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

Coding of Still Pictures

JBIG
Joint Bi-level Image Experts Group

JPEG
Joint Photographic Experts Group

TITLE: JPEG Pleno Use Cases and Requirements for Light Field Quality Assessment v1.0

SOURCE: Carla Pagliari (Instituto Militar de Engenharia-IME)
Eduardo A. B. da Silva (Universidade Federal do Rio de Janeiro-UFRJ)
Peter Schelkens and Saeed Mahmoudpour (Vrije Universiteit Brussel-VUB and imec)

PROJECT: JPEG PLENO

STATUS:

REQUESTED

ACTION:

DISTRIBUTION:

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
1. Goal

In the context of the JPEG Pleno standardization process with respect to light field coding (Part 2) [1], various subjective visual quality assessment procedures have been defined and significant knowledge has been built-up with respect to challenges, good practice guidelines and methodological aspects.

This document aims at providing an overview of the current state-of-the-art in subjective and objective quality assessment for light field modalities, to identify key test data, to investigate rendering procedures, to extract guidelines and requirements for quality evaluation procedures, and to identify potential use cases, with as final goal to launch a new standardization effort focused on light field quality assessment, but in future also other plenoptic modalities such as point clouds and holography.

Light field data (aka plenoptic data) records the amount of light (the “radiance”) at every point in space, in every direction (Figure 1). 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 contributes 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. Light field or plenoptic image and its micro-images [2].](image)

During the acquisition/generation of light fields two important properties are defined: the angular (given by the baseline) and spatial resolution. The former defines the maximum distance between the change of perspective within a light field, while the latter affects visual realism.

As critical perceptual cues associated with realistic 3D perception are not provided by existing 2D displays they fail to provide a full immersive experience. In contrast, 3D displays would be able to replicate the rays of light (light field), including the directional and color components.

Therefore, acquisition/generation and visualization are key components to meet the functionalities given various use case contexts.
2. Acquisition

Light field data can be computer generated or captured by plenoptic cameras (even by an array of plenoptic cameras) or by monocular camera arrays. Each method will deal with a spatio-angular trade-off generating denser or sparser light fields. In plenoptic cameras a single image sensor is shared to capture both spatial and angular information. The camera needs a sparser sampling of the angular dimensions of the light field in order to achieve better spatial resolution, and vice-versa to have a denser sampling. Camera arrays can be arranged in planar and non-planar positions. The former having wider baselines (due to physical camera distancing) and the latter with large overlaps in camera poses (shorter baselines). One can categorize the light field cameras as:

- Plenoptic cameras;
- Camera arrays:
  - planar (sparse due to physical camera distancing)
  - non-planar
- Gantry (high sampling density and high spatial resolution);
- Synthetic.

Table 1 lists selected light field cameras and acquisition systems, available on the market.

<table>
<thead>
<tr>
<th>LF plenoptic cameras</th>
<th>LF acquisition systems</th>
<th>LF camera prototypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google 16×GoPro rig (dense panoramic light field)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google 2×DSLR rig (dense panoramic light field)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google 46-camera [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="https://home.otoy.com/capture/lightstage/">https://home.otoy.com/capture/lightstage/</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 illustrates the field of view (FOV) and the depth of field of a light field camera. Figures 3 and 4 display different light field cameras, available on the market. Figure 4 displays cameras designed to specific use cases.
Figure 2. Light field camera FOV and DOF illustration [5].

Figure 3. Light field camera from RayShaper (https://www.rayshaper.ch/).

Figure 4. Light field cameras from Raytrix (https://raytrix.de/products/).
3. Visualization

All the light that one sees is photonic energy. Light coming from a light source bounces an object and enters your eyes. The continuous streaming of photonic energy that comes from a light source bounces off an object and hits your eye (Figure 5) [6]. These reflected individual light rays represent a point on an object. Only when receiving a collection of reflected light rays coming from different directions one will be able to perceive the object. The challenge of LF displays is to convey/recreate the light rays that come from the scene to the viewer’s eyes.

