ISO/IEC JTC1/SC29/WG1 N74020
74th Meeting – Geneva, Switzerland – 15-20 January 2017

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC 1/SC 29/WG1
(ITU-T SG16)

Coding of Still Pictures

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

TITLE: JPEG Pleno – Scope, use cases and requirements Ver.1.7

SOURCE: Peter Schelkens (Editor)

PROJECT: JPEG Pleno

STATUS: Draft

REQUESTED ACTION: Review by WG1 and Innovations AhG

DISTRIBUTION: WG1

Contact:
ISO/IEC JTC 1/SC 29/WG1 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

Editorial Comments

This is a living document that goes through iterations. Proposals for revisions of the text can be delivered to the editor Peter Schelkens, by downloading this document, editing it using track changes and sending it to peter.schelkens@vub.ac.be.

If you have interest in JPEG Pleno, please subscribe to the mailing list, via the following link:
https://listserv.uni-stuttgart.de/mailman/listinfo/jpeg-innovations

JPEG Pleno

Scope, Use Cases and Requirements

V1.6 dd. October 20th , 2016

1 Scope

JPEG Pleno targets a standard framework for the capture, representation, and exchange of new imaging modalities such as light field, point cloud and holographic imaging. Additionally, it also targets to define new tools for improved compression while providing advanced functionality and support for — but not limited to — image manipulation, metadata, image access and interaction, privacy and security (JPEG Pleno Executive Summary [1]).

The applications that may benefit from these emerging imaging 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 applications should have different associated functionalities, notably depending on the relevant type of data, processing and addressed scenario.

In this document, use cases and requirements are collected for the representation of these new imaging modalities.

2 Target imaging modalities

JPEG Pleno targets new imaging modalities able to provide richer scene representations [1]. The new representations may not be directly visualizable, like images in the past, but rather may need to be rendered to recover a displayable version of the imaging content.

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

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 2D image. However, partial and truncated spherical and cylindrical configurations have also been referred to an omnidirectional in some cases.

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 contents via various cloud based storage and social websites. Having a unified format will help in creation of an ecosystem across multiple software and hardware platforms.

A **point cloud** is a set of data points in a given coordinate system (Figure 1). Such data is often acquired with a 3D scanner or LIDAR and subsequently used to generate and represent a 3D surface. 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 data (aka plenoptic data) records the amount of light (the “radiance”) at every point in space, in every direction (Figure 2). This radiance can be approximated and captured by either an array of cameras (resulting in wide baseline light field data) or by a light field camera that uses microlenses to sample each individual ray of light that contribute to the final image (resulting in narrow baseline light field data). Furthermore, light field data can be synthetically generated from computer models, and combinations of captured imagery, point clouds and synthetically rendered light fields can be fused together.

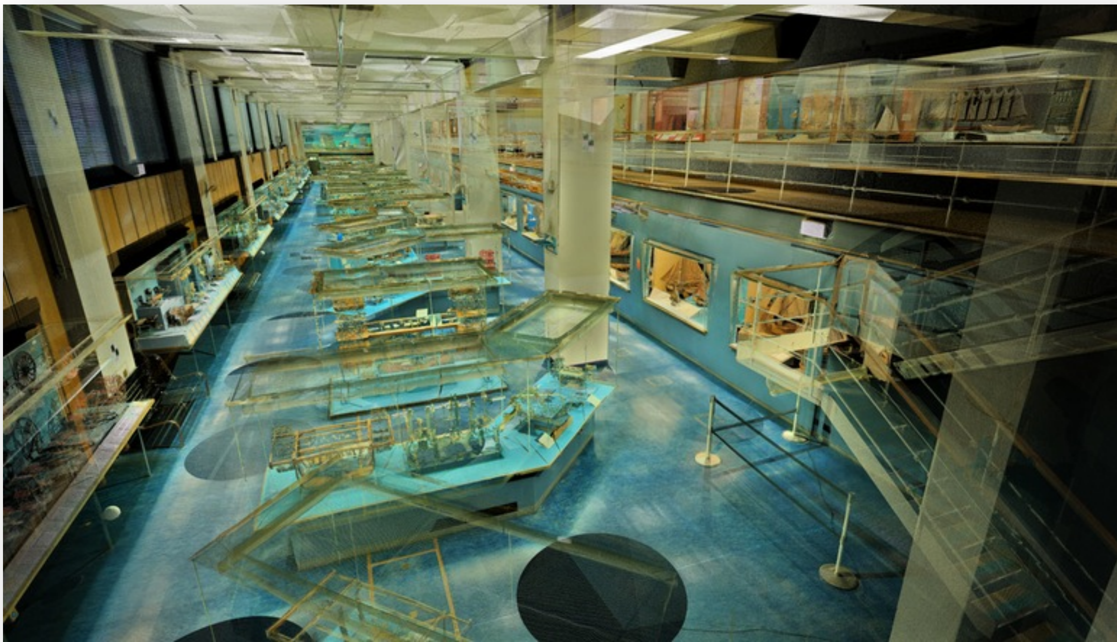

Figure 1 – Point cloud example: Shipping Gallery in London’s Science Museum
(Source: Science Museum London and ScanLAB Projects Ltd)

