ISO/IEC JTC 1/SC 29/WG 1
(ITU-T SG16)

Coding of Still Pictures

JBIG	JPEG
Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE: JPEG Pleno Call for Proposals on Light Field Coding

SOURCE: WG1

PROJECT: JPEG Pleno

STATUS: Second version (2.0)

REQUESTED ACTION: Provide feedback and response

DISTRIBUTION: Public

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

JPEG Pleno Call for Proposals on Light Field Coding

This document contains a Call for Proposals (CfP) issued in the context of the JPEG Pleno standardization activity, a new work item initiated by JPEG Committee[1] which aims at developing a next generation image coding standard that moves beyond coding of 2D flat content by taking advantage of plenoptic representations.

This call addresses in particular the following components of the JPEG Pleno framework:

- Coding technologies for content produced by lenslet light field cameras;
- Coding technologies for content produced by high-density arrays of cameras;
- System-level solutions associated with light field coding and processing technologies that have a normative impact;

Additionally, contributions are encouraged in the form of:

- Use cases and requirements not yet identified or refinements of already identified requirements for the final version of the call;
- Representative datasets for potential applications identified (or new if not already identified) with conditions allowing usage for standardization purposes as well as organization of special sessions and grand challenges in scientific events;
- Rendering solutions for light fields to serve evaluation purposes;
- Subjective evaluation methodologies and test-bed implementations that can be used to assess various requirements identified (or new requirements if not already identified);
- Objective evaluation methodologies and test-bed implementations that can be used to assess the various requirements identified (or new requirements if not already identified).

This document is structured in two parts. It starts with the rationale behind this new work item, followed by listing the content modalities currently under consideration in JPEG Pleno. In particular, the new features offered by JPEG Pleno in addition to those offered in past JPEG standards are discussed and examples of potential applications that can benefit from this new standard are presented.

JPEG Pleno standardization will proceed in a step by step approach with well defined milestones and deliverables. Each subsequent step serves to enhance and enrich the JPEG Pleno standard by offering solutions for additional modalities, coding tools or system components. In a first phase, static light field coding technologies and associated system level components are called for.

[1] ISO/IEC JTC 1/SC 29/WG 1 and ITU-T SG16

1. JPEG Pleno Framework

1.1. Rationale

Tremendous progress has been achieved in the way consumers and professionals capture, store, deliver, display and process visual content. We have been witnessing an ever-growing acceleration in creation and usage of images in all sectors, applications, products and services. This widespread and ever growing use of images has brought new challenges for which solutions should be found. Among many, one can mention image annotation, search and management, imaging security and privacy, efficient image storage, seamless image communication, new imaging modalities and enhanced imaging experiences. These challenges are just examples to which the scientific community, industry, service providers and entrepreneurs have responded in the past.

During the past 25 years, the Joint Photographic Experts Group (JPEG) has been an example of such efforts, and it has offered image coding standards which can cope with some of the above challenges. This work has resulted in a series of successful and widely adopted coding algorithms and file formats leading to the JPEG, JPEG 2000, and more recently, the JPEG XR, JPEG XT, JPEG Systems and JPEG XS families of image coding standards.

Digital photography markets have known a steady and exponential evolution over the last decade as it concerns supported resolutions, which was mainly driven by Moore's law. However, we have reached the era of nano-electronics and simultaneously we are observing the maturing of micro- and nano-photonic technologies which are giving rise to an unprecedented and heterogeneous range of new digital imaging devices. HDR and 3D image sensors, burst-mode cameras, light-field sensing devices, holographic microscopes and advanced MEMS (e.g. DMD and SLM) devices enable new capture and visualization perspectives that are driving a paradigm shift in the consumption of digital photographic content: we are moving from a planar, 2D world, towards imaging in volumetric and contextually aware modalities. This paradigm shift has the potential to be as disruptive for the photographic markets as the migration from analogue film to digital pictures in the 1990's.

Emerging sensors and cameras will allow for the capture of new and rich forms of data, along with the dimensions of space (e.g. depth), time (including time-lapse), angle and/or wavelength (e.g. multispectral/multichannel imaging). Among these richer forms of data, one can cite omnidirectional, depth enhanced, point cloud, light field and holographic data.

1.2. Plenoptic modalities

JPEG Pleno will be able to cope with various modalities of plenoptic content under a single framework and in a seamless manner. The currently identified modalities include:

Omnidirectional imaging evolves around content which is generated by a single camera or multiple cameras, enabling a wider field-of-view and larger viewing angles of the surroundings. It's often captured as a 360° panorama or complete sphere and mapped to a mono or stereo 2D image. However, partial and truncated spherical and cylindrical configurations have also been referred to as "omnidirectional" in some cases. Efficient projection solution(s) to map captured content on a sphere to an image for further compression and its (their) signalling are among open questions which require standardization.

Depth-enhanced imaging provides many new forms of interactivity with images. It is currently implemented in various types of file formats. Most capture devices today come with their own software solutions that process and share depth-enhanced content via various cloud based storage and social websites. Having a unified format will help to create an ecosystem spanning multiple software and hardware platforms.

Point cloud imaging refers to a set of data points in a given, often 3D coordinate system. Such datasets are usually acquired with a 3D scanner or LIDAR and can subsequently be used to represent and render 3D surfaces. Combined with other sources of data (like light field data, see hereunder), point clouds open a wide range of new opportunities for immersive browsing and virtual reality applications.

Light field imaging is defined based on a representation where the amount of light (the "radiance") at every point in space and in every direction is made available. This radiance can be approximated and captured by either an array of cameras (resulting in wide baseline light field data) or by a single light field camera that uses micro-lenses to sample individual rays of light that contribute to the final image (resulting in narrow baseline light field data). Combinations of the two capture approaches are also possible.

Holographic imaging mainly concerns holographic microscopy modalities that typically produce interferometric data and electro-holographic/computer-generated holographic (CGH) displays that use holographic patterns to reproduce a 3D scene. Considering the maturing of the underlying technologies that are enabling macroscopic holographic imaging systems, it is expected that in the near future this type of imaging data will become ubiquitous. In terms of functionality, holographic data representations will carry even more information than light field representations to facilitate interactive content consultation.

1.3. Goal of the standard

JPEG Pleno intends to provide a standard framework to facilitate capture, representation and exchange of omnidirectional, depth-enhanced, point cloud, light field, and holographic imaging modalities. It aims to define new tools for improved compression while providing advanced functionalities at the system level. It also aims to support data and metadata manipulation, editing, random access and interaction, protection of privacy and ownership rights as well as other security mechanisms.

JPEG Pleno will provide an efficient coding format that will guarantee the highest quality content representation with reasonable resource requirements in terms of data rates, computational complexity and power consumption. In addition to features described next, supported functionalities will include low latency, some degree of backward and forward compatibility with legacy JPEG formats, scalability, random access, error resilience, privacy protection, ownership rights, data security, parallel and distributed processing, hierarchical data processing and data sharing between displays or display elements. The associated file format will be compliant with JPEG Systems specifications and will include signalling syntax of associated metadata for the capture, creation, calibration, processing, rendering, and editing of data as well as for user interactions with such data.

1.4. New features offered by the standard

JPEG Pleno opens doors to new approaches for representing both real and synthetized worlds in a seamless manner. This will enable a richer user experience for creating, manipulating and interacting with visual content. Such new experiences will, in turn, enable a wide range of novel and innovative applications that are either difficult or impossible to realize with existing image coding formats.

JPEG Pleno will facilitate this evolution by offering a collection of new features, complementing those already offered by existing JPEG standards such as scalability, random access, error resilience, high compression efficiency and many more.