So, how to reproduce these directional light rays as accurately as possible? Pixels in 2D displays convey the same color light in all directions, despite the viewer's point of view. In contrast, 3D (holographic) displays convey unique (or quite) color rays in each direction in a bundle of rays (3D pixels/hogel - hologram element). In order to reproduce these rays, the display needs acuity (spatial resolution), a number of possible viewing angles of the light rays and their associated depth. LF displays cannot reproduce an infinite number of light rays, while having to be able to address the vergence-accommodation conflict.

![Figure 5. Plenoptic function [6].](image)

3.1. Displays' limitations

All light field displays can emit a limited number of rays, thus dictating the Field of View (FOV), the spatial, the angular and the depth resolution. Considering the two-plane parametrization [7], the spatial resolution is given by the $u,v$ dimensions (2D image size) while the angular is given by the $s,t$ dimensions (2D array view spacing: horizontal and vertical). The vertical and horizontal spacing between adjacent views are the view’s baselines. For any display system the baseline is the distance between the extremes of the Field of View (FOV) in the observer display distance. If the acquisition system is a 2D planar array of monocular cameras, or a single or an array of lenslet-based cameras, the horizontal and vertical FOV will be limited to 180° in both horizontal and vertical directions, respectively. In a camera, the FOV is the maximum area of a sample that a camera can image and it is related to the focal length of the lens and the sensor size.

The display resolution and FOV will establish a volume where the scene objects are exhibited. So, the observer will have to be at a ‘valid’ point of view to ‘see’ the object. The visualization techniques and output

---

1 hogel holographic element; a piece of hologram that has homogeneous diffraction properties and is small enough to appear as a single point to the viewer [Lucent1994]
devices may allow extended depth of field, refocusing, and 3D views. Horizontal parallax only (HPO) and vertical parallax only (VPO) displays limit the practical application of 3D display technology and the assessment of the discomfort issue.

Table 2 presents the display’s characteristics relevance (high, medium, low, NA- not applicable) for quality assessment procedures.

**Table 2 - Visualization devices – display’s characteristics relevance for quality assessment.**

<table>
<thead>
<tr>
<th>Features</th>
<th>Display type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>light field</td>
</tr>
<tr>
<td>Comfort/wearability</td>
<td>NA</td>
</tr>
<tr>
<td>Comfort/noise</td>
<td>medium</td>
</tr>
<tr>
<td>Immersion/interaction</td>
<td>high</td>
</tr>
<tr>
<td>Immersion/FOV</td>
<td>high</td>
</tr>
<tr>
<td>Immersion/resolution</td>
<td>high</td>
</tr>
<tr>
<td>Multiple viewers</td>
<td>high</td>
</tr>
<tr>
<td>Eye-tracking</td>
<td>NA</td>
</tr>
<tr>
<td>Full-parallax (FP)</td>
<td>high</td>
</tr>
<tr>
<td>Horizontal-parallax only (HPO)</td>
<td>high</td>
</tr>
<tr>
<td>Vertical-parallax only (VPO)</td>
<td>high</td>
</tr>
<tr>
<td>Vergence-accommodation conflict</td>
<td>low</td>
</tr>
<tr>
<td>Depth perception</td>
<td>high</td>
</tr>
</tbody>
</table>

Table 3 lists selected light field displays and near-eye displays, available on the market.
Table 3- Light field displays

<table>
<thead>
<tr>
<th>Head-set/near-eye display</th>
<th>LF display</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="https://blog.google/technology/research/project-starline/">https://blog.google/technology/research/project-starline/</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://www.lightfieldlab.com/">https://www.lightfieldlab.com/</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://www.holoxica.com/">https://www.holoxica.com/</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://holografika.com/">https://holografika.com/</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://mopic3d.com/">http://mopic3d.com/</a></td>
</tr>
</tbody>
</table>

Figures 6, 7, and 8 show examples of light field displays.

**Figure 6.** The 3D light field tablet (LUME PAD - [https://www.leiainc.com](https://www.leiainc.com)).

**Figure 7.** Light field display (Looking Glass 8k - [https://lookingglassfactory.com/8k](https://lookingglassfactory.com/8k)).
4. Use cases and functionalities

The use cases refer primarily to static light fields, corresponding to a single time sample, where spatial and angular information of a tri-dimensional scene is simultaneously captured. However, light field video use cases may be addressed in the future. Light field technology is another step in the evolution of visual consumption and light field's representations, associated use cases and functionalities are detailed in this section.

