EFFICIENT PLENOPTIC IMAGING REPRESENTATION: WHY DO WE NEED IT ?

Fernando Pereira[1], Eduardo A.B. da Silva[2]

[1] Instituto Superior Técnico – Instituto de Telecomunicações, Portugal
[2] PEE/COPPE, Universidade Federal do Rio de Janeiro, Brazil
emails: eduardo@smt.ufrj.br, fp@lx.it.pt

ABSTRACT

The 3D representation of the world visual information has been a challenge for a long time both in the analogue and digital domains. At least in the past decade, 3D stereo-based solutions have become very common. However, several constraints and limitations ended up causing a negative impact on its user popularity and market deployment. Recent developments in terms of acquisition and display devices have shown that it is possible to offer more immersive and powerful 3D experiences by adopting higher dimensional representations. In this context, the so-called plenoptic function offers an excellent framework to analyze and discuss the recent and future developments towards improved 3D imaging representations, functionalities and applications. Since they are associated to huge amounts of data, the new imaging modalities such as light fields and point clouds critically ask for appropriate efficient coding solutions. In this context, the main objective of this paper is to present, organize and discuss the recent trends and future developments on 3D visual data representation in a plenoptic function framework. This is critical to effectively plan the next research and standardization steps on 3D imaging representation and coding.

Index Terms— Plenoptic function, light fields, multiview video, depth, point clouds, meshes

1. INTRODUCTION

Light plays a vital role in our daily lives while communicating with the world around us. But while the world is made of objects, these objects do not communicate their properties directly to an observer; they rather fill the space around them with a pattern of light rays that is perceived and interpreted by the human visual system. Such a pattern of light rays can be measured, yielding the now ubiquitous images and videos. Visual information plays an increasing role in our lives and evolution and it is believed that up to 50% of the human brain is involved in some way in processing visual information.

Over the past decades, tremendous progress has been achieved in the way consumers and professionals capture, represent, code, store, distribute, display and ultimately use images and video. This motivated an ever-growing acceleration in the creation and usage of image and video in a multitude of sectors, applications, products and services providing the users adapted, powerful experiences. To make these experiences possible, visual information has to be acquired and represented according to some models and formats which allow replicating the light patterns that constitute the visual world. This process is simultaneously driven and conditioned by the available sensors, transmission and storage channels, displays, and the human visual system. This way the relevant set of functionalities can be offered for each target application and service in an efficient, effective, immersive, adaptive, and low-complexity way, always targeting the best user (visual) experience.

Over the years, images and video have been traditionally acquired, represented, coded and displayed using a rather simple model corresponding to rectangular sets of samples for some wavelength components, e.g. RGB or luminance and chrominances. The need to provide more immersive experiences has continuously raised the spatial resolution and more recently also the frame rate and sample depth. The large amount of data associated with this representation model if a raw, non-compressed representation is adopted, as well as the limited available bandwidth and storage capacities, have boosted the need for efficient coding solutions. Due to the extreme importance of interoperability for the most important applications, video coding standards are paramount. They have been changing the vision/visual-based applications landscape with an incredible explosion in numbers and variety of applications in the past decades. In practice, several generations of video coding standards have been developed with H.264/AVC (Advanced Video Coding) and HEVC (High Efficiency Video Coding) being the most recent developments [1][2] and with largest deployment.

In recent years, 3D experiences have become more popular, acknowledging that a faithful, transparent and immersive representation of the world requires more than 2D video. This has asked for 3D video acquisition, representation and coding models, notably stereo pairs to provide depth impressions (stereo parallax) and multiview video to offer some immersive 3D navigation capabilities (both stereo and motion parallaxes). With the increase on the number of views, efficient coding is even more important and thus there are currently several standard coding solutions for multiview video (MVV). The most recent multiview video coding standards are based on two different representation paradigms, which are *only texture* and *texture and depth*. These gave rise to the MV-HEVC and 3D-HEVC [3] coding standards, respectively, which build on the HEVC standard in a backward compatible way. The 3D-HEVC standard is clearly the most advanced as it finally exploits the inter-component depth and texture redundancies never exploited before. It also adopts a DBIR (depth based image rendering) approach where the receiver not only decodes the received views but may also synthesize many other views to provide a smooth parallax. However, although the most powerful, the 3D-HEVC standard only addresses linear and horizontal-only parallax camera arrangements, narrow baselines, i.e. short camera distances, and reduced viewing ranges.