Among the most compelling new features offered by JPEG Pleno which are not currently offered by existing JPEG standards one can mention:

1) **Depth of field change:** change the depth of field after capture in a flexible way;

2) **Refocus:** change of focus as well as the ability to refocus on object(s) of interest after capture;

3) **Relighting:** change of lighting, including both number of sources and the direction of lighting in an already captured or synthetized scene;

4) **Motion parallax:** change of viewing perspective from observer's position;

5) **Navigation:** ability to view a scene from different positions and directions with the ability to explore the scene by moving inside it;

6) **Enhanced analysis and manipulation:** facilitating advanced analysis and manipulation of objects within a scene, such as their segmentation, modification, and even removal or replacement by taking into account the richer information extracted from plenoptic data such as depth (either directly or indirectly).

1.5. Potential applications

The above features – both those existing in past standards and the new ones – enable applications that were either only possible with computer generated content or required sophisticated and complex capture or processing algorithms, but will now also be possible for real world scenes or used in virtual, augmented, and mixed-reality scenarios. Here we briefly describe a few illustrative examples.

Depth-enriched photography: Most digital images today are stored in the well-known legacy JPEG file format. Wouldn't it be nice to be able to select two positions in an image and measure their distance? Then you could determine, for example, whether a new sofa you plan to purchase can fit in your living room. This is just one among many possibilities enabled by depth-enriched photography. Your next smartphone enriched by adequate capture devices and capable of representing the resulting content in JPEG Pleno can offer experiences such as the above example.

Enhanced virtual and augmented reality: Today, most visual content in form of images and video from 360-degree capture devices are stitched together based on a fixed geometric mapping. This leads to artefacts due to a missing common nodal point. With depth-based processing based on plenoptic representation, parallax compensated stitching will be possible. In addition, JPEG Pleno can be used to view content from different viewpoints, as when a physical scene is observed in the real world. This means when viewing 360-degree panoramas, you can move around the scene as if you were physically there. Furthermore, real and virtual objects, such as the user's hands and body, can be included in the scene, and interactions will be possible.

Enhanced post production: All blockbusters today include special effects to make the movie more immersive or to allow actors to move beyond physical limitations. For this reason, many parts of movies are computer-generated. With plenoptic image processing, real scenes can be processed similar to synthetic scenes, accelerating the production process. Potential depth-based effects include depth-based colour grading without manual masking of objects, virtual backlots without a green screen, and scene relighting.

Mixed reality, teleconferencing, and telepresence: Wouldn't it be nice to have your relatives in the same room, so you could interact with them even though they might be thousands of kilometres away? Similar to enhanced post production, plenoptic representation offered by JPEG Pleno will enable new immersive experiences as an extension of conventional teleconferencing and telepresence in real time and with ultra low latency.

Digital holographic microscopy (DHM): Microscopes that produce holograms from which intensity and phase images can be derived, for example, cell refractive index tomography that facilitates 3D reconstruction of cells. Examples of life science applications also include monitoring the viability of cell cultures in suspensions, automating multi-well plate screening devices to measure the cell density and cell coverage of adherent cell cultures, or supporting simultaneous fluorescent and holographic cell imaging.

The above illustrative applications may ask for additional data representation requirements such as scalability, random access, and very to ultra-low latency.

1.6. Framework vision in JPEG Pleno

According to Aristotle, "The whole is greater than the sum of its parts." This is also the JPEG Pleno underlying vision as it targets more than a collection of tools to seriously and effectively take into account that there is a unifying root entity: light and its plenoptic representation with various related modalities. This is the conceptual charter for JPEG Pleno, which aims to provide a standard framework for representation of new imaging modalities such as depth-enhanced, light-field, point-cloud and holographic imaging. Furthermore, such imaging modalities should be understood as light representations all inspired by the plenoptic function, regardless of which modality was used to captured or create parts or the entire content. This mind set recognizes that conversions between different modalities are possible and sometimes useful.

The above framework requires an efficient and flexible underlying architecture serving to materialize the system components in the definition of JPEG Pleno standard. Proposals for such system level solutions are therefore very important and essential for the success of JPEG Pleno. The system level solutions should offer efficient syntax in human and/or machine readable formats, support both flexible delivery and storage, must be applicable to both streaming and download scenarios and be adaptable to a large number of application scenarios, some of which are mentioned in this document. The systems level solutions should be both modular and extensible as future enhancements, functional extensions and new features are identified. This in order to facilitate profiling for different classes of applications, if and where needed.

Coding technologies for the various modalities described above are additional important elements of this call. Because of logistics and the potential for different levels of maturity of each modality, the JPEG Committee does not expect to receive proposals for coding of all modalities mentioned above at the same time.

The JPEG Pleno framework will be standardized in a step by step approach. Each step serves to enhance and enrich the JPEG Pleno standard. In this first step, light field coding technologies and associated system level components are addressed. Thereafter, additional CfPs will be issued addressing other specific components of the JPEG Pleno framework.

2. Scope of this call for proposals

This version of the JPEG Pleno CfP will focus on static light field or LF content produced by either lenslet cameras or high-density arrays of cameras. In the text hereafter, 'camera' will always refer to a 'still image camera' unless otherwise stated. Figure 1 shows the reference end-to-end processing chain for plenoptic data as understood by JPEG Pleno for this version of the CfP.

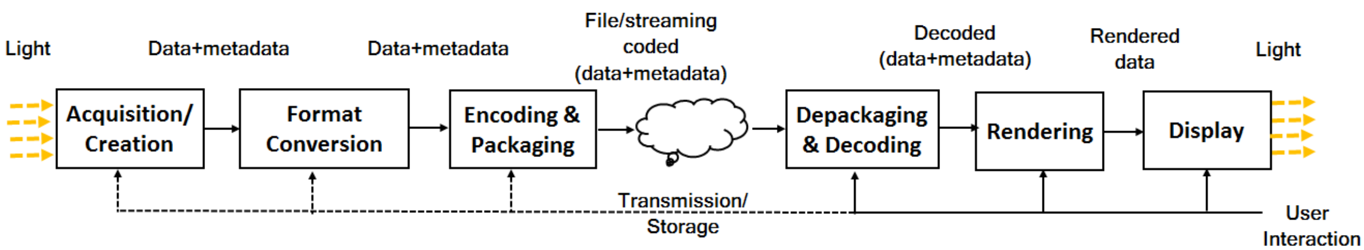


Figure 1 End-to-end processing chain for plenoptic data.

2.1. What is asked in this CfP?

This second version of the JPEG Pleno CfP invites contributions for one or more of the following **normative** items:

- **Coding technologies** for LF content produced by lenslets;
- **Coding technologies** for LF content produced by high-density arrays of cameras;
- **System-level solutions associated with LF coding and processing technology** that have a **normative** impact;

Lenslet cameras incorporate a micro-lens optical array that allows them to record a narrow baseline light field. High-density camera arrays can be realised through either a robot-operated camera manipulation system that is utilizing a regular camera that is picturing the scene from different spatial positions, or via a dense grid of several physically distinct cameras.

Additionally, in this version of the CfP, contributions on several **non-normative** items are encouraged:

- **Use cases and requirements** not yet identified or refinements if already identified up to the final version of the call;
- **Representative datasets** for potential applications identified (or new if not already identified) with conditions allowing usage for standardization and organization of special sessions and grand challenges in scientific events;
- **Rendering solutions** for LFs to serve evaluation purposes;
- **Subjective evaluation methodologies and test-bed implementations** that can be used to assess the various requirements identified (or new requirements if not already identified);
- **Objective evaluation methodologies and test-bed implementations** that can be used to assess the various requirements identified (or new requirements if not already identified).

2.2. Submission requirements

This section lists the several elements that should be delivered to JPEG Committee when submitting a proposal answering to this version of the CfP.

