{2} \rfloor) \times (\lfloor u_k - \frac{u_k}{2} \rfloor)$ respectively. The optimized partitioning for each sub-block is computed by a recursive procedure and the Lagrangian costs of the four sub-blocks are added to compute the spatial R-D cost J_s. The block can be further partitioned in the view directions, with sub-blocks of sizes defined in the right-hand side of Fig. 6. The optimized partitioning for each sub-block is computed using a recursive procedure and the Lagrangian costs of the four sub-blocks are added to compute the view R-D cost J_v. One should note that if the recursive procedure is expanded to transform into a non-recursive one, it would be equivalent to a bottom-up optimization of the tree.

Fig. 7 shows the hierarchical recursive partitioning. The algorithm keeps track of this tree, returning a *partitionString* flag that represents the optimized tree. When the lowest cost is chosen, the current value of *partitionString* is augmented by appending to it the flag corresponding to the lowest cost chosen (Fig. 7: *transform* Flag, *spatialSplit* Flag or *viewSplit* Flag). The string returned by the recursive call that leads to the minimum cost is also appended to the end of the *partitionString*, returning both the minimum cost J_N, J_s or J_v and the updated *partitionString*.

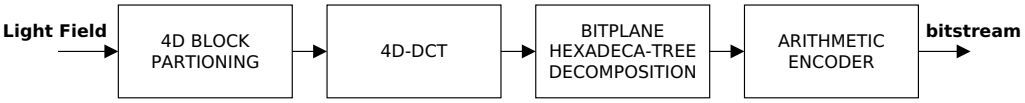

Fig. 5 – 4DTM encoder block diagram.

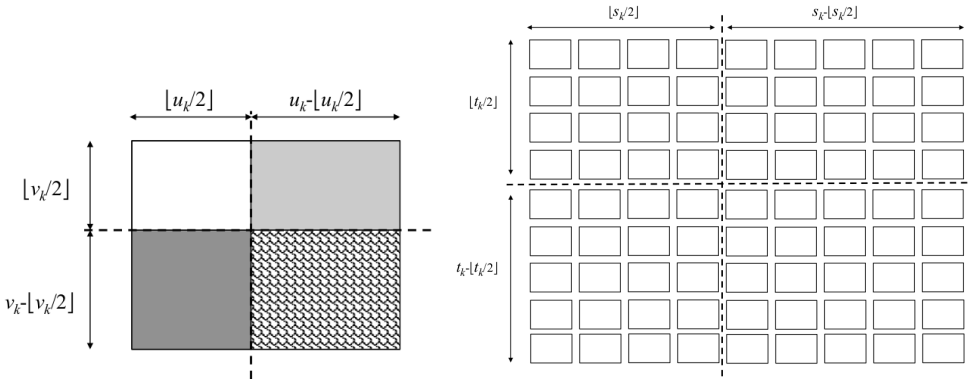

Fig. 6 – 4DTM 4D-block spatial and view partitioning.

The R-D optimized hexadeca-tree structure dictates the quantization and entropy encoding steps. This tree is built by recursively subdividing a 4D block until all sub-blocks reach a 1×1×1×1 4D block-size. Starting from a 4D block of size $t_b \times s_b \times v_b \times u_b$, and a bit plane initially set to a maximum bit plane value three operations are performed represented by a set of ternary flags: *lowerBitplane*, *splitBlock* and *zeroBlock*. The first one lowers the bit plane, where the descendant of the node is another block with the same dimensions as the original one, but represented with precision bitplane-1. Another operation splits the block meaning that a node will have up to 16 children, each one associated to a sub-block with approximately half the length of the original block in all four dimensions. The remaining operation discards the block, indicating that the node has no children and is represented by an all-zeros block, generating three Lagrangian costs J_0, J_1 and J_2, respectively. The optimization procedure is called recursively for each sub-block and the returned Lagrangian costs are added to obtain the new R-D cost and its associated flag: *lowerBitplane*, *splitBlock* or *zeroBlock*.

The 4D coefficients, flags, and probability context information generated during the encoding process are input to the arithmetic encoder, which generates the compressed representation of the light field.

Random access capability is an important feature of the 4DTM. As the light field is divided into fixed-size 4D blocks (e.g. $t_k \times s_k \times v_k \times u_k$) that are independently encoded, random access is provided. Another important feature of the 4DTM is the uniform quality of the reconstructed views. This feature is very important for applications such as refocusing.

4.2.2 4D-Prediction Mode (4DPM)

In the 4DPM, depth information and camera parameterization are used for efficient representation and coding of intermediate views from a set of reference views. As the first step, the 4DPM method takes the light field L, consisting of a set of TS views indexed by $\{(t,s)|t = 0, \dots, T-1, s = 0, \dots, S-1\}$, and partitions it into h_{max} hierarchical subsets, as indicated by the label matrix $H(t,s) \in \{h_0, \dots, h_{max}\}$, where each subset of views corresponds to a particular hierarchical level, as illustrated in Fig. 8. Reference views occupy the lowest hierarchical level $h_0 = 1$, and intermediate views (having $H(t,s) > 1$) are reconstructed based on the views on the lower hierarchical levels [16]. The hierarchical partitioning of views used in 4DPM is therefore similar to the frame referencing structures used in video codecs such as HEVC. However, the efficiency of the hierarchical coding order in 4DPM is based on inter-view redundancies in the angular arrangement of the views in the light field, in a way that can be paralleled to the inter-frame dependencies in video coding.