Recently, new sensors and displays have started emerging which go beyond the traditional representation model to provide more powerful, adaptive and immersive visual experiences, e.g. *Lytro*, *Light* and *Raytrix* cameras [4][5][6] and *Holografika* displays [7]. These unconventional developments led to revisiting the fundamentals of the vision process, notably the structure of the information in the light impinging on an observer of a scene, in a deep quest to find more powerful and complete visual representation models. In this context, it is only natural to consider the so-called *plenoptic function* which represents the intensity of light seen from any viewpoint or 3D spatial position, any angular viewing direction, over time, and for each wavelength [8][9]. It thus provides a simple although powerful 7D representation model of light where all the old and new representation models can fit depending on the assumed constraints. While new and denser samplings of the plenoptic function as performed by the emerging cameras may provide richer representations, and thus additional user capabilities, they also imply more visual data. Thus, plenoptic imaging needs appropriate representation models with associated efficient coding solutions. In summary, and directly answering to the question raised in the title of this paper, efficient plenoptic imaging representations are needed because the associated cameras and displays are emerging in the market at an increasing speed. In addition, the richer representations of the light around us that they acquire to provide more powerful capabilities, imply very large amounts of data, critically asking for efficient coding. Naturally, for a new generation of imaging technology to arise in practice, with an exploding number of users, standards and interoperability will play again a major role. This is confirmed by the emerging related standardization initiatives [10][11].

In this context, the main objective of this paper is not to provide technical novelty in terms of new algorithms but rather to discuss the recent trends on visual data representation and understand them in the context of the plenoptic function. With this purpose in mind, the next sections define and discuss the plenoptic function and present some important issues in terms of acquisition, representation, and coding. The paper ends with a reference to the relevant standardization initiatives and some final remarks.

2. PLENOPTIC FUNCTION DEFINITION

As stated above, more immersive visual experiences demand richer and more complete representation paradigms for the visual information. Ultimately, all the visual information present in the world is given by the full description of the electromagnetic field in the visible spectrum for every point in space. At first, providing such a complete description of this electromagnetic field may seem too complicated. However, some useful models may be obtained with some simplifying assumptions:

- By restricting the description to non-coherent light, it is possible to use the Fourier representation to describe a propagating wave as an infinite sum of random-phase sinusoids within the visible spectrum, each with a different energy. Mathematically, this is represented as a power density for each wavelength λ.
- The electromagnetic wave at a point (x,y,z) in space can be decomposed by a sum of wavefronts coming from all directions. Each direction can be described by an azimuth and orientation pair (θ,Φ). Therefore, for a given wavelength, the power spectral density for each point and each propagation direction is given by a function $P(x,y,z,\theta,\Phi,\lambda)$.
- If the energy of each electromagnetic wave varies in time (at a rate much smaller than the period of the sinusoidal electromagnetic waves), one can add a temporal dimension, finally yielding a function $P(x,y,z,\theta,\Phi,\lambda,t)$.