2.2.1. Proposal overview

The proposal overview shall include:

- A description of the proposal contributed as a numbered input document submitted to JPEG Committee according to its guidelines and uploaded to the JPEG document registry. In particular, the title of the document and its introduction should clearly identify which among the different categories called for is being responded to;
- Detailed presentation of the proposal at the JPEG meeting immediately following the date of submission where one of the contributors should be physically present at the meeting to present the proposal and answer to questions and requests for clarifications.

Accepted formats for the submission are Word and PDF. Presentations can be in PowerPoint or PDF.

For proposed datasets, the content should be made available to JPEG Committee members prior to or immediately after the presentation to allow them to access and examine them. The condition of use of the dataset should allow for both standardization as well as publications and presentations in scientific events including organization of grand challenges.

For proposed rendering solutions, an implementation should be made available before the next WG1 meeting, 15-20 January 2017, Geneva, Switzerland, in order to allow JPEG experts to examine them and to take a decision for their inclusion in the final version of the CfP issued after that meeting. Conditions of use for rendering solutions should allow the JPEG Committee and its affiliates (e.g. liaison organizations or test laboratories in charge of evaluations) to use and adapt the technology for the purpose of assessment of technologies submitted to JPEG Pleno.

For proposed subjective evaluation methodologies, an implementation of the test-bed should be made available before the next WG1 meeting 15-20 January 2017, Geneva, Switzerland, in order to allow JPEG experts to examine them and to take a decision for their inclusion in the next version of the CfP. Conditions of use for the test-bed should allow JPEG Committee and its affiliates (e.g. liaison organizations or test laboratories in charge of evaluations) to use and adapt the evaluation methodologies for the purpose of assessment of technologies submitted to JPEG Pleno.

For proposed objective evaluation metrics, an implementation of the metric in source code should be made available before the next WG1 meeting 15-20 January 2017, Geneva, Switzerland, in order to allow JPEG experts to examine them and to take a decision for their inclusion in the next version of the CfP. Conditions of use for the source code should allow JPEG Committee and its affiliates (e.g. liaison organizations or test laboratories in charge of evaluations) to use and adapt the software for the purpose of assessment of technologies submitted to JPEG Pleno.

2.2.2. Proposal technical description

Each proposal shall include a document with a detailed description of the proposed algorithm, systems component(s), use cases and requirements, image material, rendering solution and/or objective/subjective quality assessment methodologies. This presentation shall be in a text document format (Word or PDF). The presentation shall additionally clearly explain how the proposed algorithm and/or system architecture meets each requirement listed in Annex B. Proponents are encouraged to list all features, benefits and performance advantages of their proposed technologies, including complexity issues.

2.2.3. Software binaries

In case of proposals responding to system level solutions and/or coding technologies, proponents need to submit separate encoder and decoder executable programs (statically linked Linux executables including all required libraries and system dependencies), configurable via command line or configuration file. Binaries should preferably be optimized software meeting the performance requirements described in Annex B in order to speed up the evaluation process.
Proponents can choose to use executable compression or similar tools to prevent reverse engineering or disassembly of the submitted executable files.
Proponents shall provide the command-line parameters intended to be used for the evaluation procedures described in Sections 2.3 and 2.4. Scripts for generating the test content shall also be provided for every test case.

2.2.4. Test materials

Proponents submitting coding technologies need to submit materials for objective and subjective evaluations at pre-defined bit rates and/or compression ratios. Details regarding which content and which bit rates and/or compression ratios along with the format and procedures to follow are provided in Sections 2.3 and 2.4.

2.2.5. Technical documentation

If (part of) a proposal has been selected to be considered for inclusion in the standard, a technical description of the selected technology shall be provided by the proponents to the JPEG Pleno standard editors. This includes:

- Description of operations, algorithms and design;
- File format, codestream and bitstream syntax;
- Coding process (encoding and decoding) methodology.

2.2.6. Verification model source code

Proponents agree to release source code to serve as part of a Verification Model (VM), written in a high-level language such as C or C++ if they are selected to be potentially part of the standard. Source code shall be documented and understandable. Assembly language or GPU code is not accepted.
All libraries used by the source code shall be either public or provided as source code with ISO/IEC and ITU-T compliant terms. Make files or project files need to support compilation on Windows, MacOS and Linux platforms.

2.3. Evaluation of technical proposals for lenslet light field coding

This section describes the evaluation details of technical proposals addressing coding technologies for static lenslet LF. In the following details, are given about the assumed processing flow used to obtain the coded and decoded LF, starting from the LF camera raw sensor data. Furthermore, the assessment methodology to be used to evaluate the coding technologies submitted in response to this CfP is presented.

2.3.1. Processing Flow

Many methods exist to convert lenslet LF camera raw sensor data to what are usually called subaperture images. In this CfP, a simple and practical approach is used, relying on the Matlab implementation of the Light Field Toolbox v0.4[2]. This toolbox converts lenslet LF camera raw sensor data to subaperture images following the process described in [1, 2].
The end-to-end coding-decoding chain of the lenslet LF– which we will refer to hereafter as a *lenslet image* – is provided in Figure 2 and includes the following steps:

1. First, the 10-bit precision raw sensor data is pre-processed by applying demosaicking and devignetting.
2. The 10-bit precision data is clipped to 8 bit by dropping the two least significant bits. It is currently under consideration if the full processing should happen at 10 bits and not at 8 bits. The Final CfP to be issued after January 2017 JPEG Committee meeting will state the final decision.
3. The RGB 4:4:4, 8-bit non-uncompressed lenslet image (at point **A** in Figure 2) is coded to produce a bitstream by means of a coding algorithm. This can either be an anchor or a proposed encoder.
4. The bitstream is subsequently decoded to obtain the decoded lenslet image in RGB 4:4:4 (at point **B**) by means of a decoding algorithm. Again this can either be an anchor or a proposed decoder.
5. Subsequently, before rendering, the decoded lenslet image in RGB may be converted to the set of subaperture images (at point **C**), i.e. a stack of 2D images where each image represents a different viewpoint. This representation is typically more suitable for rendering and viewing by the Matlab toolbox explained above. An example of a decoded lenslet image and the corresponding subaperture images as provided by the above mentioned toolbox is shown in Figure 3.

The stack of subaperture images is composed of 2D low-resolution RGB images plus a weighting image. The weighting image carries the confidence associated with each pixel, which can be useful in filtering applications that accept a weighting term [3]. For example, for the Lytro images in the test set, the resulting dimension of the subaperture image stack is $15 \times 15 \times 434 \times 625 \times 4$. Here 15×15 represents the number of perspective views, 434×625 represents the resolution of each view and 4 corresponds to the RGB and weighting image components.