Reference view, depth and view prediction information is encoded by default with JPEG 2000. However, the JPEG Pleno file format syntax allows also support for a whole series of other codec technologies such as JPEG-1, JPEG LS, JPEG XS, and moreover in principle 4DPM can be combined with any still image encoder. Different hierarchical configurations can be selected depending on the characteristics of the light field data and the desired bit rate. For example, a micro-lens based plenoptic image with high inter-view redundancy can be efficiently coded using a single reference view at h_0, while for wide baseline light fields, such as those obtained with camera arrays, it may be more suitable to select multiple reference views to occupy the lowest hierarchical level h_0. For low

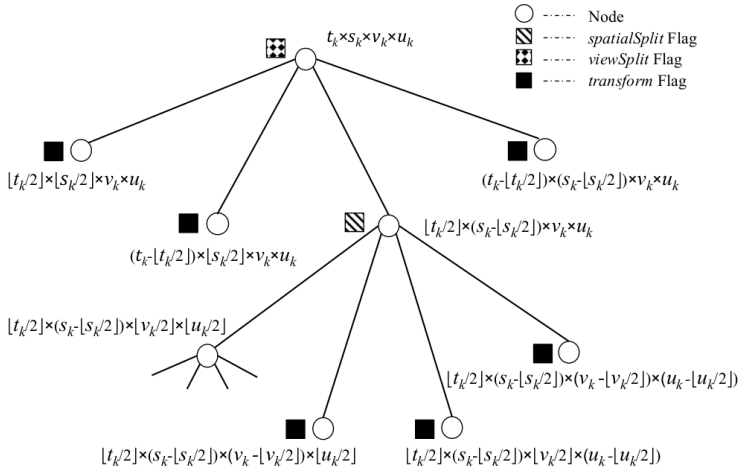

Fig. 7 – 4DTM 4D-block hierarchical spatial and view partitioning.

bit-rate applications, one may choose to use fewer reference views compared to high bit-rate applications.

For a given intermediate view the texture prediction of 4DPM consists of two sparse prediction stages. In the first stage, the depth information is used to warp the pixels of the reference views to their corresponding locations in the camera position of the target intermediate view. Each warped reference view produces a distinct pattern of disoccluded pixels. The aggregation of this information over all warped reference views can be used to infer a useful segmentation of the predicted intermediate view. The segmentation used in 4DPM is generated by the split of the set of pixels into various occlusion classes. Each occlusion class defines a specific subset of warped reference views. These subsets are used to obtain the optimal linear predictor the least-squares sense involving only the relevant warped reference views at each pixel location of the predicted view. The occlusion classes are specially designed to handle the depth-based prediction of wide baseline light fields where disocclusions in warped reference views need to be considered during view merging.

The view merging stage is followed by inpainting, after which the second sparse prediction stage is applied, involving a 2D convolution with a spatially sparse filter, having the template chosen adaptively by a greedy sparse filter design. The intermediate-depth views are also synthesized using a depth-based view warping algorithm, allowing each already processed hierarchical level to act as a source of reference views for subsequent hierarchical levels. Since each hierarchical level contributes a set of reference views for the prediction of higher hierarchical levels, the views on the higher hierarchical levels become more efficiently coded compared to the lower hierarchical levels.

The 4DPM uses sample-based warping and merging operations, which are very efficient when the disparity information is precise enough, differing in this respect from the block-based approaches used in the multi-view extensions of HEVC, namely MV-HEVC and 3D-HEVC. The occlusion classes provide a localized prediction for each of the regions, with each region having an arbitrary contour based on depth features. Therefore, the contours of the regions follow the borders of different depth levels in the scene, and this makes the predictors to be specifically designed for different objects or group of objects. The latter adjustment of the predicted view using a 2D convolution with an adaptively chosen filter template provides an adjustment of the predicted view using global features of the image (such as global illumination conditions).

The entire texture and depth components of any intermediate view are reconstructed using the previously mentioned prediction stages. The view prediction stage of the texture component is optionally followed by coding of the view prediction error. When the application requires low distortion of L, the view prediction error is usually encoded at every $H(t, s) > 1$, while for applications targeting low rates and moderate distortion, the view prediction error may not be encoded at all.

4.2.3 Part 2 Performance Evaluation

The coding tools described in Section 4.2 are currently implemented as a verification model software while design activities for the reference software are ongoing. Different technologies, initially proposed as a response to the call for proposal for the JPEG Pleno standardization, have gone through several exploration studies and core experiments aiming at critically evaluating the coding performance against state-of-the-art encoders and determine actions for the performance and quality improvement of the JPEG Pleno coding framework. Details about the experimental analysis, obtained results, and discussions are reported in several research papers [13], [17].