Thus, provided that the light is incoherent, all the visual information in the world can be represented by the 7D scalar function $P(x,y,z,\theta,\Phi,\lambda,t)$ usually referred to as the *plenoptic function* [8] illustrated in Fig. 1. However, although the plenoptic function is conceptually simple, its 7 dimensions imply that an enormous amount of data is associated to its representation. Therefore, in practical situations it is essential to reduce its dimensionality and appropriately sample it. One usually starts by not using the full wavelength information. For the visible spectrum, the wavelength dimension is removed and instead three plenoptic functions are used, e.g. one for each R, G and B color channel. Another sampling and dimensionality reduction example is given by multiview video. There, the camera position (x,y,z) is usually replaced by an index k in an one-dimensional array of monocular cameras. This reduces the plenoptic function to $P(k,u,v,t)$, where the position (u,v) on each camera plane is equivalent to the ray orientation (θ,Φ). It is important to note that the sampling of the plenoptic function dimensions has to be such that aliasing is avoided. There are many works that theoretically study the spectrum and bandwidth of the plenoptic function [12][13].

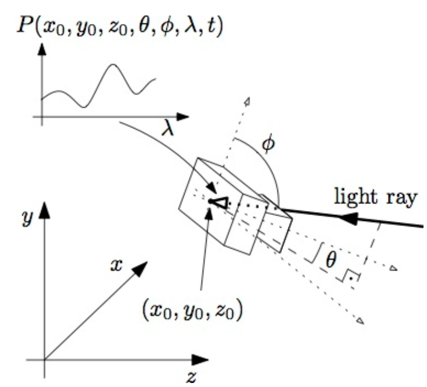

Fig. 1. Visualizing the plenoptic function.

Such high dimensionality representation enables the plenoptic function to provide many interesting functionalities. Once sampled and recorded, this rich information can be processed to render images in multiple, desired viewing conditions. For example, light rays can be combined *a posteriori* to generate an image focused at any depth [14]. It is possible to select objects by their depth using *n*-dimensional digital filtering [15][16]. Virtual camera movement can be easily produced, and the rich information in the plenoptic representation can be used to perform tasks such as highly realistic relighting [17]. Also, by dropping all but the spatial position (x,y,z) a binarized plenoptic function $P(x,y,z)$ may yield the 3D structure of an object.

The processing chain used in plenoptic imaging can be summarized as in Fig. 2. There, one first acquires some version of the plenoptic function with reduced dimensionality using some acquisition model closely related to the selected sensors (naturally, some metadata may be also acquired or inserted). After acquisition, this information may be converted to a (raw data) representation format as the acquisition and representation formats do not have to be the same. For example, to acquire the plenoptic function one can use a LIDAR laser scanner, recording photon counts and time intervals [18], but the representation model may be point cloud based, thus consisting of spatial coordinates and maybe some optical properties. In all cases, since the amount of data generated is huge, one must compress the raw data generated for storage or transmission. In the next sections, the plenoptic imaging acquisition, representation and coding stages will be addressed. Although extremely important in plenoptic imaging, the rendering stage will not be addressed due to space constraints. The interested reader is referred to [14][15] for some examples.

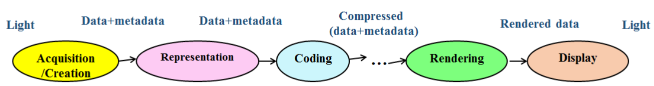

Fig. 2. Plenoptic imaging data flow.

3. PLENOPTIC FUNCTION ACQUISITION AND REPRESENTATION

While the 7D plenoptic function can fully express the light information available in a real scene, practical constraints have led to strongly reducing its sampling dimensionality. However, with the trend to increase visual immersion, less powerful dimensionality reductions become relevant. The sampling of the plenoptic function will involve measuring radiance information for specific wavelengths or bandwidths, positioned in a 5D space, with or without time. To perform this task, several types of sensors may be used, naturally in different ways and contexts:

Texture

- **Traditional cameras -** Measure color information represented as RGB or luminance and chrominance data, the so-called *texture*, for each view.
- **Arrays of traditional cameras - I**nclude a more or less dense array of cameras, with regular or irregular, linear or non-linear arrangements, with single (horizontal or vertical only) or double/full parallax.
- **Microlens array (aka light field) cameras -** Include a microlens array into the optical path of a monocular camera, thus providing directional data for each sampled *xy* position; in practice, these devices behave like multiple small cameras in a single box and may consider or not the temporal dimension.
- **360º (static or video) cameras -** Increase the field of view up to 360º in the horizontal plane, or even cover (approximately) the entire sphere.