[2] http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4

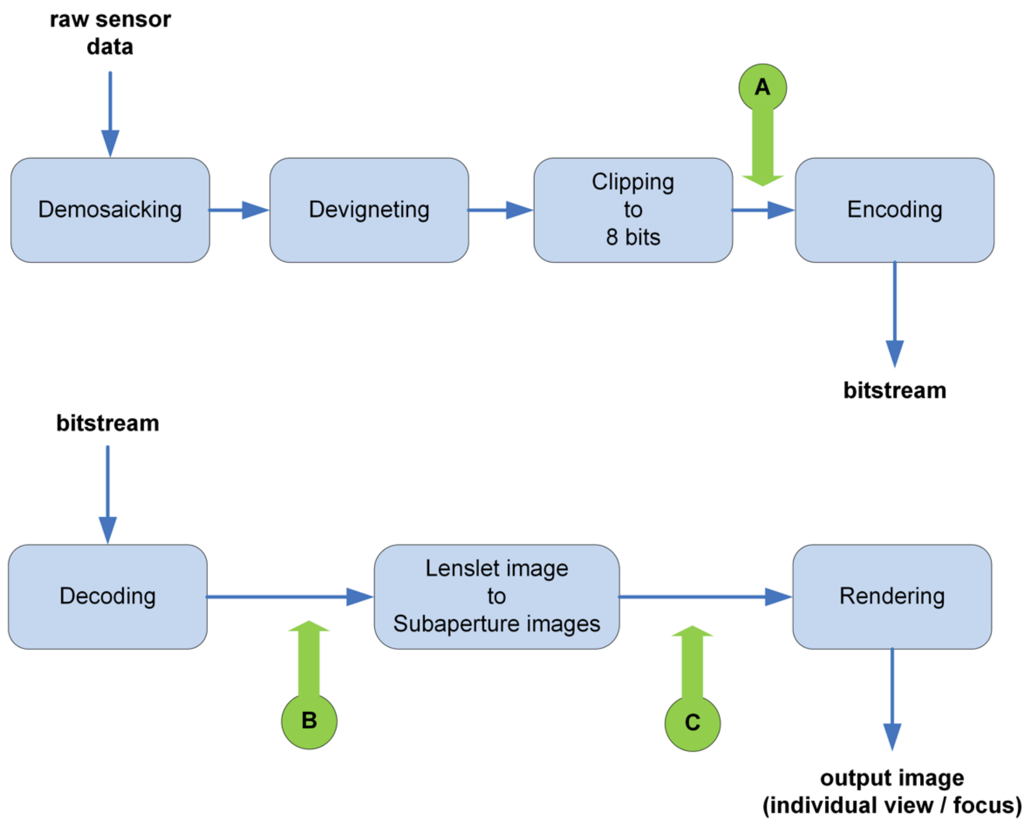
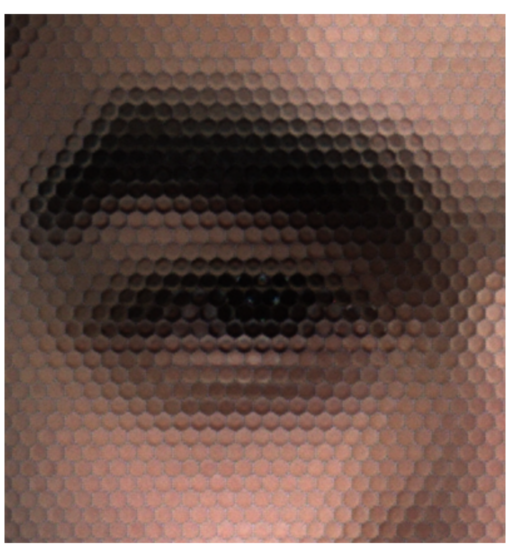
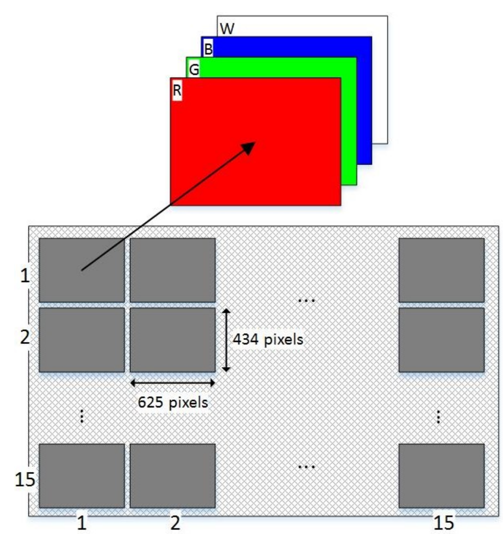


Figure 2 End-to-end processing chain showing compression and decompression of lenslet LF.



a) Crop of an uncompressed lenslet image in RGB obtained after demosaicing and devigneting of raw sensor data.



b) Subaperture images as produced by the LF toolbox. For the Lytro test data, the dimensions of LF structure are $15 \times 15 \times 434 \times 625 \times 4$

Figure 3 Example of a lenslet image and its corresponding subaperture image stack.

The processing flow to generate the reference LF data is illustrated in Figure 4 below.
The original lenslet image at point $\mathbf{A_{Ref}}$, which equals point **A** in Figure 2, is obtained as described above in steps 1 and 2. Then the non-compressed reference lenslet image is converted to the reference subaperture images (at point $\mathbf{C_{Ref}}$). This stack of 2D perspective views will serve as reference data for full reference visual quality assessments and subjective quality assessment used in this evaluation.

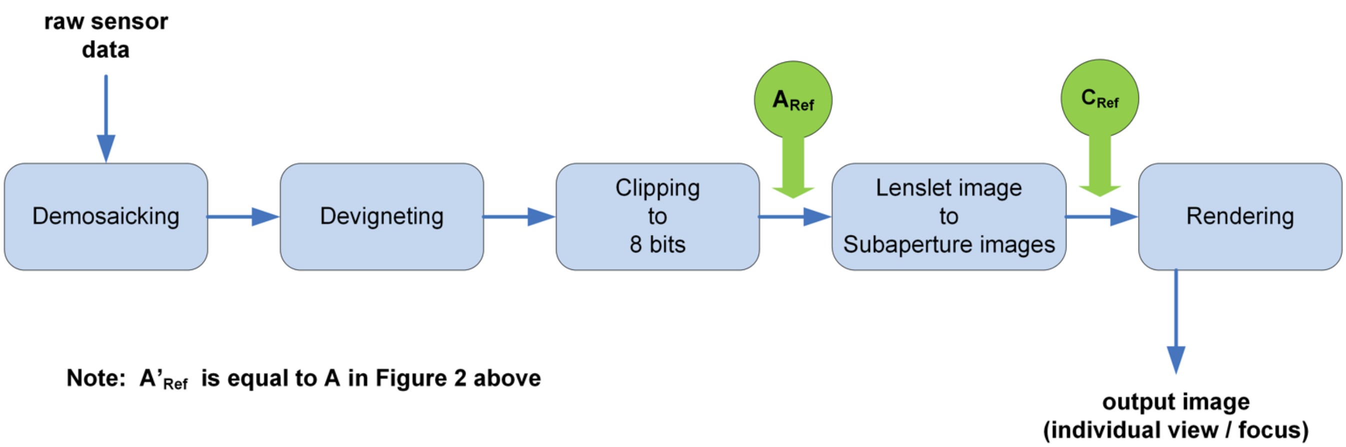


Figure 4 Processing chain for generation of the reference subaperture image stack. $\mathbf{A_{Ref}}$ is equal to **A** in Figure 2.

2.3.2. Objective performance evaluation

Objective quality assessment will be carried out to evaluate how the quality of the contents has changed as a result of the applied processing. Quality changes are induced through coding with the proposed technology at predefined bit rates and/or compression ratios and after rendering of the decoded content according to specific features against which the technology is to be assessed (e.g. change in depth of field, refocusing, relighting, motion parallax). The objective metrics are selected from state-of-the-art methods available in the literature and for which there is sufficient evidence that they provide good prediction of subjective quality.

The distortion for a proposed coding solution will be measured relative to the non-compressed reference image data at various bit rates corresponding to compression ratios defined in Table 1. More specifically, the objective quality metric will be computed between the subaperture images at point **C** created using the decoded LF at point **B** and the reference subaperture images at point $\mathbf{C_{Ref}}$ created from the content available at point $\mathbf{A_{Ref}}$. Both the objective quality of the individual views (2D low-resolution RGB images) and the global quality of all subaperture images will be evaluated.