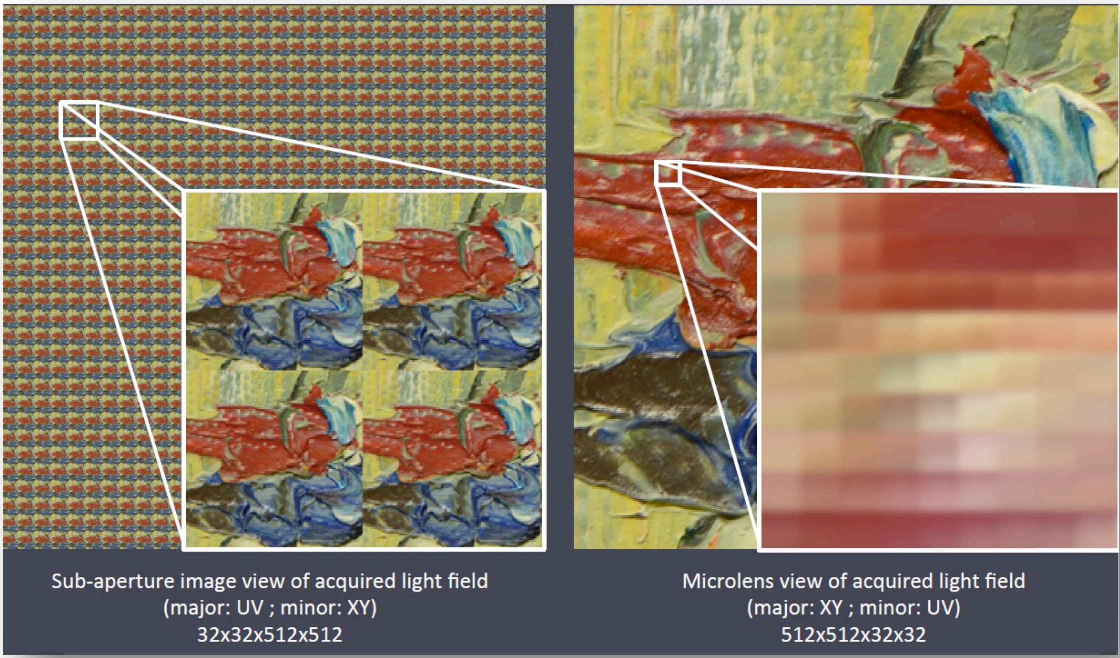

Figure 2 – Light field or plenoptic image and its micro-images

In contrast to light fields, which are based on geometric optics or a ray-based model to describe light propagation, **holograms** are based on a wave-based light propagation model and are typically represented by complex numbered, interferometry data. Holographic data is currently mainly generated in the context of microscopy applications (life sciences and non-destructive testing). However, considering the maturing of the underlying hardware technologies that are enabling macroscopic holographic imaging systems, it is to be expected that in a near future also this type of data will be flooding our imaging markets. In terms of functionality, holographic data representations will carry even more information than light field representations to facilitate interactive content consultation (for example in museums, galleries, and so on). It is important to realise that the holographic data generated for these applications will not be directly retrieved from holographic capturing systems as the case in microscopy, but rather using for example classical images, light field images and point cloud data as input and deploying computer-generated holography techniques to generate the holograms.