Depth

- **Time of flight (ToF) range imaging cameras -** Measure the distance by means of the time-of-flight of a continuous light signal between the camera and the subject for each point of the image resulting into depth maps.
- **LIDAR cameras -** Measure the distance to a target by illuminating it with laser pulses and analyzing the reflected light again resulting into depth maps. ToF and LIDAR are typically appropriate for shorter (< 10 m) and longer distances, respectively.
- **Infrared cameras -** Capture the object distance using infrared radiation, typically in wavelengths as long as 14 µm. For example, the Kinect 2 works by projecting a grid of infrared light points onto a scene, then measuring how long it takes for the light from each of those points to reflect back to the camera's sensor.
- **Structured light projectors -** Capture depth by solving a correspondence problem between a reference image pattern and the captured pattern.

These sensors may directly measure the data for the final representation and coding format, e.g. array of views, or may cooperate to create this data, e.g. time-of-flight cameras may be combined to create point clouds. There are sometimes many processing steps between the sensors and the final scene representation data [19] which may take different forms:

1. **Regular Sampling** – These representation formats adopt regular sampling grids/densities, such as in conventional multiview video, which means than no explicit coding of the sample positions is needed (as it is implicit). Here a few cases may be distinguished:

- ***Multiview (only) video (aka super multiview video) -*** High density, wide range, array of conventional, monocular views with horizontal or both horizontal and vertical parallaxes, with linear or arc arrangement. For static objects, images may also be obtained with a sequence of frames taken with regular camera moving around an object or a sequence of frames taken with a fixed, regular camera when an object rotates around itself.
- ***Multiview video and depth -*** Same as above but including now depth data for at least some (if not all) of the views. As the depth may compensate for texture, the higher is the number of depth views, the lesser may be the number of textures views for the same final rendered quality. However, the depth accuracy is typically a problem to reach high quality synthetic views, especially for arbitrary camera arrangements and wide baselines.
- ***Microlens array (aka light field) imaging and video -*** Frame based model with a 2D structure of so-called *micro-images* or *sub-aperture images* each with a resolution corresponding to the number of directions for which radiance/color is measured. This is the output of a so-called *light field camera* and may consider or not the temporal dimension. Super multiview image/video and microlens array image/video are not that different in the sense that both result into a 2D array of views although with different resolutions, baselines, viewing ranges, complexities, costs, etc.; however, while one image of the multiview set may be displayed without much processing, this is not the case for the microlens images which need appropriate rendering. However, light field cameras are emerging at an incredible pace with increasing resolutions and increasingly creative ways to sample the plenoptic function, which give rise to corresponding new representation models.

2. **Irregular Sampling** – It is also possible to adopt a representation model without regular sampling which means that the sample positions may have to be explicitly coded, such as:

- ***Irregular array of conventional cameras -*** Output of an array of conventional cameras with both horizontal and vertical parallaxes with an irregular arrangement. This case may correspond to an irregular array of lens in a single camera thus in a 2D plane. Alternatively, it can correspond to an irregular array of independent cameras to provide some free navigation capabilities, e.g. the cameras around a stadium.
- ***Point cloud with texture -*** A point cloud is a set of data points in some coordinate system; in a 3D coordinate system, these points are usually defined by *xyz* coordinates, and often are intended to represent the external surface of an object. To represent the field of light this positioning information has to be complemented with texture or point cloud attributes like colors and normal directions. Point clouds need to be sufficiently dense to allow high quality rendering. In practice, most currently available point clouds have a single color allocated to each position assuming scenes with Lambertian lighting and thus without too much specular reflections. This may differentiate *point clouds* from so-called *light field* representations where direction-dependent color is typically available. In fact, no *a priori* color restrictions are made in light fields, but these may be present in point clouds, in that case restricting its representative power. A way to go around this restriction is the so-called Bidirectional Reflectance Distribution Function (BRDF) which determines for each 3D position how the light propagates the available Lambertian color in the different viewing directions.
- ***Mesh with texture -*** While point clouds can be directly rendered, it is also possible to convert them into a connected polygon mesh or triangle mesh model. In practice, 3D meshes add connectivity between the points of the point cloud, which then become the vertices of the mesh patches. The color for the mesh patches may be simply obtained by interpolating from the colors available for the vertices. Meshes are easier to edit and good for synthetic data and thus for applications like gaming; however, the complexity of their conversion may be problematic for real-time applications such as telepresence [20]. Also some views may be used to bring texture to the geometric data associated to the mesh [19].