The evaluation of the proposed coding algorithms will be performed according to conventional objective metrics: PSNR_Y, PSNR_YUV, SSIM_Y and SSIM_YUV as defined in the following:

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

MSE(k,l) is defined for the k^{th} and l^{th} viewpoint, respectively in the horizontal and the vertical direction as:

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

m and *n* correspond to dimensions of one individual view (one 2D low-resolution image) in units of pixels. For Lytro images m = 434 pixels and n=625 pixels. *I(i,j)* and *R(i,j)* are the values of the pixels in position *(i,j)* of the corresponding colour channel (Y, U or V) for the respective evaluation and reference views.

The PSNR_YUV value for each individual view is computed as:

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

The average PSNR_Y_ value is computed as the mean of all PSNR_Y values for each individual view as:

$$PSNR_{Y_{mean}} = \frac{1}{k\,l}\sum_{k=2}^{K-1}\sum_{l=2}^{L-1} PSNR_Y\,(k,l),$$

k and l correspond to the k^{th} and l^{th} individual view while K and L correspond to the number of views in the horizontal and vertical directions respectively (so $K=L=15$). The Mean value for PSNR_YUV is then computed as:

$$PSNR_{YUV_{mean}} = \frac{1}{k\,l}\sum_{k=2}^{K-1}\sum_{l=2}^{L-1} PSNR_{YUV}\,(k,l).$$

Similarly, the $SSIM_Y$ for the k^{th} and l^{th} viewpoint will be calculated as:

$$SSIM_Y(k,l) = \frac{(2\mu_I\mu_R + c_1)(2\sigma_{IR} + c_2)}{(\mu_I^2 + \mu_R^2 + c_1)(\sigma_I^2 + \sigma_R^2 + c_2)}$$

where c_1 *and* c_2 are defined as $(0.01*L)^2$ and $(0.03*L)^2$ respectively with L being the dynamic range of the pixel values (in an 8-bit setting: $L=2^8-1$).
$SSIM_{YUV}$, $SSIM_{Ymean}$ and $SSIM_{YUVmean}$ are calculated analogous to the PSNR metric.

2.3.3. Subjective performance evaluation

Subjective quality assessments will be carried in order to assess the impact of the applied processing to the perceived quality. Quality changes are introduced by coding and decoding with the proposed technology at predefined bit rates and/or compression ratios. The decoded content is then rendered according to target features against which the technology is to be assessed (e.g. change in depth of field, refocusing, relighting, motion parallax). The subjective evaluation methodology will be based on a double stimulus with reference approach, carried out in an environment adapted to the use cases for which the technology is evaluated.

Formal viewing tests are performed to assess the subjective quality of the resulting rendered 2D images obtained from each coding proposal when compared to a coding anchor and the non-coded reference. Based on past experience, the users will make a relative assessment of the quality for two stimuli. Here each stimulus will be presented to the viewer in the form of a video showing the results of navigating the LF image through several view points and focal planes.

The visual comparison between the individual images rendered from the subaperture images (point **C**) obtained from decoded lenslet images (point **B)** and their corresponding images (point $\mathbf{C_{Ref}}$) from the reference (point $\mathbf{A_{Ref}}$) (see Figures 2 and 4 above). At least 3 target views and 3 focus points will be randomly selected for visual testing by the test organizers. Test subjects will be viewing a video sequence showing the results while navigating between views and focus points. The selected focus points will be rendered using the Matlab toolbox as adopted for this evaluation. Please note that the rendered images will be corrected for colours using the Matlab toolbox *(DecodeOptions.OptionalTasks = 'ColourCorrect')* and for gamma (*1/2.2*).

2.3.4. Test materials and coding conditions

This section defines the test materials and coding conditions for the lenslet LF case.

2.3.4.1. Test Material, Anchors and Coding Conditions

The lenslet test material for this CfP is defined in Table 1 and is accessible from the JPEG website[3]. This dataset also includes the relevant camera calibration data.

The anchors are generated according to the procedure outlined in Section 2.3.1 using Legacy JPEG as the coding algorithm.

Image ID	Image name	Compression ratios			
		R1	R2	R3	R4
I01	Bikes	10:1	20:1	40:1	100:1
I02	Danger_de_Mort	10:1	20:1	40:1	100:1
I03	Flowers	10:1	20:1	40:1	100:1
I04	Stone_Pillars_Outside	10:1	20:1	40:1	100:1
I05	Vespa	10:1	20:1	40:1	100:1
I06	Ankylosaurus_&_Diplodocus_1	10:1	20:1	40:1	100:1
I07	Desktop	10:1	20:1	40:1	100:1
I08	Magnets_1	10:1	20:1	40:1	100:1
I09	Fountain_&_Vincent_2	10:1	20:1	40:1	100:1
I10	Friends_1	10:1	20:1	40:1	100:1
I11	Color_Chart_1	10:1	20:1	40:1	100:1
I12	ISO_Chart_12	10:1	20:1	40:1	100:1

Table 1 Image test sequences and compression ratios

The compression ratios are computed as ratios between the size of the sensor raw data in 10-bit precision, i.e. 5368 × 7728 × 10b = 51 854 880 bytes, and the size of the resulting coded bitstream. The compression ratios indicate in the table above are the minimum compression that has to be achieved.

Depending on the number of proponents and the size of the test dataset, a smaller number of test images may be selected for subjective testing. This to keep the extent of the test procedure manageable.

3 https://jpeg.org/plenodb/

2.3.4.2. Materials for proponents

Proponents to this evaluation are provided with:

1. LFR images (*.LFR) as taken by a Lytro camera including camera calibration data (available at JPEG Website, https://jpeg.org/plenodb/).
2. Non-compressed lenslet images (*.RGB) after clipping to 8 bit (point **A** in Figure 2).
   a. Matlab scripts to perform computation of PSNR and SSIM values.
3. JPEG anchors
   a. Compressed JPEG bitstreams (in *.JPG file format) generated according to Figure 2.
   b. Corresponding subaperture image stacks (in *.MAT file format) created from compressed JPEG bitstreams.
4. JPEG 2000 anchors
   a. Compressed JPEG 2000 bitstreams (in *.JPG file format) generated according to Figure 2.
   b. Corresponding subaperture image stacks (in *.MAT file format) created from compressed JPEG 2000 bitstreams.
5. Reference images
   a. Subaperture image stacks (in *.MAT file format) created from non-compressed lenslet images.

2.4. Evaluation of proposals for high-density camera arrays image coding

This section describes the evaluation details of technical proposals addressing coding technologies for LF images produced by high-density camera arrays. In the following, more details are given about the assumed processing flow used to obtain the coded and decoded high-density camera array LFs. Furthermore, the assessment methodology to be used to evaluate the coding technologies submitted in response to this call is presented. The evaluation procedure for high-density camera array image coding will be further completed and refined based on the response received by next WG1 meeting. Hence, the procedure below intends to provide a basic outline of the envisaged evaluation methodology.

2.4.1. Processing Flow

The end-to-end coding-decoding chain for the high-density camera array LF is provided in Figure 5 and includes the following steps:

1. First, the camera array LF is acquired, in this case with a planar camera arrangement using a X-bit RGB representation.
2. The set of original views (at point A) is coded, thus creating a bitstream for which the total bit rate should be measured.
3. The bitstream is subsequently decoded to obtain the decoded camera array LFs in RGB format (at point B) using a corresponding decoding algorithm. The encoding process may correspond either to the selected anchor or a proposed coding solution. For the reference solution, no encoding-decoding is performed at all. This decoding process must output decoded frames for all positions for which an original view is available.
4. While rendering could be applied to create additional virtual views, this step is outside the processing flow in which performance will be assessed. This approach is valid under the assumption that the camera array density is truly so high that the distance between two neighbouring views is

virtually indistinguishable for human vision and thus intermediate virtual camera positions become unnecessary. If after examination of the contributed datasets, this assumption is not confirmed and the Final CfP may consider the creation of virtual views using some pre-defined view synthesis software and additional objective quality metrics will consider the quality performance for such synthesized views.