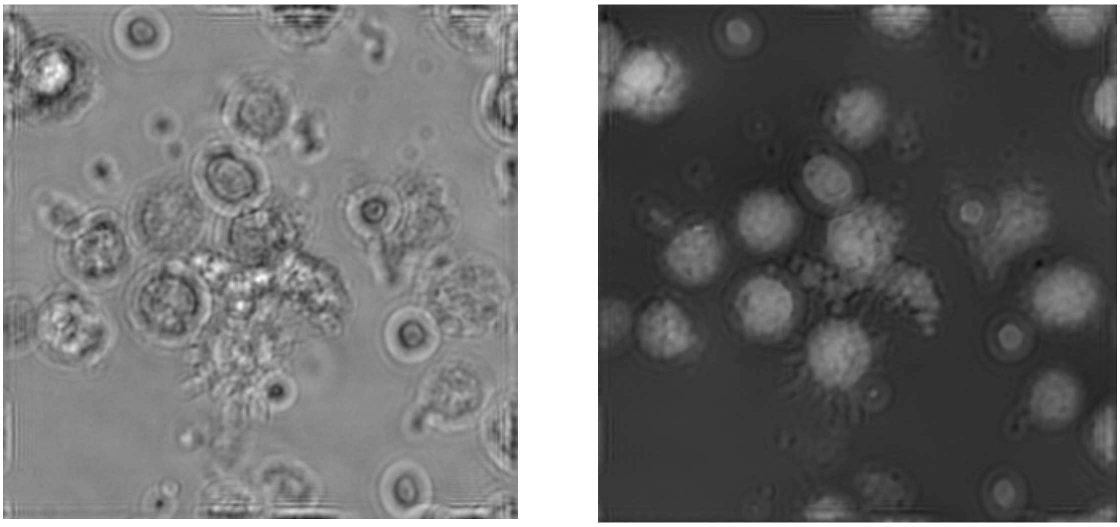

Figure 3 – Amplitude (left) and phase component (right) of an holographic microscopy image
(Source: Ovizio Imaging Systems)

These new data types can then be processed to recover additional scene information and to render this information in novel ways.

3 General requirements

This section defines general requirements, which may apply to all addressed imaging modalities, notably light field, point cloud and holographic imaging.

3.1 Representation model

The JPEG Pleno standard should adopt an as small as sufficient 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."

3.2 Colour representation

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

3.3 JPEG Backward compatibility

The JPEG Pleno standard should fit well in 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 light field representation."

3.4 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].

3.5 JPEG Systems compatibility

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

3.6 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.

3.7 Compression efficiency/functionality tuning

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

3.8 Subjective quality

The JPEG Pleno standard shall incorporate coding tools that exploit properties of the human visual system.

3.9 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.

3.10 Scalability

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

- quality (SNR);
- spatial, depth, temporal 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.

3.11 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.

3.12 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.

3.13 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.

3.14 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. Having separate standards for each of the numerous scenarios outlined in this document, the resulting system would be heavily fragmented. Instead, it would be more advantageous to engineer a single ecosystem by representing the various types of metadata outlined here (e.g. point cloud, depth map, multiple images) as part of a single data structure. This metadata framework should be able to support other metadata sources needed to support specific use cases.

The JPEG Pleno standard should leverage an object-oriented model, that could adjust itself for different combinations of these metadata categories and cover many of these distinct use cases.

3.15 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 Privacy & Security and JPEG Systems framework.

3.16 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).

3.17 Latency and real-time behaviour

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

3.18 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.

3.19 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.

4 Light field data

4.1 Use cases

This section intends to present uses cases where light field imaging may bring added value relative to more conventional imaging. With light fields, significantly more light 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 performances 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.

4.1.1 Light Field Photography

Light field 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 – and as such limit the complexity for the photographer. Fixing for example the focus distance and the proper acquisition angle can be decided upon afterwards when editing the captured material. Moreover, this opens also the avenue towards ‘living’ pictures, which will represent a new realm in digital photography.

Current sensor resolutions are approx. 40 MP resulting in effective resolutions of 2-10 MP depending on the angular resolution targeted and the microlens array configuration.

4.1.2 Video production

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

Capturing

Compared to classical 2D capture and as in computational photography, light field based video capturing aims to record a much more complete representation of the scene, such that the video can be manipulated more easily afterwards in postproduction. This essentially means capturing a variety of viewpoints while handling the full dynamic range of the illumination.

Video capture can be live, meaning that human spectators immediately consume it, or it can be recorded for offline processing. In both cases, data will typically be processed by image processing algorithms and systems. Consequently, light field 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 purpose, the quality requirements of the captured content differ. Similarly, different light field capture devices can be used, such as plenoptic cameras or multi-camera arrays.

Post-production

Post-production encompasses all the production steps after capturing. The recorded light field 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 for allowing a smooth pipeline, where high quality can be achieved in 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, requiring an interoperable way to exchange data.

Introduction of *special effects* is part of video post-production. Due to the better manipulation and processing capabilities, light fields open the door for new visual effects, such as virtual camera movement, relighting, depth of field adjustment, post-refocus, compositing, any 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.