These different representation models/formats suggest the following important questions: Is there a unique, universal representation format that may accommodate all the mentioned representation formats, eventually using some appropriate conversion? Or is there at least a very limited number of representation formats that have to be considered for coding purposes driven by the most relevant application scenarios and associated displays? These are some of the key 3D visual information representation questions to be addressed in the next years. At this stage, the appropriate representation format for which efficient coding should be provided seems to be application dependent and several approaches have been pursued. For example, while some contexts seem to privilege a multiview video and depth format, e.g. MPEG FTV (Moving Picture Experts Group, Free Viewpoint TV) for entertainment applications, other contexts seem to prefer a point cloud/mesh and texture format, e.g. virtual reality applications.

With these new sensors and representation formats, the data to be displayed is no longer the same as the data acquired (eventually after coding) as in traditional image and video (especially if a traditional 2D display is still used). For example, the light field images cannot be directly shown in a 2D display and thus a view has to be rendered from the rich data acquired; the same happens with point clouds and texture, and even for super multiview video when a synthesized view is considered.

4. PLENOPTIC IMAGING CODING

It is well known that coding standards play a paramount role in the explosion of the imaging applications and thus it is expectable that the same will happen with plenoptic imaging. Considering the representation formats listed in the previous section, several situations may happen, notably regarding the coding solutions:

1. **Appropriate standard coding solutions are already available** - This is the case for multiview video coding and multiview video plus depth coding if linear and horizontal-only parallax camera arrangements with narrow baselines and reduced viewing ranges are considered.
2. **Available coding standards may be used although not fully adapted** - This is the case of microlens array images which may be coded with any image codec even if the codecs may not be able to exploit the redundancy between the micro-images
3. **No standard coding solutions are available** - This is the case of point clouds for which no coding standard is available although there are some relevant coding solutions in the literature.
4. **No coding solutions are available at all** - This may be the case for point clouds with view-dependent texture for which there seem to be no coding solutions available at all.

To be more precise, the coding situation for each representation model/format previously mentioned is:

1. Regular Sampling

- ***Multiview (only) video (aka super multiview video) -*** The MVC (Multiview Video Coding), MV-HEVC and 3D-HEVC [3] standards may code this type of data although only for linear, horizontal camera arrangements, narrow baselines and reduced viewing ranges. These coding standards may be less efficient if less dense, non-linear camera arrangements are considered. This is the reason for MPEG to be calling for technological evidence regarding more efficient solutions for those conditions [10]. If more efficient coding solutions arise, MPEG may consider developing a new coding standard to better address those camera arrangement scenarios. This type of solution may target smooth parallax by synthesizing many views at the decoder using an epipolar-plane images (EPI)-based rendering solution [21] where the depth is generated at the decoder and does not have to be coded and transmitted.
- ***Multiview Video and Depth -*** The 3D-HEVC standard specifically addresses this type of representation format where depth data may be explicitly coded and transmitted in this case even exploiting the inter-component redundancy. It is assumed that both decoded and synthesized views will be available at the decoder. Again, this standard assumes linear, horizontal camera arrangements, narrow baselines and reduced viewing ranges. This type of solution targets smooth parallax by synthesizing many views at the decoder mainly using depth-based image rendering (DBIR) where the depth is coded and transmitted.