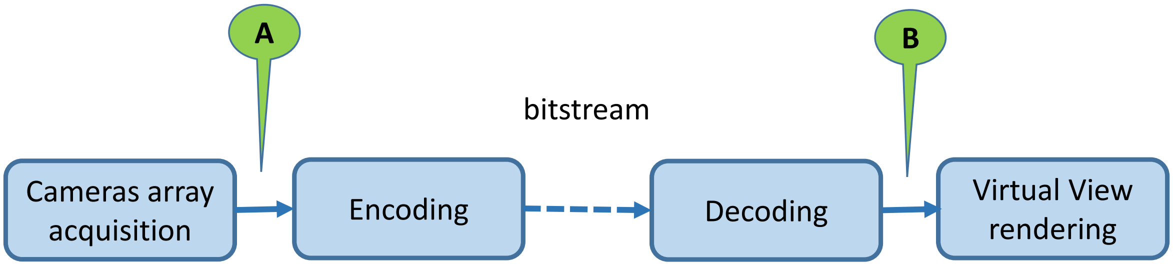


Figure 5 Processing chain for the cameras array LF images.

2.4.2. Objective performance evaluation

The objective assessment performance for the high-density camera array LF will use similar metrics as for the lenslet LF case. Notably, for each rate point, the following metrics will be measured:

1. Average PSNR_Y computed for all perspectives corresponding to positions where an input view exists (meaning that no virtual perspectives are considered). PSNR_Y is measured as defined above.
2. Average PSNR_YUV computed for all perspectives corresponding to positions where an input view exists (meaning that no virtual perspectives are considered). PSNR_YUV is measured as defined above.
3. Average SSIM_Y computed for all perspectives corresponding to positions where an input view exists (meaning that no virtual perspectives are considered). SSIM-Y is measured as defined above.

Again, it is important to stress that the objective metrics and conditions above have been selected under the assumption that the density of camera array LFs are so high that the distance between two neighbouring views is virtually indistinguishable for the human vision and thus computation of intermediate virtual camera positions become unnecessary.

2.4.3. Subjective performance evaluation

The subjective quality assessment will adopt precisely the same method as for the lenslet LF, notably the assessment protocol defined in Section 2.5.3; Only multiple perspectives are to be considered in this case. In this context:

1. The assessment protocol involves the creation of a video sequence resulting from the navigation of multiple perspectives within the high-density camera array LF.
2. The subjective assessment protocol involves double stimulus assessment with reference, where the reference is here the video created with the same process while using the decoded camera array LFs.
3. One of the stimuli to be assessed will correspond to the anchor coding conditions defined in the next section.

4. The precise sequence of navigation perspectives will be defined by experts involved in the AhG on JPEG Pleno before the subjective tests and after the proposal deadline in order to avoid optimization of proposals by proponents.
5. The rules to create the video based on the selected sequence of perspectives will be defined in the Final CfP.

2.4.4. Test materials and coding conditions

This section defines the test materials and coding conditions for the high-density camera array LFs case.

2.4.4.1. Test Material, Anchors and Coding Conditions

The high-density camera array LF material for this CfP are defined in Table 2 and will be accessible from the JPEG website[4]. The dataset also includes the relevant camera calibration data.
The anchors are generated using one of the following: i) Legacy JPEG; ii) Legacy JPEG 2000. Other anchors could be added in the final version of this call.
The precise definition of the anchor and software codec used to generate results will be detailed in the Final CfP.

Cameras array LF ID	Cameras array LF name	Compression ratios			
		R1	R2	R3	R4
J01	Content 1	10:1	20:1	40:1	100:1
J02	Content 2	10:1	20:1	40:1	100:1
J03	Content 3	10:1	20:1	40:1	100:1
J04	Content 4	10:1	20:1	40:1	100:1
J05	Content 5	10:1	20:1	40:1	100:1
J06	Content 6	10:1	20:1	40:1	100:1
J07	Content 7	10:1	20:1	40:1	100:1
J08	Content 8	10:1	20:1	40:1	100:1
J09	Content 9	10:1	20:1	40:1	100:1
J10	Content 10	10:1	20:1	40:1	100:1
J11	Content 11	10:1	20:1	40:1	100:1

Table 2 High-density camera array LF and compression ratios.

[4] https://jpeg.org/plenodb/

The compression ratios are computed as ratios between the total size of the camera array LF in X-bit precision and the size of its corresponding coded bitstream file. The compression ratios indicate the minimum compression that has to be achieved, meaning that the rate should not be above that limit.
Depending on the number of proponents and the size of the test dataset, a smaller number of LF content may be selected for the subjective test procedure in order to keep the size of the test procedure manageable.

2.4.4.2. Materials for proponents

Proponents to this evaluation are provided with:

1. Original images for the various perspectives of the high-density camera array LF
2. Calibration data for the corresponding acquisition
3. Anchors
    a. Compressed bitstreams (in a file format that will be defined in the final version of this call) generated according to Figure 5.
    b. Decoded version of the anchor LF in a format that will be defined in the final version of this call for the selected test material.

More details about anchors will be provided in the final version of the CfP.

2.5. Evaluation of system level solutions proposals

This section describes the procedure used to assess system level solutions submitted as a response to the call.
Evaluation of such responses mainly reply on a descriptive analysis that should be submitted by proponents in a written form, followed by an interactive session between proponents and JPEG experts in order to allow for questions and further clarifications to take place.
In addition to a detailed description of components and general architecture of the system level solutions, components proponents should include in their response how at a functional level each requirement identified in the call will be fulfilled by their respective system level solution and motivate the advantages of such an approach if/when comparable to other alternatives.

2.6. Timeline for Call for Proposals

The following schedule is planned for the development of JPEG Pleno specifications from the initial CfP to publication of the standard. This schedule is subject to changes either in response to received responses to this or next versions of the CfP, or as a consequence of subdividing the JPEG Pleno specifications into multiple parts:

06/16	First version of the CfP on Light Field Coding
10/16	Second version of the CfP on Light Field Coding
01/17	Final version of the CfP on Light Field Coding
04/17	Anchor evaluation results
05/17	Submission deadline for responses to CfP
07/17	Responses evaluation results available
10/17	Working draft 1 (WD) and core experiments

01/18	Working draft 2 (WD) and core experiments
04/18	Committee Draft (CD) and validations
10/18	DIS
01/19	IS

2.7. IPR conditions (ISO/IEC Directives)

Proponents are advised that this call is being made in the framework and subject to the common patent policy of ITU-T/ITU-R/ISO/IEC and other established policies of these standardization organizations. The persons named below as contacts can assist potential submitters in identifying the relevant policy information.

2.8. Contribution to Standardization

Proponents are informed that based on the submitted proposals, a standard specification will be created. If they submit a proposal and (part of) the proposed technology is accepted for inclusion in the standard, they will hence have to attend subsequent WG1 meetings and contribute to the creation of the different standard documents. Within this process, evolution and changes are possible as several technologies may be combined to obtain a better performing solution.

2.9. Further information

2.9.1. JPEG Pleno Ad hoc Group

A JPEG Pleno Ad Hoc Group has been established between 73rd and 74th JPEG meetings in order to continue activities and progress with the planned work. All interested parties are requested to register to the email reflector of the AhG.

E-mail reflector: jpeg-pleno@listserv.uni-stuttgart.de
In order to subscribe to the mailing list, please follow the link:
https://listserv.uni-stuttgart.de/mailman/listinfo/jpeg-pleno
and follow the steps of the e-mail being received.