4.1. Immersive user experiences

Immersive light field data is going to play an important role in future VR, AR and MR platforms. Light field content reproduces the parallax one sees in the real world, providing an immersive experience. Depth perception offered by systems which employ at least two views of a scene could bring the real world to many applications. Accurate and reliable 3D reconstructions are provided by light field data, an alternative high-performance imaging system to stereo and multi-view systems.

The key difference between an AR and a VR device is that while the former has to superimpose/combine a virtual content onto the real world, the latter has not.

There are several real-life applications of VR, such as education, real estate, healthcare, marketing, and travel. Immersive VR devices, such as PC-connected headsets (HTC Vive [8], Samsung Odyssey+ [9], PlayStation VR [10], Oculus Quest [11], and Oculus Rift [11]) and standalone headsets (Samsung Gear VR [9], Google Daydream [12], and Google Cardboard [12]) aim to bring realistic visual experience. Users create VR experiences by moving their heads so that their view perspective is shown by the headset. Figure 9 shows a child wearing an HTC Vive [8] headset.
Many factors play an important role when creating a realistic visual experience. To bring a realistic immersive experience to a user a system may be able to deliver high resolution images with large FOV, refreshing at fast frame rate. Head mounted displays, or wearable displays, or near-eye displays create a virtual image in the field of view (FOV) of one or both eyes, integrating one or multiple image sources in the device. One challenge to be tackled is the vergence-accommodation conflict, that causes visual fatigue, in near-eye displays. One solution is that the light field display design incorporates light field optics with the eye models, thus mitigating the vergence-accommodation conflict [13]. The wearable device displays a light field directly in front of the viewer’s eyes, who can focus on the various depths of a scene.

One can list advertising, gaming, maintenance, manufacturing, navigation, and retail as potential applications. People can use portable devices, such as smartphones or tablets, to run AR applications, without the need of specific devices. Pokemon Go is an example of an AR mobile game. In contrast, one can use special smart AR glasses or headsets to create AR experiences, by adding digital objects to the real world.

Light field data compression state-of-the-art technology will allow multiple viewers experience and not only single viewer experience with standalone devices.
4.2. 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.

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 analyzing 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.
- 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.

Figure 10 illustrates 3D plant analysis and metrology using light fields.

![Figure 10](image-url)

**Figure 10.** 3D plant analysis and metrology using light fields using Raytrix cameras [14].
4.3. Medical Imaging

The possibility to change the plane of focus of a captured image allows the viewer/surgeon to clearly see the anatomy, regardless of the plane in which the object is located. It also allows the viewer to interact with the light field image changing the focus plane and the viewing angle, for example. It provides accurate imaging of multi-plane anatomy, which is an invaluable asset to surgery or analysis. Figure 11 illustrates medical imaging using light field cameras.

![Medical image](image1)

**Figure 11.** Medical image [5]

4.4. Light field telepresence

The “sense of presence” is the ability to experience the virtual presence of another person. Light field technology provides smooth motion parallax without the need of glasses. Such a telepresence system also may need to replicate audio, while providing lifelike sizes to visually achieve a sense of presence. This system may use Horizontal Parallax-Only (HPO) LF displays as viewers tend to mostly move horizontally. Critical requirements are the system latency, the angular (view density) and spatial resolutions, the frame rate, and the true-to-scale (true-to-life) visualization. Figure 12 shows two “telepresence” light field displays.

(a) Holografika display [15]  
(b) Google’s Starline Project [16]

**Figure 12.** Telepresence light field displays.
5. Light Field Quality Assessment Solutions/Schemes

Quality assessment of light fields remains a complex task due to the diversity of light field acquisition procedures, distortion types, and rendering techniques. This section summarizes the existing subjective and objective quality assessment solutions for light fields with the aim to identify the existing limitations and to extract the requirements for quality assessment procedures. The quality assessment aspects in terms of the acquisition, assessment, rendering, and distortion types are summarized in Table 4.