- ***Microlens array (aka light field) imaging and video* -** This is a rather interesting case as still luminance and chrominance rectangular arranged data has to be coded although there is now a quasi-periodic structure with high redundancy within each image associated to the micro-images; naturally, this redundancy may be exploited. It is in principle possible to code light fields with any available image and video standard coding solution, e.g. Lytro Illum-like light fields [4]. However, these coding solutions will not be able to exploit the redundancy between the many micro-images within each light field image. For some of the light field cameras in the market, several specific coding solutions have been developed to increase the compression efficiency of standard solutions, notably HEVC [22][23][24][25][26][27][28].

2. Irregular Sampling

- ***Irregular array of conventional cameras* -** Considering that any type of camera arrangement is possible, the simplest coding solution for this type of data is the independent coding of each camera view, e.g. using HEVC. However, this should be an inefficient solution as typically there is some degree of redundancy between the various irregularly arranged cameras/views. While the 3D-HEVC standard may be also used, this should not be an efficient solution as this codec is well prepared only to exploit the inter-view redundancy for specific types of camera arrangements as mentioned above. MPEG is currently calling for technological evidence of more efficient coding solutions for cameras with arbitrary positioning and wide baselines [10].
- ***Point Cloud with Texture* -** While there are some point cloud coding solutions in the literature, there is no standard available for point cloud coding. Both MPEG and JPEG (Joint Photographic Experts Group) are currently working on this topic although at rather early stages. The most popular point cloud encoder is the one available at the Point Cloud Library (PCL) [29], many times used as benchmark in the literature. While textures may be coded with available coding standards, point cloud attributes may need specific coding solutions such as the solution in [20]. However, to fully model the plenoptic function data in an alternative way to the light field representation models, each point cloud position must have associate color values for all directions; the authors could not find any point cloud coding solution fulfilling this requirement.
- ***Mesh with Texture* -** There are many solutions available for the coding of meshes, especially triangular, both static and dynamic [30]. There are also some standard coding solutions available, notably in MPEG-4 Part 2, Visual, with 3D Mesh Coding (3DMC) for the compression of generic, static meshes, and MPEG-4 Part 16, Animation Framework eXtension (AFX). For example, the FAMC (Frame-based Animation Mesh Compression) tool codes animated meshes on a time basis considering the attributes (positions, normal vectors, etc.) of the vertices composing a mesh. However, as for the point clouds, there are no coding solutions available which also consider view-dependent color data.

The situation described above clearly shows that while visual data is becoming available in many configurations, targeting a richer sampling of the plenoptic function, efficient coding is lagging behind, especially regarding standard coding solutions which are critical for the markets to grow. This highlights again the need to develop efficient plenoptic imaging coding solutions, notably standard solutions, thus explaining the reason for the growing number of standardization initiatives in this area as will be shown in Section 5.

5. PLENOPTIC IMAGING CODING STANDARDIZATION INITIATIVES

It is well known that coding standards play a critical role in the explosion of visual data applications and the creation of big markets. Regarding plenoptic imaging, MPEG and JPEG are the two standardization bodies which seem to be willing to become protagonists in terms of coding solutions.