2.9.2. Use cases and requirements

WG1N72006 - "JPEG Pleno – Scope, use cases and requirements" contains currently identified use cases and requirements for JPEG Pleno. This document is regularly updated while progressing towards the final CfP. The JPEG Pleno AhG has the mandate to further update the use cases and requirements between face to face meetings which will then have to be approved by JPEG experts during its face to face meetings. This document can be obtained via one of the contacts below.

2.9.3. Contacts

Touradj Ebrahimi (JPEG Convener)
Email: Touradj.Ebrahimi@epfl.ch

Peter Schelkens (JPEG Coding & Analysis Subgroup Chair)

Email: pschelke@etro.vub.ac.be

Fernando Pereira (JPEG Requirements Subgroup Chair)
Email: fp@lx.it.pt

Siegfried Foessel (JPEG System Subgroup Chair)
Email: siegfried.foessel@iis.fraunhofer.de

Thomas Richter (JPEG Test & Quality Subgroup Chair)
richter@rus.uni-stuttgart.de

David McNally
david.mcnally@epfl.ch

2.9.4. References

[1] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens and F. Pereira et al. Objective and subjective evaluation of light field image compression algorithms. 32nd Picture Coding Symposium, Nuremberg, Germany, December 2016.

[2] I. Viola, M. Rerabek and T. Ebrahimi. A new approach to subjectively assess quality of plenoptic content. SPIE Optics + Photonics, San Diego, California, USA, 2016.

[3] E. Upenik, M. Rerabek and T. Ebrahimi. A Test bed for Subjective Evaluation of Omnidirectional Visual Content. 32nd Picture Coding Symposium, Nuremberg, Germany, December 2016.

Annex A – Use cases

This section intends to present uses cases where LF imaging may bring added value relative to more conventional imaging. With LFs, significantly more information is acquired, thus capturing a richer visual scene structure with textural and geometric information. With more information available, it should be possible to reach better performance for uses cases currently utilizing 'classical' imaging modalities – for example digital photography or video post-production – and to address new uses cases like industrial inspection.

A.1 Light field photography

LF photography promises to bring new experiences to the photographer. With one shot, the photographer will capture all information related to the scene (assuming a ray-based light model). This modality limits the complexity of the photographic process for the photographer as the focus distance and the proper acquisition angle can be decided upon when editing the captured material. Moreover, this opens also the avenue towards interactive pictures, which will represent a new realm in digital photography.

A.2 Video production

Video production encompasses content capturing, editing and transcoding for final distribution.

A.2.1. Capturing

LF based video capture aims to record a much more complete representation of the scene compared to that of classical 2D capture. This then allows for the captured video to be manipulated more easily during postproduction. As a consequence, LF based video implies capturing of a variety of viewpoints while handling the full dynamic range of the illumination.
Video capture is performed either live (for immediate delivery and consumption) or the content is for offline processing and later delivery. In both cases, data will typically be processed by image processing algorithms and systems. Consequently, LF capture needs both to be able to handle large amounts of data and to enable easy access to the stored content. Depending on the application, the quality requirements of the captured content differ. Similarly, different LF capture devices can be used, such as plenoptic cameras or multi-camera arrays.

A.2.2. Post-production

Post-production encompasses all the production steps after capturing. The recorded LF data needs to be pre-processed, possibly reconstructed, combined or manipulated in another manner and finally composed into a pre-distribution format.
Post-production is characterized by repetitive access to the captured or intermediately processed content. Intermediate storage between the individual processing steps is fundamental to allowing for a smooth pipeline offering high quality which can be achieved in a reasonable amount of time. Depending on the application sector, a large variety of processing programs are used, each having different capabilities.

Often, a combination of different tools is used which requires an interoperable way to exchange data. Introduction of special effects is part of post-production. Due to the better manipulation and processing capabilities, LFs open the door for new visual effects such as virtual camera movement, relighting, depth of field adjustment, post-refocus, compositing and many more.
These visual effects can then be used to create videos that are difficult to achieve with traditional capture techniques. This includes for instance the change of illumination to generate a certain mood, the replacement of a green screen by a virtual backlot, or the correction of deficiencies having occurred during capture.

A.3 Industrial imaging

Metrology based on LF imaging may be useful for numerous types of applications. With more information available, a better analysis, decision and control performance can be achieved, particularly increasing robustness to difficult environmental conditions (e.g. unfocused, low light, rain, fog, snow, smoke, and glare), unstructured scenes and the constraints of an unstable or moving platform.
Among the relevant analysis functions that may be performed, there are many computer vision related functions: mapping, modelling, segmentation, localization, depth measurement, tracking, classification, object recognition, and also biometrics related functions, e.g., face, gait, and palm print recognition.
Current industrial LF cameras range from 1-40 MP, while offering frame rates of up to 180 fps.
Examples of relevant applications domains are:

- **Robotics –** Robotics deals with the design, construction, operation, and application of robots, as well as computer systems for their control, sensory feedback, and information processing. These technologies deal with automated machines that can take the place of humans in dangerous environments or manufacturing processes. In this context, better analysis for better decisions, e.g., controlling the actions of a robot, moving a robot around, etc., are key needs. LF based vision may be a critical development in this area in terms of sensing the visual world.
- **Non-destructive testing -** Non-destructive testing (NDT) is a type of analysis techniques used in industry to evaluate the properties of a material, component or system. Because NDT does not damage the article being inspected, it is a highly valuable technique that can save both money and time in product evaluation, troubleshooting, and research.
- **3D fluid analysis –** Measuring and analysing accurately fluid dynamics is important for many application domains.
- **3D plant and animal analysis –** Non-invasive analysis is important to deal with plants and animals, e.g., to control their growth and well-being. LFs may improve through more accurate analysis means the performance of these processes.

Figure 5 Example of plant analysis utilizing a LF based camera[5]

Although not-really an industrial application, similar control and surveillance may be performed with humans in general, e.g. for surveillance, and elderly and young people, e.g., for well-being monitoring.

A.4 Visualization

Current display technology enables the visualization of LFs in several different ways. 2D Displays can present a 2D view of the LF to the human spectator. 3D displays can reproduce the LF as a scene with stereoscopic depth perception viewed from a fixed position. By combining a 2D or 3D display with a head tracker device, the presentation of content can be adapted to the movement of the human spectator. Such head tracker can be used both with traditional displays as well as with special head-mounted displays.
Methods that use current display technology for LF data visualization reduce the dimensionality of the LF in some form to adapt to the display technology. Moreover, they still suffer from the vergence-accommodation conflict, which may induce viewing discomfort. New displays developed specifically for 3D LF visualization are able to reproduce the complete LF content, such that the spectator can enjoy a three-dimensional environment without suffering from the vergence-accommodation conflict as in traditional stereo 3D displays. In order to accomplish the full reproduction of the LF, displays must achieve very high resolutions, frame rates and data processing capability. For example, Alpaslan et al [1] reported a display formed by combining several tiles of high-resolution microdisplays. The tiled LF display has 2500 views (50 horizontal x 50 vertical) and uses high-density pixel technology, with 10μm pixel pitch. The display has an angular pitch of 0.6° with 80x32 elemental images, generating 6.4 megapixels. LF displays for near-eye devices also use microdisplays, such as the ones presented in [2] and [3]. Such displays are used for Virtual and Augmented Reality, and their pixel pitch is in the order of microns ([2] utilizes 8.01

[5] **Source: raytrix and Max Planck Institute of Molecular Physiology [3]**

μm pixel pitch, while [3] uses a microdisplay with 12 μm pixel pitch). The proposed near-eye displays utilize LCD or OLED panels with HD resolution for each eye ([2] uses a 1080p display, while [3] uses a 720p display). LF displays with larger form factor targeting entertainment and TV broadcast are proposed in [4] and [5]. Such systems generate a large number of pixels (33 megapixels in [4] and 10 megapixels in [5]), with various viewing angles (measured viewing angle of 24° is reported in [4], and a 50° field-of-view is described in [5]). LF displays with high angular and spatial resolution allow for more natural 3D viewing.