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Plenoptic camera</th>
<th>Planar monocular camera array</th>
<th>Non-planar monocular camera array</th>
<th>Gantry</th>
<th>Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datasets (acquisition/generation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>Full-reference (FR) (Objective QA)</td>
<td>Reduced-reference (RR) (Objective QA)</td>
<td>No-reference (NR) (Objective QA)</td>
<td>Single Stimulus (SS) (Subjective QA)</td>
<td>Double Stimulus (DS) (Subjective QA)</td>
</tr>
<tr>
<td>Rendering</td>
<td>Views assessed as independent images</td>
<td>Refocused views with varying focal lengths</td>
<td>Pseudo-video assembled from views (scanning patterns: raster scan, serpentine, spiral, etc.)</td>
<td>Change of perspective (parallax) rendering using eye tracking</td>
<td>Type of display (2D, light field, near-eye)</td>
</tr>
<tr>
<td>Types of distortion assessed</td>
<td>Added noise</td>
<td>Added geometric distortion</td>
<td>Compression artifacts</td>
<td>Blurring</td>
<td>Reconstruction (view interpolation)</td>
</tr>
</tbody>
</table>

5.1. Subjective Quality assessment procedures

Existing subjective quality assessment solutions are constrained by lack of availability of high-end light field displays and limitations in acquisition of high-resolution 6DoF immersive content. Table 5 summarizes the subjectively-annotated light field datasets that are publicly available. The datasets are characterized by the stimulus type (double or single stimulus (DS/SS)), rating method, light field visualization method (passive or active), and display types used in the subjective experiment. Literature is briefly analyzed (Annex A) the impact of different visualization techniques for light field quality assessment.
Table 5: Subjectively-annotated light field datasets

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Stimulus Type</th>
<th>Rating method</th>
<th>Visualization</th>
<th>Display Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tamboli et al. (2016)</td>
<td>DS</td>
<td>Pair-wise</td>
<td>Passive</td>
<td>2D</td>
</tr>
<tr>
<td>SMART (2017)</td>
<td>SS</td>
<td></td>
<td>Active</td>
<td>3D</td>
</tr>
<tr>
<td>MPI-Lightfield (2017)</td>
<td></td>
<td></td>
<td>Light field</td>
<td></td>
</tr>
<tr>
<td>VALID (2018)</td>
<td></td>
<td></td>
<td>Light field</td>
<td></td>
</tr>
<tr>
<td>Win5-LID (2018)</td>
<td></td>
<td></td>
<td>Light field</td>
<td></td>
</tr>
<tr>
<td>LFDD (2020)</td>
<td></td>
<td></td>
<td>Light field</td>
<td></td>
</tr>
</tbody>
</table>

5.2. Objective quality assessment procedures

Objective quality assessment solutions can be divided into three categories based on the availability of the reference light field: Full-reference (FR), Reduced Reference (RR), and No reference (NR or blind). The suitability of objective (2D) image quality assessment (IQA) methods (PSNR, SSIM, etc) for light fields has been extensively analyzed in literature (Annex B). While quality evaluation of sub-aperture views using IQA methods can exploit spatial quality degradation, inter-view quality properties and angular quality degradation cannot be well captured. To this end, new objective quality assessment methods have been recently proposed to integrate both spatial and angular quality degradation aspects in the assessment. Table 6 summarizes the existing FR and NR light field quality assessment solutions.
Table 6: Objective quality assessment solutions for light fields.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Goal</th>
<th>Approach</th>
<th>Benchmarks Datasets</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tian et al. [23] (2020)</td>
<td>To design a Full-reference QA method for light fields</td>
<td>Multi-scale Gabor feature extraction</td>
<td>MPI-LF and SMART</td>
<td>Higher quality prediction accuracy</td>
</tr>
<tr>
<td>Meng et al. [25] (2020)</td>
<td>To design a Full-reference QA method based on the HVS and LF angular-spatial characteristics</td>
<td>A dual-fan filter is used to constrain the parallax range. The difference of Gaussian (DoG) operator was used for feature extraction.</td>
<td>VALID, SHU</td>
<td>Higher quality prediction accuracy</td>
</tr>
<tr>
<td>Zhou et al. [26] (2020)</td>
<td>To design a new No-reference (Blind) QA method for light fields based on tensor theory</td>
<td>Measures spatial-angular quality considering global naturalness, local frequency properties, and angular quality consistency. Trained a regression model.</td>
<td>MPI-LF, SMART, Win5-LID</td>
<td>Higher quality prediction accuracy</td>
</tr>
<tr>
<td>Shi et al. [27] (2019)</td>
<td>To design a new No-reference (Blind) QA method for light fields based on measuring angular consistency degradation</td>
<td>Extracting the global and local features from LFs. Features extracted based on Image naturalness, Gradient direction, and Weighted local binary pattern (LBP). Trained a regression model.</td>
<td>VALID, MPI-LF, SMART, Win5-LID</td>
<td>Higher quality prediction accuracy</td>
</tr>
</tbody>
</table>