The JPEG Committee provided a Common Test Conditions (CTC) document [14], as well as a publicly available

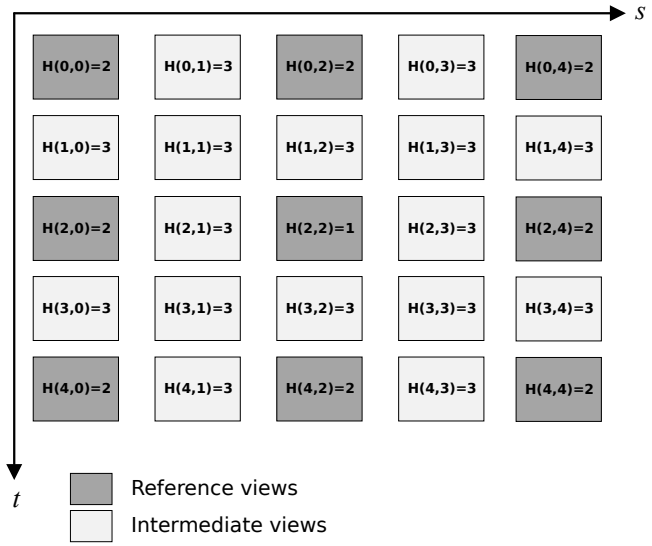

Fig. 8 – 4DPM hierarchical partitioning into reference and intermediate views, with $h_{max} = 3$, when processing the views (t, s) of L for which $H(t, s) = 3$.

JPEG Pleno database [18] to fairly evaluate experimental results of the different proposed codecs. All the experimental comparisons performed in this paper followed the procedures as outlined in the JPEG Pleno Common Test Conditions (CTC) document [14], which also defines all the metrics used for the evaluation (PSNR-Y, PSNR-YUV and SSIM). For completeness, we specify below the equation that was used to calculate the PSNR-YUV, this formula is specified in [14] and [19]:

$$PSNR_{YUV} = (6 \times PSNR_Y + PSNR_U + PSNR_V)/8$$

Also, the precise configuration of the x265 codec is defined in the CTC document. Please note that x265 was used instead of HM reference software since the latter was to slow for the experiments to be completed. For the 4DTM codec only results for the lenslets-based data sets are reported, as its coding design is suitable for very high-density light fields.

Low variation of the quality of the reconstructed views is required in refocusing applications, for example. Fig. 9 depicts the values of average, maximum, minimum, and confidence interval of one standard deviation of PSNR-YUV for the data sets Bikes, Danger de Mort, Fountain&Vincent2 and Stone Pillars Outside, respectively [18]. Due to the lenslet light field camera acquisition process, in all lenslet data sets, the four views at the corners are much darker than the rest. When coding these data sets using the 4DTM the corners views are multiplied by 4 before encoding and divided by 4 after decoding. This procedure decreases the corner views coding error by 4, increasing their metric values. Therefore, in order to show fairer results, these corner views have been removed from the PSNR-YUV Max computation (Fig. 9).

Table 1 shows the Bjøntegaard delta rate (BD-BR (%)) and the Bjøntegaard delta PSNR (BD-PSNR (dB) [20] regarding the PSNR-Y and PSNR-YUV results obtained for

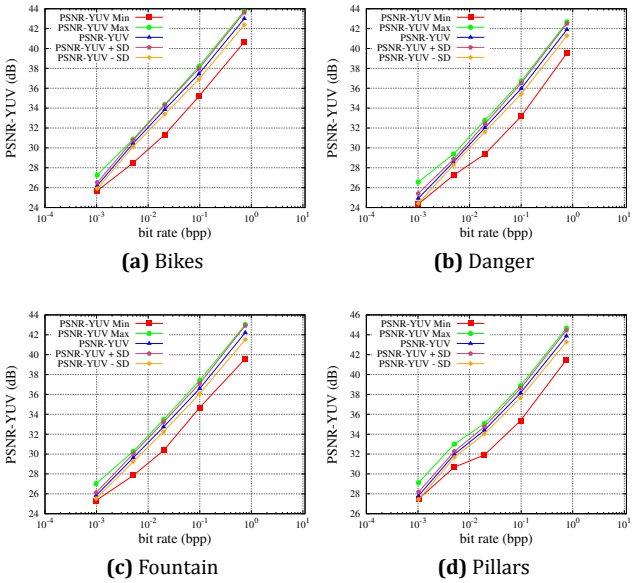

Fig. 9 – 4DTM: PSNR-YUV max, min, average and standard deviation values of the reconstructed views.

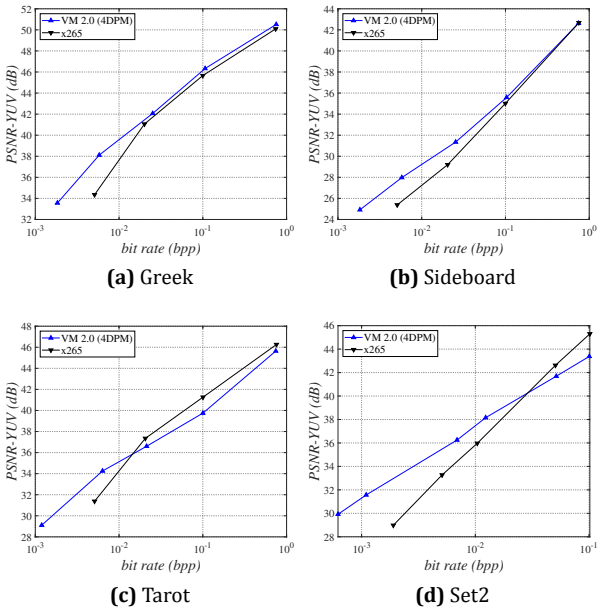

Fig. 10 – 4DPM: PSNR-YUV values of the reconstructed high density camera-array images.

the lenslets data sets: Bikes, Danger de Mort, Fountain&Vincent2 and Stone Pillars Outside [14]. Both coding modes outperform the x265 codec implementation of the H.265/HEVC standard [14] (Table 1). For these very high-density light fields, the 4DTM presents bit-rate savings for the same PSNR-YUV values when comparing with the 4DPM. The former also outperforms the latter when assessing the BD-PSNR(dB) values for the same average PSNR-YUV difference in dB for the same bit rate, as reported in Table 1. When comparing only the luminance results (PSNR-Y), the 4DTM is outperformed by the 4DPM in both BD-rate and BD-PSNR values for the Fountain&Vincent2 data set (Table 1). The 4DTM codec does

not employ any type of prediction structure, nor uses depth maps, which is enough for very high-density light fields. Its design relies on a very simple R-D governed pipeline. These results have been reported in [17].