Transcoding

Transcoding is the process of transforming the video content into a representation suitable to the end user. With today’s 2D technology, resizing, cropping, colour adjustment and re-encoding are the most often performed operations. Light fields offer in addition the possibility to generate different viewing points from one scene, allowing for instance the adjustment of the 3D-stereo baseline depending on the viewing conditions.

4.1.3 Industrial imaging

Metrology based on light field 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 light field 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. Light field 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. Light fields may improve through more accurate analysis means the performance of these processes.

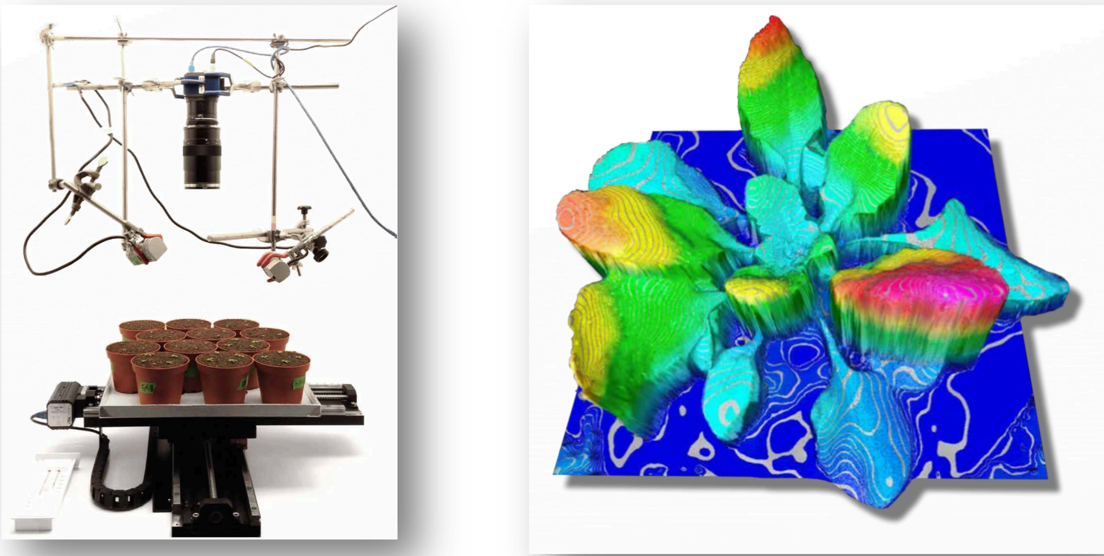

Figure 4 - Example of plant analysis utilizing a light field based camera
(Source: raytrix and Max Planck Institute of Molecular Physiology [3])

- 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.

4.1.4 (Bio)medical imaging

To be completed

4.1.5 Visualization

Current display technology enables the visualization of light fields in several different ways. 2D Displays can present a 2D view of the light field to the human spectator. 3D displays can reproduce the light field 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 light field data visualization reduce the dimensionality of the light field 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 light field visualization are able to reproduce the complete light field 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 light field, displays must achieve very high resolutions, frame rates and data processing capability. For example, Alpaslan et al [10] reported a display formed by combining several tiles of high-resolution microdisplays. The tiled light field 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. Light field displays for near-eye devices also use microdisplays, such as the ones presented in [11] and [12]. Such displays are used for Virtual and Augmented Reality, and their pixel pitch is in the order of microns ([11] utilizes 8.01 µm pixel pitch, while [12] uses a microdisplay with 12 µm pixel pitch). The proposed near-eye displays utilize LCD or OLED panels with HD resolution for each eye ([11] uses a 1080p display, while [12] uses a 720p display). Light Field displays with larger form factor targeting entertainment and TV broadcast are proposed in [13] and [14]. Such systems generate a large number of pixels (33 Megapixels in [13] and 10 Megapixels in [14]), with various viewing angles (measured viewing angle of 24° is reported in [13], and a 50° field-of-view is described in [14]). Light field displays with high angular and spatial resolution allow for more natural 3D viewing.

4.2 Specific Requirements

4.2.1 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 light field representation
- universal light field representation

4.2.2 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.