Objective quality assessment challenges:
- spatial-temporal quality degradation, distortion types interactions, diagnostic and metrological quality
- impact of content types (resolution, baseline, specularity and transparency, dynamic range)
6. Dataset selection

Current JPEG Pleno dataset presents different scene geometries and spatio-view geometry diversity. It has been selected according to the criteria detailed in [28]. The current JPEG Pleno light field dataset [29] is composed by the following data:
- Lenslets: Bikes, Danger de Mort, Foutain&Vincent2 and Stone Pillars Outside;
- HDCA: Set2 2K sub, Tarot and Lab1; and
- Synthetic: Greek and Sideboard.

6.1. Current JPEG Pleno datasets

6.1.1. Lenslets
- Bikes, Danger de Mort, Foutain&Vincent2 and Stone Pillars Outside: 625×434 pixels in 15×15 views (13×13 “useful” views) [28]:
  - The light field images Bikes (I01) and Stone Pillars Outside (I04), belong to a category of the dataset named Urban, presenting a high level of spatial information and moderate depth of field. The thin objects (bicycle rims) are very challenging to disparity estimation algorithms, as are the reflections present in the Bikes scene. Both, Bikes and Stone Pillars Outside, have objects at different depths, with the latter presenting low to medium degrees of texture.
  - Danger de Mort (I02) (dataset) images are part of the category Grids that have grid patterns with close details with wide depth of field range.
  - Fountain& Vincent 2 (I09) is part of the category People, displaying one person and a fountain, which is very close to the camera presenting a high level of spatial complexity. The dataset presents low to high degrees of spatial information in different fields of depth.

6.1.2. HDCA
- Set2 2K sub: 1920×1080 pixels in 33×11 views [28]:
  - The HDCA Table Top I (Set 2) images present fine details, reflections and regular patterns, according to the JPEG Pleno Specifications for High-density Camera Array (HDCA) Data Sets document (at the time it was Set 2 and not Set 2 2K sub).
  - Table Top I images show several objects of different sizes placed at different depths, as well as a repetitive pattern tablecloth and a not so regular patterned curtain background at the farthest scene depth. Part of a table is located at the closest depth, while the bottles are responsible for the reflections and also for some of the fine details that are present in the scene. Moreover, there are houseplants presenting a high level of spatial detail in the low to medium contrast scene.
Laboratory 1: 1936×1288 pixels in 31×31 views [28].
  - The dataset PoznanLaboratory1 has a spatial dimension of 1936×1288 pixels in 31×31 views. It presents repetitive patterns, highly textured and low textured regions, occlusions (high perspective distortions) and depth variations.

Tarot: 1024×1024 pixels in 17×17 views [28].
  - The dataset Tarot Cards and Crystal Ball (small angular extent) has spatial dimension of 1024×1024 pixels in 17×17 views, presenting non-Lambertian surfaces, highly textured and low textured regions, occlusions and depth variations.

6.1.3. Synthetic

Synthetic: Greek and Sideboard: 512×512 pixels in 9×9 views [28].
  - Non Lambertian surfaces;
  - Highly textured and low textured regions; repetitive patterns;
  - Depth variations;
  - Occlusions;
  - Intensity level variation in the same view, as well as inter-views.

6.2. Additional datasets

Additional datasets with different types of LF acquisition might include:

- synthetic generated LF, lenslet and HDCA;
- different/complex scene geometries;
- different color gamut;
- different degrees of texture;
- different bit-depths;
- different parallax levels;
- different spatial resolution;
- different number of views;
- test data with specularity and transparency.

Bear in mind that other types of plenoptic cameras are being tested by MPEG-I dense LF (ISO/IEC JTC 1/SC 29/WG 04 N0056).