References

[1] Alpaslan, Z. Y., El-Ghoroury, H. S., “Small form factor full parallax tiled light field display,” Proceedings of Electronic Imaging, IS&T/SPIE Vol. 9391, February 9, 2015

[2] Wang, J., Xiao, X., Hua, H. and Javidi, B., “Augmented reality 3D displays with micro integral imaging” in Display Technology, Journal of , vol.11, no.11, pp.889-893, Nov. 2015

[3] Lanman, D. and Luebke, D. “Near-eye light field displays”. In ACM SIGGRAPH 2013 Emerging Technologies (SIGGRAPH '13). ACM, New York, NY, USA, 2013

[4] Arai, J., "Three-dimensional television system based on integral photography," in Picture Coding Symposium (PCS), 2012 , vol., no., pp.17-20, 7-9 May 2012

[5] T. Balogh, "The HoloVizio system," Proc. SPIE 6055, Stereoscopic Displays and Virtual Reality Systems XIII, 60550U (January 27, 2006); doi:10.1117/12.650907

Annex B – Technology Requirements

B.1 Generic JPEG Pleno requirements

B.1.1 Representation model

The JPEG Pleno standard should adopt as small as possible a number of representation models. These shall fit the various addressed imaging modalities and the various stages in the processing pipeline.

Note: From [1], "As the acquisition of image data changes to support new applications and functionalities, it is expected that the representation model must also adapt"; also "if technically feasible, it would be highly desirable to have a unified format for these new imaging modalities, which could facilitate further exchange and interoperability between the new formats."

B.1.2 Colour representation

JPEG Pleno shall support high dynamic range colour definitions, wide colour gamut, XYZ colour space, ICC profiles, transparency and opacity.

B.1.3 JPEG Backward compatibility

The JPEG Pleno standard should fit well within the JPEG standard ecosystem, notably by providing some degree of backward compatibility with JPEG, JPEG 2000, etc.

A standard is backward compatible when the new specification includes the old one. This means that any devices implementing the new standard can also interpret all data compliant with the old version of the standard. An old device, however, only compliant with the old version of the standard might not be able to interpret the data compliant with the new version of the standard [2].

Note: From [1], "As we move towards the development of standards that will support these new representation formats, it will be essential to consider interoperability with widely deployed image formats, such as JPEG and JPEG 2000. Although interaction and manipulation features may be limited, enabling such compatibility would allow any existing media browser or device to view conventional images, e.g., derived from a LF representation."

B.1.4 JPEG Forward compatibility

The JPEG Pleno standard should fit well in the JPEG standard ecosystem, notably by providing some degree of forward compatibility with JPEG, JPEG 2000, etc.

A new standard is considered to be forward compatible when devices only compliant with the old version of the standard are nevertheless able to interpret the data conforming with the new standard. However, it might be possible that the obtained results are not as good as when using a device compliant with the new version of the standard [2].

B.1.5 JPEG Systems compatibility

The systems elements of the JPEG Pleno standard shall comply with the relevant JPEG Systems specifications.

B.1.6 Compression efficiency

Considering the large amount of raw data associated to the new imaging modalities, the JPEG Pleno standard shall provide the highest possible compression efficiency, given a set of functionalities supported when compared to the raw representation and considering the available state-of-the-art solutions in the literature. Means to reach near lossless and lossy coding up to high quality shall be provided. Means to reach lossless coding should be provided.

B.1.7 Compression efficiency/functionality tuning

The JPEG Pleno standard should provide the means to trade off compression efficiency for other functionalities/features, notably random access, scalability and complexity.

B.1.8 Random access

The JPEG Pleno standard shall provide efficient methods to allow random or partial access to subsets of the complete compressed image data (e.g. parts of an image, selected directions). The JPEG Pleno standard should provide random access with fine granularity.

B.1.9 Scalability

The JPEG Pleno standard should provide tools to achieve scalability in terms of:

- quality (SNR);
- spatial, depth and spectral resolution;
- number of viewing angles;
- viewing angle range;
- complexity;
- content (object).

These types of scalability shall result in a very flexible scaling of the imaging information.

B.1.10 Editing and manipulation

The JPEG Pleno standard shall provide means for editing and manipulation such as change of depth of field, refocusing, relighting, change of viewpoint, navigation, enhanced analysis of objects.

B.1.11 Error resilience

The JPEG Pleno standard should provide the tools to achieve error resilience, both in terms of bit errors and packet losses, for a large set of networks and storage devices. Error resilience should consider graceful degradation and graceful recovery, associated to all or only parts of the content.

B.1.12 Low complexity

The JPEG Pleno standard should allow for low complexity in encoding and decoding, while simultaneously enabling low end-to-end content processing complexity (postproduction, rendering…). Complexity needs to be considered in terms of computational complexity, memory complexity and parallelisation.

B.1.13 Metadata

The JPEG Pleno standard shall provide appropriate content description tools for efficient search, retrieval, filtering and calibration of content from the various imaging modalities. Strategies and technical solutions shall fit within the JPSearch and JPEG Systems frameworks.

The JPEG Pleno standard should be able to support the use cases and requirements by a single metadata standard framework. All modalities supported by the JPEG Pleno framework should use the same underlying metadata signalling syntax.

The JPEG Pleno standard should leverage an abstraction model that could adjust itself for different combinations of these metadata categories.

B.1.14 Privacy and security

The JPEG Pleno standard shall provide means to guarantee the privacy and security needs associated with content from the various imaging modalities. Strategies and technical solutions shall fit within the JPEG Systems Privacy & Security.

B.1.15 Support for parallel and distributed processing

The JPEG Pleno standard should facilitate distributed processing or parallelization of the decoding process. Solutions should account for the data consumption patterns of the successive processing steps (e.g. rendering).

B.1.17 Latency and real-time behaviour

The JPEG Pleno standard should facilitate low latency and real-time processing implementations.

B.1.18 Support for hierarchical data processing

The JPEG Pleno standard should facilitate hierarchical data processing, enabling fast and easy data segmentation. Solutions should allow for further adaptation/partial decoding.

B.1.19 Sharing of data between displays or display elements

JPEG Pleno should support signalling syntax to enable sharing of content between different displays or display elements.

B.2 Specific light field coding requirements

B.2.1 Representation format

JPEG Pleno shall support

- a relevant set of spatial and angular resolutions
- multiple colour/spectral components

- 6 degrees of freedom

Furthermore, JPEG Pleno should support

- capture and display dependent/independent LF representation
- universal LF representation

B.2.2 Support for calibration model in metadata

JPEG Pleno shall support the signalling of a calibration model in metadata to enable correct content transformation in the imaging pipeline.

B.2.3 Synchronization of data between sensors

JPEG Pleno should support signalling syntax to enable synchronization of content captured by different sensors.

B.2.4 Support for different (non)linear capturing configurations

The JPEG Pleno standard shall support the representation of content produced by various linear/nonlinear configurations and signalling the associated metadata: micro-lens arrays on sensors, sensor arrays of different regular or irregular layouts, rotating sensors/objects…

B.2.5 Carriage of supplemental depth maps

JPEG Pleno should support carriage of supplemental depth maps as part of the codestream or file format.

References

[1] JPEG Pleno Abstract and Executive Summary, Doc. ISO/IEC JTC 1/SC 29/WG1 N6922, Sydney, Australia, Feb. 2015.

[2] "Information Technology: Scalable Compression and Coding of Continuous-Tone Still Images, ISO/IEC 19566-1 JPEG Systems Part1: Packaging of Information using Codestreams and File Formats", PDTR-2, 2015.