5.1 MPEG Free Viewpoint TV

For a long time MPEG has been developing 3D video coding standards starting with a specific profile in MPEG-2 Video in the nineties. The most recent and powerful standard developed, 3D-HEVC [3], was defined to operate in rather specific conditions, notably linear and horizontal-only parallax camera arrangements, narrow baselines and reduced viewing ranges. This provides only limited user experience, and it is one of the reasons why 3D video is clearly not exploding in the market. As new acquisition and display solutions are quickly emerging, MPEG decided in early 2015 to further explore new technologies by issuing a Call for Evidence to assess the availability of more powerful representation technologies in the areas of *super multiview video* (SMV) and *free navigation* (FN). For SMV, the objective is to substantially reduce the rate required to reconstruct the full set of input views at the receiver compared to 3D-HEVC, the current state-of-the-art on 3D video compression. For FN, the objective is to substantially improve the rendering quality at arbitrary virtual view positions in 3D space; this may be achieved through an alternative (to 3D-HEVC) representation format, in which case compression efficiency must also be considered. Keeping its tradition, there is no intention to standardize any post-processing tools. While SMV clearly aims at higher compression efficiency exploiting at best the information in all camera views, improved view synthesis is an additional cornerstone for FN in large baseline camera arrangements. To better understand the technological landscape in these domains, MPEG is asking that, by February 2016 [10], companies with coding technologies performing better than 3D-HEVC bring such information to MPEG, with evidence taken under well-defined test conditions. If the collected technology significantly outperforms currently available MPEG technology, a Call for Proposals (CfP) will be issued targeting the development of standards allowing increased compression performance beyond 3D-HEVC.

5.2 JPEG PLENO

JPEG has decided in October 2014 to launch a so-called JPEG PLENO innovation activity. This acknowledges that new imaging sensors already in the market are enabling new capturing and visualization capabilities that will finally result in a paradigm shift in the production and consumption of digital photographic material. JPEG PLENO targets three imaging modalities: light field, point-cloud, and holographic data [11]. While light field and point cloud data are direct representations of the plenoptic function, holographic data samples the interference patterns between a reference wave and an object wave corresponding to the reference wave diffracted by the scene; such data may be physically created or computer generated. Although these are all representations of the visible information in the world, the current technical limitations associated to the creation and visualization of holograms seem to indicate that holographic data may be a lower priority in JPEG PLENO. Its emerging practical applications tend to be limited to holographic microscopy where the relevant objects are not very deep and a large look-around effect is not required. JPEG PLENO has been stimulating significant positive reactions but the precise standardization schedule has still not been defined. While new functionalities are a major target, JPEG PLENO still considers some backward compatibility requirements with JPEG's legacy formats to create a powerful JPEG ecosystem.

5.3 Joint JPEG and MPEG Activities on Plenoptic Imaging

Acknowledging that other plenoptic representation exploration works were being developed within MPEG, such as point cloud coding, MPEG decided in October 2015 to create an Ad Hoc Group (AhG) on Light Fields. This AhG showed that MPEG had a clear intent to revise its global positioning regarding 3D visual data with a larger breadth, notably beyond the previous MPEG FTV scope. Since in February 2016 both JPEG and MPEG had ongoing activities related to plenoptic function representations, it was decided to try joining efforts in a so-called Joint AhG for Digital Representations of Light/Sound Fields for Immersive Media Applications. The addition of 'sound' to the scope of this AhG is clearly an MPEG endeavor as sound wave fields are not within JPEG scope. However, it shows that the plenoptic function may be applied both to light and sound fields. The AhG mandates hint that the plenoptic representation joint activity should consider many areas, notably uses cases, functionalities, representation models, sensor and display technologies, performance assessment, etc. It is expected that at some stage JPEG and MPEG exploration activities may identify precise industry needs in terms of uses cases and requirements, thus finally launching the usual technology specification process.

6. FINAL REMARKS

This paper has presented, organized and discussed the recent trends and future developments on 3D visual data representation using a plenoptic function framework. It is expected that this type of analysis contributes to better understanding the current *status quo* and allows better planning the future development of powerful representation model(s) for 3D visual information.

7. REFERENCES

[1] T. Wiegand *et al*., "Overview of the H.264/AVC Video Coding Standard", IEEE T-CSVT, vol. 13, no.7, pp. 560-576, Jul. 2003.

[2] G. J. Sullivan *et al*., "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE T-CSVT, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[3] G. Tech, *et al*., "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding", IEEE T-CSVT, vol. 26, no. 1, pp. 35-49, Jan. 2016.