The performance of the 4DPM, which relies on the hierarchical depth-based view reconstruction scheme, is illustrated in Fig. 10. The PSNR-YUV results are reported for the high-density camera-array (HDCA) images Greek, Sideboard, Tarot, and Set2 [18]. Table 2 shows the Bjøntegaard delta rate (BD-BR (%)) and the Bjøntegaard delta PSNR (BD-PSNR (dB) [20, 19] regarding the PSNR-Y and PSNR-YUV results obtained for the HDCA images. The HDCA images correspond to light fields with wide baselines, offering less redundancy between the views, when compared to the lenslet-based light fields. The 4DPM outperforms x265 especially at the low rates due to the efficient depth-based view reconstruction scheme. The comparisons have been conducted according to the Common Test Conditions specifications and validated by ISO/IEC SC29 WG1 experts.

Table 1 – Bjøntegaard delta rate (BD-BR (%)) and BD-PSNR (dB): Bitrate savings and average PSNR difference comparison (reference in **boldface**).

PSNR-YUV BD-BR (%)				
	Bikes	Danger	Fountain	Pillars
4DTM vs **x265**	-32.15	-36.37	-6.22	-30.24
4DPM vs **x265**	-0.32	-13.65	-5.86	-13.07
4DTM vs **4DPM**	-31.47	-21.97	-11.15	-10.51
PSNR-Y BD-BR (%)				
	Bikes	Danger	Fountain	Pillars
4DTM vs **x265**	-19.02	-24.18	-13.04	-20.23
4DPM vs **x265**	0.80	-14.87	-8.59	-11.39
4DTM vs **4DPM**	-22.37	-7.06	3.08	-0.83
PSNR-YUV BD-PSNR (dB)				
	Bikes	Danger	Fountain	Pillars
4DTM vs **x265**	1.41	1.46	0.54	0.96
4DPM vs **x265**	0.31	0.62	0.33	0.43
4DTM vs **4DPM**	0.95	0.69	0.26	0.31
PSNR-Y BD-PSNR (dB)				
	Bikes	Danger	Fountain	Pillars
4DTM vs **x265**	1.06	1.11	0.11	0.73
4DPM vs **x265**	0.29	0.77	0.52	0.49
4DTM vs **4DPM**	0.72	0.24	-0.18	0.07

4.3 JPEG Pleno Part 3: Conformance testing

JPEG Committee has established JPEG Pleno Part 3 for describing the conformance testing methodology used to ensure that an application complies with the standard. It provides a framework to enable conformity testing and includes, for instance, a description of expected functionality, quality, and speed performance testing. Since the JPEG Pleno standard includes a considerable bulk of features, which can be used for enabling diverse applications, a series of conformance tests are being defined.

Table 2 – Bjøntegaard delta rate (BD-BR (%)) and BD-PSNR (dB): Bitrate savings and average PSNR difference comparison (reference in **boldface**).

PSNR-YUV BD-BR (%)				
	Greek	Sideboard	Tarot	Set2
4DPM vs **x265**	-33.47	-29.59	22.37	-32.34
PSNR-Y BD-BR (%)				
	Greek	Sideboard	Tarot	Set2
4DPM vs **x265**	-5.12	-17.06	68.23	-1.28
PSNR-YUV BD-PSNR (dB)				
	Greek	Sideboard	Tarot	Set2
4DPM vs **x265**	0.80	0.91	-0.71	1.41
PSNR-Y BD-PSNR (dB)				
	Greek	Sideboard	Tarot	Set2
4DPM vs **x265**	0.17	0.46	-1.56	0.32

Two groups of conformance tests are currently available. These groups retain a subset of features from JPEG Pleno Part 2 standard, which depends on the encoding mode. The first group of tests corresponds to the 4DTM tool. The second group of tests addresses the 4DPM tool.

During the decoder conformance testing, a procedure is executed to evaluate the conformance according to the conformance class the decoder belongs to, given a specific encoding mode. conformance classes define the constraints to which a given JPEG Pleno implementation is subjected.

The procedures are designed based on abstract test suites (ATS) specifications, which define general tests for components of parts 1 and 2 of the JPEG Pleno standard. Each ATS defines test purposes, methods, and references of the portion of the document that is being tested. The practical embodiment of the ATS are the executable test suites (ETS). Commonly ATS are embodied into one ETS, which consists of codestreams, reference decoded images, a textual description of the contents of the codestream and tolerance values for MSE and peak error.