4.2.3 Synchronization of data between sensors

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

4.2.4 Support for different (non)linear capturing configurations

The JPEG Pleno standard shall support the representation of content produced by various linear/nonlinear configurations: microlens arrays on sensors, sensor arrays of different regular or irregular layouts, rotating sensors/objects…

4.2.5 Storage of depth maps as intermediate data

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

5 Point cloud Data

5.1 Use cases

5.1.1 Virtual and Augmented Reality

Advances in 3D capture and reconstruction enable real-time generation of highly realistic 3D visualisation for 3D tele-presence (cfr. Figure 1). 3D Point clouds are an efficient representation as they can be seamlessly integrated and rendered in 3D virtual worlds enabling a convergence between real and virtual realities. As point cloud capture and reconstruction from single or multiple calibrated cameras are simpler compared to 3D mesh reconstruction, it makes the representation particularly suitable to such real-time application. In this case, point clouds are reconstructed, compressed transmitted and rendered in real-time as in video conferencing systems, enabling conversational style communication.

When the original image grid and camera parameters are still present, point clouds can be defined on the image grid and this can be referred to as an organized point cloud (8-bit per colour component and up to 12-bit depth for each grid entry). When each point is defined for explicit xyz coordinates without the original image grid structure (i.e. all data has been fused in the reconstruction stage to a full 3D reconstruction), the point cloud is considered as unorganized. In this case, the number of points (resolution) is arbitrary and depends both on the inputs and the segmentation/reconstruction algorithm. xyz coordinates can then be represented by 32-bit floating point values. However, often these are converted to 9-20 bit integers by quantizing the coordinate values. Point clouds with quantized coordinates may be used to represent occupancy grids, which are common in robotics but are also used in virtual and augmented reality. Occupancy grids may be thought of as voxelized point clouds whose quantized coordinates represent occupied volumetric elements or voxels. A floating point representation of the coordinates is often used as input format for graphics APIs, but an integer representation is more suitable for voxelized point clouds as well as scanned 3D point cloud data, as the 3D scanners often have a fixed depth granularity/resolution and accuracy.

In addition, the colours are represented by 8-12 bits for each component. For highly realistic rendering a 3D world scene, support for colour and additional attribute information such as normals, material, reflectance, radiance and transparency properties, is desired. For example, normals can be represented by 3 floating point components (e.g. in a spherical coordinate system), while material reflectivity can be represented by material properties (such as on a 0-1 floating point scale) and transparency/opacity by point specific attributes.

The 3D point clouds are generally rendered at frame rates of 5 to 30 fps (possibly at a varying frame rate) and the number of vertices can range up to 1 million points. To support these frame-rates and transmission in current networks, bit rates between 1 and 60 Mbps should be targeted, requiring lossy encoding support.

As such applications introduce stringent requirements on the end-to-end media chain, JPEG Pleno requirements for real-time and low complexity are of importance to this use case. In addition to support immersive 3D networking, view selectivity, region of interest and resolution granularity/scalable bit stream are of importance [4].

5.1.2 Motion Capture

Motion Capture with or without markers can result in point clouds corresponding to each marker. These type of motion capture data are relevant for motion analysis and synthesis for various applications such as sports analysis, video game production, choreography, movie production. Point clouds are a useful representation for data resulting from multiple sensors or from marker-based/markerless motion reconstruction. Efficient and interoperable compression and storage methods are a necessity in such systems. In such scenarios the number of of points can range from 100-10000 or even more points.

5.1.3 Geographic Information Systems

Geographic information is often represented as point clouds (resulting from airborne SAR, Lidar) and are often too massive to be handled efficiently in current geo-ICT infrastructures [5]. Therefore, an efficient compression and data representation that is standardized is critical to enable efficient point cloud data management systems (PCDMS) [5]. A point cloud data management system enables users to query subsets of point cloud data of interest at a specified Level of Detail (LoD). PCDMS for geographic information systems have a specific set of user requirements that can be found in [5] appendix A. PCDMS combines two technologies, a database management system and underlying storage/binary format for storing and transmitting point cloud data. Some of the requirements for PCDMS need to be supported directly in this underlying compressed format. Firstly, fast region/spatial selectivity and resolution granularity (LoD) are particularly important for PCDMS and need support from the underlying compressed format. In addition, the compressed storage format should have the ability to scale to huge clouds with billions of points. This use case may be further developed in a liaison with ISO/TC211. An example of a GIS point cloud dataset is the AHN2 dataset that contains 640 billion points mapping the Netherlands height data in 6-10 points per square meter [6].