Table 7 summarizes potential additional datasets.
Table 7: Potential additional datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>year</th>
<th>Capturing device</th>
<th>#scenes and resolution</th>
<th>Purpose</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rerabek and T. Ebrahimi (EPFL)</td>
<td>2016</td>
<td>Lytro Illum (625 x 434 x 15 x 15)</td>
<td>118</td>
<td>General</td>
<td><a href="http://plenodb.jpeg.org/lf/epfl/caldata-B514309630.tar">http://plenodb.jpeg.org/lf/epfl/caldata-B514309630.tar</a></td>
</tr>
<tr>
<td>4D Light Field (HCI &amp; Konstanz University)</td>
<td>2016</td>
<td>8 bit LF (512x512x9x9) 16bit ground truth disparity maps in 2 resolutions (512x512 and 5120x5120px)</td>
<td>24 synthetic, densely sampled 4D LF with accurate disparity</td>
<td>Evaluation for Depth Estimation on 4D Light Fields</td>
<td><a href="https://lightfield-analysis.uni-konstanz.de/">https://lightfield-analysis.uni-konstanz.de/</a></td>
</tr>
<tr>
<td>Ziegler et al. (Fraunhofer IIS)</td>
<td>2017</td>
<td>Sony Alpha 7RII (7952x5304px) 21×99 / 21×101</td>
<td>9</td>
<td>Densely sampled light field images, wide baseline</td>
<td><a href="https://www.iis.fraunhofer.de/en/ff/amm/dl/lightfielddataset.html">https://www.iis.fraunhofer.de/en/ff/amm/dl/lightfielddataset.html</a></td>
</tr>
<tr>
<td>W. Ahmad et al. (CAU Germany)</td>
<td>2018</td>
<td>Lytro Illum, Ratyrix R29</td>
<td>31</td>
<td>Comparison of plenoptic 1.0 (Lytro Illum) and 2.0 (Raytrix R29)</td>
<td><a href="https://figshare.com/articles/data">https://figshare.com/articles/data</a> set/The_Plenoptic_Dataset/61154873</td>
</tr>
<tr>
<td>Light-Field Intrinsic Dataset (Fraunhofer IIS and etc)</td>
<td>2018</td>
<td>Canon EOS 6D and 5D as well as Sony Alpha 7 RII (4m horizontally and 0.5 m in vertical direction with a precision error of 80 mm.)</td>
<td>8 synthetic (dense 35x35 and sparse 11x11) 4 real (3D (151x1) and 4D)</td>
<td>light-field intrinsic decomposition Specularity, Albedo and Shading</td>
<td><a href="http://ffid.mpi-inf.mpg.de/">http://ffid.mpi-inf.mpg.de/</a></td>
</tr>
<tr>
<td>Stanford Multiview Light Field Dataset</td>
<td>2019</td>
<td>Single hand-held Lytro Illum 3 Lytro Illum cameras were rigidly mounted together</td>
<td>4211 LFs: (3-5 poses of 850 scenes) 3042 LFs:</td>
<td>Lambertian/non-Lambertian surfaces, occlusion, specularity, subsurface scattering,</td>
<td><a href="http://lightfields.stanford.edu/mvlf/">http://lightfields.stanford.edu/mvlf/</a></td>
</tr>
<tr>
<td>Zakeri et al. (Fraunhofer IIS)</td>
<td>2019</td>
<td>Sony alpha 7RII (Non-planar arm moves on a cylindrical path for a field of view of 125 degrees)</td>
<td>2 LFs: (5168x3448 with disparity less than 1.6 p) Mannequin: 9600x5. Sofa: 9600x3.</td>
<td>Non-planar Light field Representation</td>
<td><a href="https://fordatis.fraunhofer.de/handle/fordatis/147">https://fordatis.fraunhofer.de/handle/fordatis/147</a></td>
</tr>
<tr>
<td>TEDDY (Fraunhofer IIS)</td>
<td>2020</td>
<td>Sony alpha 7RII</td>
<td>6 LFs: (5168x3448 x11x11) (3984+2240x50x50)</td>
<td>A High-Resolution High Dynamic Range Light Field Dataset</td>
<td><a href="https://fordatis.fraunhofer.de/handle/fordatis/158">https://fordatis.fraunhofer.de/handle/fordatis/158</a></td>
</tr>
<tr>
<td>CIVIT Dataset (Tampere)</td>
<td>2018-2020</td>
<td>Optronis CP70-12-C-188 with Nikon Fx 35 mm f/1.8 lens. Camera max resolution 4080 x 3072 pixels</td>
<td>6 including Lambertian, complex, synthetic 1280+720 pixels 100-200 camera views, the disparity range between adjacent views is 1 pixel</td>
<td>ICME 2018 and 2020 grand challenge on densely sampled light field reconstruction</td>
<td><a href="https://civit.fi/densely-sampled-light-field-datasets/">https://civit.fi/densely-sampled-light-field-datasets/</a></td>
</tr>
<tr>
<td>Dingcheng et al (Cambridge Fraunhofer IIS)</td>
<td>2020</td>
<td>Sony Alpha a7II camera (mounted 4m horizontal and 0.5m vertical movement)</td>
<td>11 synthetic (512x512x9x9) 4 real (3984 x2440 x50x50)</td>
<td>An objective benchmark of light field view interpolation 5 interpolation methods including DNN-based</td>
<td><a href="https://www.cl.cam.ac.uk/research/rainbow/projects/lightfield-benchmark/">https://www.cl.cam.ac.uk/research/rainbow/projects/lightfield-benchmark/</a></td>
</tr>
</tbody>
</table>
7. Requirements for quality assessment procedures for light field modalities