4.4 JPEG Pleno Part 4: Reference software

For the purpose of testing JPEG Pleno encoder and decoder implementation and to help those implementing adequately understand the algorithms and methods of Parts 1 and 2, the JPEG Committee provides a reference software in JPEG Pleno Part 4 (ISO/IEC 21794-4). The source code of this reference software is informative only. It is entitled Jpeg Pleno Model (JPLM) and has been developed to ensure a fast proliferation of the JPEG Pleno standard and to be utilized in the conformance testing procedure described in Section 4.3.

The JPLM was designed with a focus on maintainability, integrability, and extensibility rather than encoding and decoding speeds. These non-functional requirements were prioritized because the JPEG Pleno is a complex mul-

tiparty framework and, therefore, the JPLM design should favour successful bug-fix actions, allow the addition of new capabilities or functionality, and enable the aggregation of subsystems (in case of creation of new features via future amendments, for instance). In other words, JPLM architecture has been designed having in mind future support for point cloud and holography extensions.

JPLM complies with JPEG's recommended practices for software development and maintenance.This compliance requires a permissive software license and, therefore, JPLM uses a BSD 3-clause license. Moreover, JPLM implements unit tests extensively. These tests ensure that implemented functionalities remain during software development. These tests also form part of the development pipeline that adopts continuous integration (CI) tools.

4.5 JPEG Pleno point cloud

4.5.1 Current status

Support for point cloud coding will form a future part of the JPEG Pleno standard and the JPEG Committee is currently working towards this goal. As a necessary first step in this direction, the JPEG Committee is currently working on determining the best objective and subjective assessment methodologies for this activity. This will support the later evaluation of point cloud coding solutions to be added to the JPEG Pleno standard.

The JPEG Committee plans to compare different competing coding solutions against anchor point cloud codecs used with well-defined parameter settings to reference encoded contents at a variety of quality levels. To be consistent the state of the art for point cloud coding, the V-PCC and G-PCC codecs in use by the MPEG Committee have been selected as anchor codecs [21].

Point cloud quality assessment is essential to selecting and evaluating potential submissions to form part of a new standard in this area. Point positions may be arranged on a regular grid (for example on a voxel grid) or irregularly arranged in 3D space. In addition to geometrical information, attributes such as colour, surface normal, texture and bump maps or scientific data such as temperature may be associated with individual points or the point cloud as a whole. The manner in which the point cloud is rendered, including whether the points are rendered as individual points or are connected by a mesh and how the mesh is formed can also greatly change subjective judgments of quality [22]. Point cloud data is 3D in nature, however current 3D display systems such as 3D monitors or Head-Mounted Displays (HMDs) often have lower resolution and narrower colour gamut than high-level 2D monitors typically used for subjective quality assessment. This further complicates subjective assessment of quality.

4.5.2 Objective quality assessment

The JPEG Committee measures the performance of proponent codecs against the anchors using a variety of objective measures including bit rate, and for random access, the ratio of bits decoded to access a region of interest versus bits required to decode the entire point cloud. However, here we will expand on the measures of geometric fidelity employed by the JPEG Committee. The measures described here have been chosen to be consistent with the MPEG Committee, in line with the desire of the JPEG Committee to be consistent with other efforts in this field. These are as follows:

Average Point to Point Distance (D1): This measure starts with the identification of corresponding points on a reference and processed point cloud by a nearest neighbour algorithm. The Euclidean distance between the corresponding points on the reference and processed point clouds is then calculated. The set of distances computed in this way for the entire processed point cloud are then arithmetically averaged as the quality measure [21].

Average Point to Plane Distance (D2): For this measure, a nearest neighbour algorithm identifies corresponding points between the reference and processed point cloud. A tangent plane is then fitted to the neighbourhood of the point on the reference cloud taking into account any normal information available and the normal of the plane is used to compute the smallest distance between the point on the processed cloud and the plane fitted to the reference cloud. The set of distances computed for the entire processed point cloud are then arithmetically averaged as the quality measure [21].

Average Plane to Plane Similarity: This metric is based on the angular similarity of tangent planes of corresponding points between the reference and processed point clouds. A nearest neighbourhood algorithm identifies corresponding points in the reference and processed point clouds and tangent planes are fitted to the local areas of the corresponding points on both the reference and processed point clouds. The minimum angle between the two planes is then computed and used to compute an error value between the tangent planes [23] for every point on the processed point cloud. The mean square of the error values is then computed and the logarithm of a ratio of a constant and this mean square value is used as the error measure (akin to a peak signal to noise ratio).

Point to Point Attribute Measure: This measure starts with the identification of corresponding points on a reference and processed point cloud by a nearest neighbour algorithm in line with the *D1* metric. For colour attributes, the MSE for each of the three colour components between the corresponding point is calculated in YCbCr space. The set of colour distances computed in this way for the entire processed point cloud are then arithmetically averaged as a final PSNR quality measure [21].

4.5.3 Subjective testing protocols

Despite many years of effort, current point cloud objective quality metrics do not correlate well with subjective judgments [24]. The best judge of point cloud visual quality continues to be human visual appraisal. The JPEG Committee has adapted an existing subjective testing protocol [24] for the evaluation of point cloud quality. This protocol involves the following steps:

- A view path is determined for each reference point cloud in the form of a smooth path of movement around the point cloud that allows the observer to view the point cloud from a variety of viewpoints appropriate for the content.
- For each point cloud, a 2D video is created where the frames are the projections of the point cloud onto a virtual camera image plane as it moves along the complete view path. Each video lasts for a duration of 24 seconds at a rate of at least 30 frames per second.
- The videos of the reference and the processed point cloud are presented side-by-side, synchronized, on a 2D monitor. Subjects are asked to judge the degree of impairment of the processed point cloud relative to the reference point cloud on a 5 point degradation/impairment scale according to the established Double Stimulus Impairment Scale (DSIS) system outlined in Recommendation ITU-R BT.500.13 [25].
- Collected data has outliers removed and is averaged into mean ppinion scores according to Recommendation ITU-R BT.500.13 [25].