5.1.4 Cultural heritage

Point cloud scans are one of the means to archive and visualisation cultural heritage objects and collections. Examples are the Shipping Gallery of the London Science Museum that was scanned before it was replaced by a new exhibition (the captured point cloud is composed of 256 Gb of data) and the Shipping Gallery in the Deutches Museum [7]. Another example is the project Culture 3D Cloud [8] where pictures taken by smart phones of cultural heritage objects are transformed into 3D point clouds that can be viewed from any angle. Datasets from 3D cultural heritage can have over millions of points and similar attributes as defined in §5.1.1**Error! Reference source not found.**

5.1.5 Large scale 3D maps

Rich 3D maps of indoor and outdoor environments can be created using devices that provide localization combined with depth and colour measurements of the surrounding environment. Localization could be achieved with GPS, inertial measurement units (IMU), cameras, or combinations of these and other devices, while the depth measurements could be achieved with time-of-flight, radar or laser scanning systems. Example mapping systems are already commercially available and come in various forms such as mobile vehicles (www.mitsubishielectric.com/bu/mms/) or flying drones (www.sensefly.com/home.html). There are also lightweight mobile platforms under development (www.google.com/atap/project-tango/). The resulting 3D maps, which are composed of 3D point clouds with various attributes, could be used for a variety of applications, including finding your way around indoor spaces, navigation of vehicles around a city, space planning, topographical surveying or public surveying of infrastructure and roads, augmented reality, immersive online experiences, and much more.

5.2 Specific Requirements

5.2.1 Representation format

*Organized point cloud*s keep the original camera parameters and bear similarity to multiple RGB + depth images (8-16 bit per colour component, 8-16 bit for the depth component) and a fixed NxK resolution that corresponds to the original RGB + D image, enabling traditional coding methods to be re-used. On the other hand *unorganized point clouds* have explicit xyz attributes and are independent of the camera parameters. *Unorganized point clouds* are often represented in an octree structure which enables sparse representation of the points in the 3D space. Typically, the 32-bit floating point x, y and z coordinates can be converted to linearly quantized integers (9-20 bits per component) and the associated colour information is represented by 8-16 bit per colour component. In addition, the number of points is arbitrary.

In both cases it is important to store – in addition to colour - extra *attribute values* of each point (3 floats for normal components, optionally material/reflectivity attributes).

Summarized:

- The 3D point cloud representation shall/should include 3D position (X,Y,Z) with a specification of its accuracy and dynamic range.
- The 3D point cloud representation should/shall support multiple attributes being associated with each 3D position including colour, reflectance, normal vectors and transparency.
- The 3D point cloud representation should support connectivity information such as the connectivity among 3D points and the order that 3D points were captured.

5.2.2 Compressed Data format

The compressed data format shall support methods to dynamically achieve *resolution granularity* (or level of detail LoD or resolution) based on *viewing angle (VA)* and spatial *region of interest (ROI)* selectivity. Optionally a scalable or progressive bit stream shall be supported. In practice, point clouds can exist of billions of points and based on the viewing angle, distance and region of interest of the user it shall be possible to only decode and retrieve subsets of the data at the right LoD, i.e. random access support. In addition, the compressed data format should support the coding of attributes in the various use cases (colour, normal, brdf, opacity).

Summarized:

- The 3D point cloud compression format should/shall provide a compact representation of the 3D point cloud representation format.
- The 3D point cloud compression format should/shall support reconstruction of 3D position and associated attributes with varying precision, including (mathematically) lossless.
- The 3D point cloud compression format should/shall facilitate efficient access to a subset of 3D points in the representation, e.g., within a specified volume.
- The 3D point cloud compression format should/shall support dynamic resolution granularity based on viewing angle and spatial region of interest (ROI) selectivity.
- The 3D point cloud compression format should/shall support scalable and progressive reconstruction.
- The 3D point cloud compression format should support the coding of the following attributes: colour, normals, Bidirectional Reflectance Distribution Functions (BRDFs), opacity.
- The 3D point cloud compression format should support selection of horizontal and full parallax.

6 Holographic data

6.1.1 Use cases

6.1.2 Holographic (tomographic) microscopy

Digital holographic microscopes (DHM) produce holograms from which – by using hologram reconstruction algorithms – intensity and phase images can be derived. DHM enables to extraction of 3D structural information, to perform post-acquisition refocusing of the images.

Holographic microscopy techniques can be categorized into two groups: (1) two-dimensional (2D) holographic microscopy, which measures complex amplitude maps, consisting of both the amplitude and phase delay maps and (2) 3D holographic microscopy (also known as optical diffraction tomography or holotomography), which measures three-dimensional (complex) refractive index maps. Representative holographic images obtained with 2D and 3D holographic microscopy are shown in Figure 5.