[4] Lytro page, https://www.lytro.com/

[5] Light page, https://light.co/camera

[6] Raytrix page, http://www.raytrix.de/

[7] Holografika page, http://www.holografika.com/

[8] E. H. Adelson, and J. R. Bergen, "The Plenoptic Function and the Elements of Early Vision, The MIT Press, Cambridge, Mass., 1991.

[9] S. J. Gortler *et al*., "The Lumigraph", in Computer Graphics Proceedings, pp. 43–54, Proc. ACM SIGGRAPH'96, New Orleans, USA, Aug. 1996.

[10] MPEG, Call for Evidence on FTV, Doc. MPEG2015/N15095, Warsaw, Poland, Jun. 2015.

[11] JPEG, JPEG PLENO Abstract and Executive Summary, Doc. WG1 N6922, Sydney, Australia, Feb. 2015.

[12] M. N. Do, D. M-Maillet, M. Vetterli, "On the Bandwidth of the Plenoptic Function", IEEE T-IP, vol. 21, no. 2, pp. 708-717, Feb. 2012.

[13] C. Gillian, *et al*., "On the Spectrum of the Plenoptic Function", IEEE TIP, vol. 23, no. 2, pp. 502-516, Feb. 2014.

[14] R. Ng., Digital Light Field Photography, Ph.D. Thesis, Stanford University, Jul. 2006.

[15] D. G. Dansereau, Plenoptic Signal Processing for Robust Vision in Field Robotics, Ph.D. Thesis, The University of Sydney, Australia, Jan. 2014.

[16] C. U. S. Edussooriya *et al.,* "Five-Dimensional Depth-Velocity Filtering for Enhancing Moving Objects in Light Field Videos", IEEE T-SP, vol. 63, no. 8, pp. 2151-2163, Apr. 2015.

[17] V. Masselus *et al*., "Relighting with 4D Incident Light Fields", ACM Trans. on Graphics, vol. 22, no. 3, pp. 613-620, Jul. 2003.

[18] LIDAR, http://pcl.physics.uwo.ca/science/lidarintro/

[19] A. Collet, *et al*., "High-Quality Streamable Free-Viewpoint Video", ACM Trans. on Graphics, vol. 34, no. 4, Aug. 2015.

[20] C. Zhang *et al*., "Point Cloud Attribute Compression with Graph Transform, IEEE ICIP, Paris, France, Oct. 2014.

[21] C. Brites *et al*., "Epipolar Plane Image based Rendering for 3D Video Coding", IEEE MMSP, Xiamen, China, Oct. 2015.

[22] M. Magnor, B. Girod, "Data Compression for Light-Field Rendering", IEEE T-CSVT, vol. 10, no. 3, pp. 338-343, March 2000.

[23] S. C. Chan *et al*., "The Data Compression of Simplified Dynamic Light Fields", ICASSP'03, Hong Kong, Apr. 2003.

[24] C. L. Chang *et al.,* "Light Field Compression Using Disparity-Compensated Lifting and Shape Adaptation" IEEE T-IP, vol. 15, no. 4, pp. 793–806, Apr. 2006.

[25] S. Kundu, "Light Field Compression using Homography and 2D Warping", ICASSP'12, Kyoto, Japan, Mar. 2012.

[26] A. Aggoun *et al.,* "Immersive 3D Holoscopic Video System", IEEE Multimedia, vol. 20, no. 1, pp. 28–37, Jan.-Mar. 2013.

[27] C. Conti *et al.,* "New HEVC Prediction Modes for 3D Holoscopic Video Coding", IEEE ICIP 2012, Orlando, USA, Oct. 2012.

[28] C. Conti *et al.,* "Inter-Layer Prediction Scheme for Scalable 3-D Holoscopic Video Coding", IEEE SPL, vol. 20, no. 8, pp. 819-822, Aug. 2013.

[29] Point Cloud Library (PCL), http://pointclouds.org/

[30] J. Peng *et al*., "Technologies for 3D Mesh Compression: a Survey", Journal of Visual Comm. and Image Representation, vol. 16, no. 6, pp. 688–733, Dec. 2005.