The point cloud activity issued a Call for Evidence on Point Cloud Coding in early 2020 and next steps for this activity are to evaluate submissions to the call and progress to a Call for Proposals for static point cloud coding solutions that may form a new part in the JPEG Pleno standard.

4.6 JPEG Pleno Holography

4.6.1 Current status

Holograms deliver a realistic three-dimensional viewing perception with no vergence-accommodation conflict (VAC), since they record and reproduce 3D scenes by representing both the amplitude and phase of light. Recently, digital holography has received considerable attention and has become popular in digital microscopy and interferometry. Moreover, holographic displays and printing are attracting increasingly more interest. However, since the quality of digital holograms depends on their pixel pitch and resolution [26], they contain a massive amount of data. Hence, efficient compression is necessary to realize holographic imaging services. JPEG Pleno Holography is the first international standardization activity targeting digital hologram compression.

The JPEG Pleno Holography initiative has to solve a number of problems that need to be addressed before standardization can effectively be launched.

First, several test data have been collected for experiments. These holograms were optically captured via interferometry or synthesized via Computer-Generated Holography (CGH) techniques to cover microscopic (holographic microscopy and tomography) and macroscopic applications (display and printing). Several hologram databases [27] are currently available via the JPEG website.

Second, holograms are typically represented as complex-valued data. Hence, codecs have to account for this characteristic and real & imaginary or amplitude & phase representations need to be supported.

Third, since no advanced holographic display exists currently, alternative quality assessment procedures need to be designed to evaluate the decoded holograms [28].

To address these issues, the JPEG Committee has been conducting exploration studies aiming at (1) defining objective and subjective quality evaluation procedures for holography and (2) designing a numerical reconstruction software to be used as a tool for hologram reconstruction. Currently, the common test conditions (CTC) are being defined and a call for proposals (CfP) is scheduled for 2020.

4.6.2 Numerical reconstruction software

Before discussing the test procedures, this section addresses the numerical reconstruction software that was designed to facilitate testing. Indeed, to assess the performance of holographic data coding solutions, it is necessary to compare the visual quality and/or signal-to-noise ratios of scene images numerically reconstructed from the decoded and original reference holograms.

To this end, a Numerical Reconstruction Software for Holography (NRSH) has been developed in Matlab, enabling the reconstruction of different types of holograms, including complex and real-valued, monochromatic and colourful holograms. Utilizing one unique well-validated software environment for this process guarantees that no unexpected distortions are introduced, which might be the case when using proprietary software.

Fig. 11 shows the reconstruction of a complex-valued hologram from two different viewing angles using NRSH software. Depending on the type of input hologram, three methods can be used to simulate the propagation of light waves: the Angular Spectrum, Fresnel and Fraunhofer diffraction formulas. Furthermore, to reconstruct the hologram from different viewing angles, it is possible to simulate a synthetic pupil aperture in the observer's

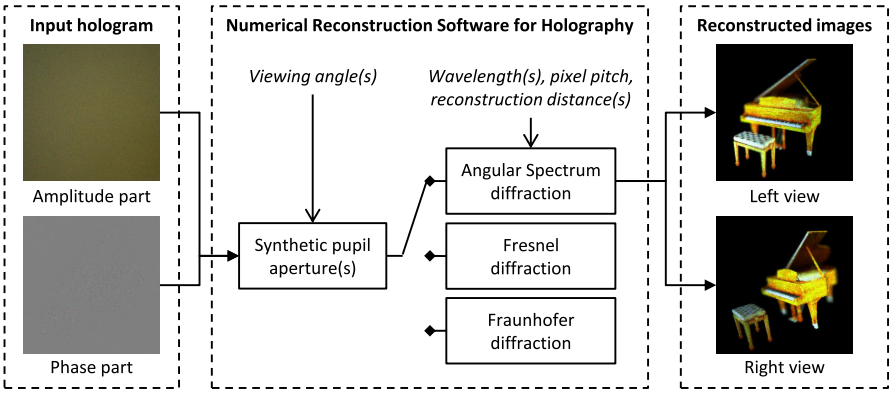

Fig. 11 – Reconstruction of complex-valued hologram *Piano8K* from two different viewing angles using the Numerical Reconstruction Software for Holography (NRSH).

plane, using either a pixel-based or angle-based declaration. The latter allows for declaring not only square but also rectangular apertures.

NRSH supports two operating modes: batch mode and a Graphical User Interface (GUI) mode. The software accounts for the main reconstruction parameters such as wavelength, pixel-pitch, reconstruction distance as well as the dimension and position of the synthetic pupil aperture.

4.6.3 Objective quality assessment

Three main strategies are foreseen for quality testing. A first approach measures the quality of the holograms in the hologram plane. This is the easiest approach and basic measurements made are based on mean squared error (MSE) and peak signal-to-noise ratio (PSNR). These measurements report the global quality of the hologram but do not return any information on how measured errors translate into distortions in the reconstructed holograms.