Figure 5.a shows the holographic image of a human red blood cell measured with Diffraction phase microscopy [16], a 2-D holographic microscopy technique. The image in Fig. 1A represents the optical phase delay map of the cell, which provides topographic information. Although the amplitude map of a sample can also be retrieved, it is not shown here. Figure 5.b presents the 3-D refractive index tomogram of white blood cells, measured by 3-D holographic microscopy. The 3-D holographic microscopy technique is an optical analogy to X-ray computed tomography (CT). 3-D holographic microscopy reconstructs 3-D refractive index distributions of a sample from multiple 2-D holograms measured with various angles of illuminations.

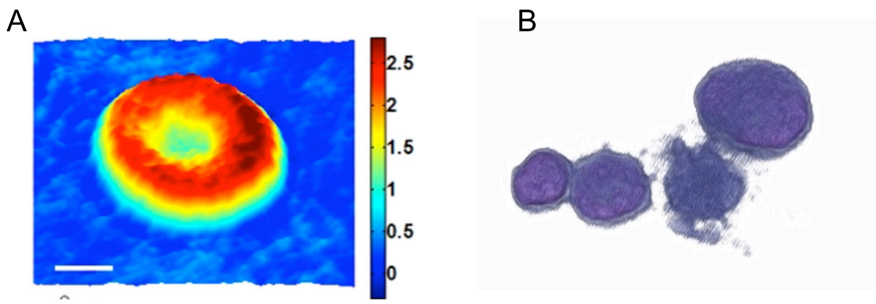

Figure 5 (a) 2-D holographic image of a human red blood cell. Scalar bar, 2 □m. (b), 3-D holographic image of human white blood cells.

Typically, the data format for 3D holographic microscopy is also single precision (IEEE Standard 754). For example, the HT-1H system, a commercialized 3D holographic microscopy system, has 67 Mega voxels; the imaging volume is 84.3 x 84.3 x 41.4 μm with a resolution of 112 x 112 x 356 nm for x, y, and z direction, respectively. A single tomogram of 67 Mega voxels will have a size of 268 Mbyte. When a sequence with 512 frames is recorded with 3D holographic microscopy, the data size will be approximately 134 GB.

Examples of life science applications for holographic microscopy are given in the next paragraphs.

Monitoring of cell density and viability of cell cultures in suspensions, such as the one cultivated in bioreactors in the biopharmaceutical industry: in this set-up the DHM is connected to a closed loop fluidic system that interfaces the DHM with the bioreactor/cell culture. The images of the cells passing through the flow cell are recorded in a continuous way, stored and post-processed during the cell growth process that can take multiple weeks. An image analysis and classification procedure is deployed to determine automatically the cell density and the cell viability in real time, during the whole cell growth process. The hologram images have currently one component (bit depths of up to 12 bit per component) – multi-wavelength acquisition is expected though – and a minimum resolution of 4 MP. In a typical session, movies of 30 frames per hour, for a total of 6.2 GB of holograms per day, or 130 GB for a typical monitoring session of 3 weeks.

Automated multiwell plate screening devices measure the cell density and the cell coverage of adherent cell cultures, such as the one cultivated in multiwell or multi-layer recipients in the biopharmaceutical industry. The devices have a motorized stage, enabling to move the recipient in front of the microscope objective. This way, multiple images of the culture area can be recorded and stitched together to create large panorama images. An image analysis procedure is performed on each image to compute the cell density and the cell culture coverage area off the growth surface. For a typical multiwell containing 48 wells and using an objective magnification of x10, 250 images are necessary to acquire its surface. To scan all the 48 wells, 12000 images are necessary, which amounts to 96 GB of hologram data.

Simultaneous fluorescence and holographic cell imaging, allowing the comparison of the optical holographic signature of a sample with its fluorescence response. The fluorescence images are classical hologram images, with additional image channels for each fluorescence mode. The fluorescence channel images can be overlaid on the actual hologram image with a user chosen colour. In a typical acquisition session, the scientist can record several hundreds of fluorescence/hologram images per hour, which amounts to a total of several GB of data per hour.

6.1.3 Metrology of small objects

Holography is utilised in non-destructive industrial inspection to measure accurately the depth profiles of objects. Nanometer scale accuracy can be achieved on the condition that phase unwrapping procedures succeed in unambiguously unwrapping the measure phase information.