Therefore, to evaluate the reconstruction quality of the decoded holograms, in a second measurement strategy, the hologram is back-propagated to the object plane such that visually consumable reconstruction of the hologram is being produced. Note that the NRSH software is deployed in this context. Please also note that some coding solutions operate in the object plane as well, particularly in the context of encoding shallow scenes.

The NRSH software will not reconstruct solely one viewing position but typically multiple, focusing the reconstruction on the foreground, middle and background of the scene in addition to multiple viewing angles (left, centre, right, top, bottom). This allows for assessing the impact of the codecs under test at different reconstruction points. If the codecs prioritize e.g. lower frequencies as classical codecs do, depth information will be removed and reconstructions at larger viewing angles will be poor.

Besides classical MSE, PSNR and SSIM metrics, other metrics can be deployed as well in this domain. A wide set of objective metrics is currently in the process of being evaluated. However, it is important to realize that numerical reconstruction gives rise to speckle noise. This noise is typically jeopardizing the correct operation of more advanced quality metrics. A solution to handle this problem is to deploy denoising filters, but yet again these are not solely impacting the speckle noise but also intrinsic relevant information about the scene [29, 30].

Finally, a third approach will deploy application-specific tools that extract for example refractive index information or phase profiles from microscopic holograms of respectively, biological samples or interferometric measurements. It will be studied how these measurements are impacted by the codecs under test.

4.6.4 Subjective testing protocols

As mentioned earlier, the availability of holographic displays is still poor. Moreover, testing on such (experimental) displays is time-consuming and intensive for test subjects. However, recent research has demonstrated that light field or regular 2D displays can be used instead and still provide sufficient insight in the reconstruction quality of compressed holographic content [31].

The strategy followed for 2D display is that a few numerical reconstructions of the hologram, at different depths and angles, are compared for both the reference and compressed hologram in a Double Stimulus Impairment Scale (DSIS) [25] side-by-side comparison test.

For light field displays, a reconstruction of the hologram is provided for every supported viewing angle, converting the hologram into a light field for a particular depth of focus. The test subject is subsequently consuming in an autostereoscopic fashion a light field reconstruction of the hologram. Also here the evaluation procedure is based on a Double Stimulus Impairment Scale (DSIS) [25], though the reconstructed light fields of the reference and compressed holograms are displayed sequentially to the test subject.

Therefore, in a first subjective testing campaign when evaluating proposed technologies, JPEG follows these two subjective evaluation strategies.

5. CONCLUSIONS AND FUTURE PLANS

This paper has presented the current status of the upcoming International Standard for light field, holography and point-cloud compression, namely ISO/IEC 21794 JPEG Pleno. The parts currently under final standardization phases are the ones addressing the container file format (Part 1) and the compression of light field data (Part 2). The conformance testing (Part 3) and the reference software (Part 4) parts are under development. Ongoing activities on JPEG Pleno for point cloud compression have identified a set of requirements that appears not being addressed by existing point cloud coding solutions and that are currently under investigation through core experiments. The JPEG Committee is currently completing the exploration activities on holography coding, defining the common test conditions for evaluating holography compression tools, and preparing the call for proposal for compression of holographic data.

ACKNOWLEDGEMENTS

This research activity has been partially funded within the Cagliari2020 project (MIUR, PON04a2 00381), the French government through the National Research Agency (ANR) Investment referenced ANR-A0-AIRT-07, the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-00001, Development of Holo-TV Core Technologies for Hologram Media Services), Samsung R & D Institute Brazil (SRBR), Portuguese projects FCT UID/EEA/50008/2020, PTDC/EEI-PRO/2849/ 2014 - POCI-01-0145-FEDER-016693, SAICT-45-2017-POCI-01-0145-FEDER-031527 and PTDC/EEI-COM/31527/2017 and European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement Nr. 617779 (INTERFERE).

REFERENCES

[1] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[2] "JPEG Pleno Scope, Use Cases and Requirements," Doc. ISO/IEC JTC 1/SC 29/WG1 N74020, 74th JPEG Meeting, Geneva, Switzerland, 2017. [Online]. Available: https://jpeg.org/jpegpleno/documentation.html

[3] "JPEG Pleno Point cloud use cases and requirements," Doc. ISO/IEC JTC 1/SC 29/WG1 N86012, 86th JPEG Meeting, Sydney, Australia, 2020. [Online]. Available: https://jpeg.org/jpegpleno/documentation.html

[4] "JPEG Pleno Holography use cases and requirements," Doc. ISO/IEC JTC 1/SC 29/WG1 N86016, 86th JPEG Meeting, Sydney, Australia, 2020. [Online]. Available: https://jpeg.org/jpegpleno/documentation.html

[5] C. M. Vest, *Holographic interferometry*. John Wiley and Sons, New York, USA, 1979.

[6] M. Dellepiane, F. Ponchio, M. Corsini, R. Scopigno, M. Callieri, G. Ranzuglia, and P. Cignoni, "3D models for cultural heritage: Beyond plain visualization," *Computer*, vol. 44, pp. 48–55, 2011.

[7] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, p. 55–81, 2015.

[8] "JPEG Pleno Call for Proposals on Light Field Coding," Doc. ISO/IEC JTC 1/SC 29/WG1 N74014, 74th JPEG Meeting, Geneva, Switzerland., 2017. [Online]. Available: https://jpeg.org/jpegpleno/documentation.html

[9] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Part 1 - ISO/IEC FDIS 21794-1," Jan. 2020, WG1N86056, 86th JPEG Meeting, Sydney, Australia.