6.1.4 Macroscale Holography: Entertainment and Medical

For macroscale scenes, like it is the case for medical and entertainment use cases, the holograms can not immediately be recorded using the typical interferometric set-up, but have to be generated from regular image sources (2D camera's, CT, MRI), computer generated graphics and computational imaging devices (e.g. light field and depth sensing cameras, and LIDAR scanners). Hence, the physical process of light propagation needs to be modelled and simulated on a computer, allowing for the generation of computer generated holograms (CGHs). Unfortunately, generating holograms for 3D objects requires significant amounts of computation time, especially when occlusions are taken into consideration. Therefore, many solutions have been proposed over the years for accelerating this process. These algorithms can essentially be divided into three categories: (1) polygonal methods, which represent the virtual objects in the scene using meshes consisting of triangles, (2) plane based method expresses the 3D object as the aggregation of planar planes, and (2) point cloud methods describing objects as a collection of self-luminous points.

Dynamic digital holographic systems require – even for humble screen sizes – pixels rates in the order of magnitude 10^{12} ~ 10^{14} pixels/second (≈ 106 HD screens) and data rates in the order of Tbps. This translates into exascale computing power with an associated power consumption of more than 10 MW. By limiting the reconstruction to a significantly reduced region and viewing angle, utilizing eye-tracking, etc. this computational complexity can probably be reduced with factor 10000. Nonetheless – the holy grail of holography – a digital holography system enabling full-parallax, large viewing angle holographic video remains a huge challenge.

6.2 Specific requirements

6.2.1 Representation format

JPEG Pleno shall support complex amplitude representations. Potential instantiations of of complex amplitude representations, are amplitude-phase, complex, shifted-difference representations, or refractive index [9].

6.2.2 Space-bandwidth product

The JPEG Pleno shall support sufficiently high space-bandwidth products for capture and displays systems to meet image size and viewing angle requirements.

6.2.3 Colour representation

In addition to the generic colour representation requirements, JPEG Pleno shall support colour specifications typically needed for holographic rendering systems.

6.2.4 Horizontal / full parallax

JPEG Pleno should enable the signalling and selection of horizontal and full parallax support.

6.2.5 Decouple capture and display

JPEG Pleno shall incorporate signalling syntax or provide a mechanism to describe capturing and/or rendering conditions in order to transform content from capturing reference to rendering reference in the processing pipeline.

6.2.6 Display specific processing

JPEG Pleno shall support signalling of information necessary for display specific processing steps.

6.2.7 Additional processing steps

JPEG Pleno should not jeopardize twin removal, noise reduction … at display side.

6.2.8 Metadata

The following metadata should be signalled [15]:

- Capture information (image resolution, channel content, number of images, time stamp)
- Microscope configuration (microscopy types: 2D/3D, reflective/transmissive, objective lens parameter: NA(numerical aperture), focal length, DOF(depth of focus), eye lens parameter: magnification);
- Calibration data and/or image;

- Channel information (channel types, illumination wavelength);
- Fluorescence metadata information (fluorochrome, exposure time);
- Information about additional processing steps (e.g. brightness/contrast correction, intensity ratio between reference and object beams);
- Image sensor (CCD/CMOS) information: pixel pitch, bit-depth, gamma value.

7 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.

[3] "A Small Revolution: The 3D Light Field Camera", Deutches Museum Poster, 2015.

[4] Wanmin Wu, Ahsan Arefin, Gregorij Kurillo, Pooja Agarwal, Klara Nahrstedt, and Ruzena Bajcsy. 2011. Color-plus-depth level-of-detail in 3D tele-immersive video: a psychophysical approach. InProceedings of the 19th ACM international conference on Multimedia (MM '11).

[5] Peter van Oosterom, Oscar Martinez-Rubi, Milena Ivanova, Mike Horhammer, Daniel Geringer, Siva Ravada, Theo Tijssen, Martin Kodde, Romulo Gonçalves, Massive point cloud data management: Design, implementation and execution of a point cloud benchmark, Computers & Graphics, Volume 49, June 2015, Pages 92-125, ISSN 0097-8493.

[6] "Actueel Hoogtebestand Nederland (AHN), http://www.ahn.nl/pagina/open-data.html, 2015.

[7] http://www.pointclouds.org/news/2014/10/20/navvis-digitizes-deutsches-museums-shipping-section/

[8] http://c3dc.fr/

[9] Frederic Dufaux; Yafei Xing; Beatrice Pesquet-Popescu; Peter Schelkens, "Compression of digital holographic data: an overview", Proc. SPIE. 9599, Applications of Digital Image Processing XXXVIII, 2015.

[10] 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

[11] 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

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

[13] 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

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

[15] M.K. Kim, Digital Holographic Microscopy: Principles, Techniques, and Applications, Springer, 2011.

[16] Y. Park, G. Popescu, K. Badizadegan, R. R. Dasari, and M. S. Feld, Opt Express **14**, 8263, 2006.