[10] "Information technology — JPEG 2000 image coding system – Part 1: Core coding system," Recommendation ITU-T T.800 |International Standard ISO/IEC 15444-1, 2016.

[11] "Information technology — JPEG 2000 image coding system – Part 2: Extensions," Recommendation ITU-T T.801 |International Standard ISO/IEC 15444-2, 2004.

[12] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Part 2 - ISO/IEC FDIS 21794-2," April 2020, WG1N87033, 87th JPEG Meeting, Erlangen (online), Germany.

[13] P. Schelkens, P. Astola, E. A. B. da Silva, C. Pagliari, C. Perra, I. Tabus, and O. Watanabe, "JPEG Pleno light field coding technologies," in *Applications of Digital Image Processing XLII*, A. G. Tescher and T. Ebrahimi, Eds., vol. 11137, International Society for Optics and Photonics. SPIE, 2019, pp. 391 – 401. [Online]. Available: https://doi.org/10.1117/12.2532049

[14] F. Pereira, C. Pagliari, E. A. B. da Silva, I. Tabus, H. Amirpour, M. Bernardo, and A. Pinheiro, "JPEG Pleno Light Field Coding Common Test Conditions V3.3," Doc. ISO/IEC JTC 1/SC 29/WG1 N84049, 84th JPEG Meeting, Brussels, Belgium, 2019. [Online]. Available: https://jpeg.org/jpegpleno/documentation.html

[15] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[16] P. Astola and I. Tabus, "WaSP: Hierarchical Warping, Merging, and Sparse Prediction for Light Field Image Compression," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, Nov 2018, pp. 1–6.

[17] C. Perra, P. Astola, E. A. B. da Silva, H. Khanmohammad, C. Pagliari, P. Schelkens, and I. Tabus, "Performance analysis of JPEG Pleno light field coding," in *Applications of Digital Image Processing XLII*, A. G. Tescher and T. Ebrahimi, Eds., vol. 11137, International Society for Optics and Photonics. SPIE, 2019, pp. 402 – 413. [Online]. Available: https://doi.org/10.1117/12.2528391

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

[19] K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström, G.J. Sullivan, and A. Touparis, "Summary information on BD-rate experiment evaluation practices," doc. JVET-R2016, Alpbach (online), Austria. [Online]. Available: http://phenix.int-evry.fr/jvet/doc_end_user/current_document.php?id=10162

[20] G. Bjøntegaard, "Calculation of average psnr differences between rd-curves," *ITU-T VCEG-M33*, 2001. [Online]. Available: https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/

[21] "Common Test Conditions for PCC," Doc. ISO/IEC JTC1/SC29/WG11/n18665, 127th MPEG Meeting, Gothenburg, Sweden., 2019.

[22] E. Alexiou, M. Bernardo, L. da Silva Cruz, L. Dmitrovic, C. Duarte, E. Dumic, T. Ebrahimi, D. Matkovic, M. Pereira, A. Pinheiro, and A. Skodras, "Point cloud subjective evaluation methodology based on 2d rendering," in *International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, Italy*. IEEE, 2018.

[23] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *IEEE International Conference on Multimedia and Expo (ICME), San Diego, USA*. IEEE, 2018.

[24] L. da Silva Cruz, E. Dumic, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany*. IEEE, 2019.

[25] "Methodology for the subjective assessment of the quality of television pictures," Recommendation ITU-R BT.500-13, 2012.

[26] D. Blinder, A. Ahar, S. Bettens, T. Birnbaum, A. Symeonidou, H. Ottevaere, C. Schretter, and P. Schelkens, "Signal processing challenges for digital holographic video display systems," *Signal Processing: Image Communication*, vol. 70, pp. 114–130, Feb. 2019.

[27] "JPEG Pleno Database on Holography," https://jpeg.org/plenodb/.

[28] P. Schelkens, T. Ebrahimi, A. Gilles, P. Gioia, K.-J. Oh, F. Pereira, C. Perra, and A. M. Pinheiro, "JPEG Pleno: Providing representation interoperability for holographic applications and devices," *ETRI Journal*, vol. 41, no. 1, pp. 93–108, 2019.

[29] E. Fonseca, P. T. Fiadeiro, M. V. Bernardo, A. Pinheiro, and M. Pereira, "Assessment of speckle denoising filters for digital holography using subjective and objective evaluation models," *Applied Optics*, vol. 58, no. 34, pp. G282–G292, Dec. 2019.

[30] T. Birnbaum, T. Birnbaum, A. Ahar, A. Ahar, S. Montrésor, P. Picart, P. Picart, C. Schretter, C. Schretter, P. Schelkens, and P. Schelkens, "Speckle Denoising of Computer-Generated Macroscopic Holograms," in *Digital Holography and Three-Dimensional Imaging 2019 (2019), paper W3A.1*. Optical Society of America, May 2019, p. W3A.1.

[31] A. Ahar, M. Chlipala, T. Birnbaum, W. Zaperty, A. Symeonidou, T. Kozacki, M. Kujawinska, and P. Schelkens, "Suitability Analysis of Holographic vs Light Field and 2d Displays for Subjective Quality Assessment of Fourier Holograms," *arXiv:1909.12594 [eess]*, Sep. 2019, arXiv: 1909.12594.