The quality evaluation (subjective and objective) of light field images is mainly applied to assess the artifacts induced by compression and reconstruction algorithms. Therefore, additional aspects related to use cases should be considered in order to properly design subjective tests and select appropriate metrics, preferably modelled specifically for conducting quality assessment of light field images, to specific use cases. This section is intended to describe a number of use cases and corresponding aspects to be evaluated in quality assessment procedures for light field modalities, listed on Table 8.

<table>
<thead>
<tr>
<th>Table 8 - Use cases and corresponding aspects.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersive user experience</td>
</tr>
<tr>
<td>Horizontal parallax</td>
</tr>
<tr>
<td>Full parallax</td>
</tr>
<tr>
<td>Different viewing angles</td>
</tr>
<tr>
<td>Immersion</td>
</tr>
<tr>
<td>Refocusing</td>
</tr>
<tr>
<td>Interaction</td>
</tr>
<tr>
<td>View interpolation</td>
</tr>
<tr>
<td>Latency</td>
</tr>
<tr>
<td>Vergence/accommodation conflict (VAC)</td>
</tr>
<tr>
<td>Cognitive load</td>
</tr>
<tr>
<td>Visual fatigue</td>
</tr>
<tr>
<td>Metrological quality</td>
</tr>
<tr>
<td>Diagnostic quality</td>
</tr>
<tr>
<td>Depth or disparity estimation</td>
</tr>
</tbody>
</table>
Note that content should be chosen to optimally exploit features (resolution, colour gamut, dynamic range ...) of the display device.

Table 8 summarizes the quality assessment requirements that meet the different use cases (the requirements for objective and subjective assessment should be separated in a future version of the document), that are outlined as follows:

- The standard shall provide means to assess the objective and subjective qualities of the rendering of different static viewing perspectives.
  
  *Use cases from Table 8: different viewing angles*

- The standard shall provide means to assess the smoothness of view switch rendering at different rates of view perspective changes.
  
  ○ The standard shall provide means to simulate different viewer trajectories when assessing view switch rendering.
  
  *Use cases from Table 8: horizontal/full parallax, latency, interaction, immersion*

- The standard shall provide means to assess the objective and subjective qualities of refocused views rendered from the reconstructed light field.
  
  *Use cases from Table 8: refocusing, latency (codec random access when changing focus)*

- The standard shall provide means to assess the latency of rendering refocused views.
  
  *Use cases from Table 8: refocusing, latency (codec random access when changing focus), interaction, immersion*

- The standard shall provide means to assess the objective and subjective qualities of interpolated views rendered from the reconstructed light field.
  
  *Use cases from Table 8: view interpolation, interaction, immersion*

- The standard shall provide means to assess the objective quality of depth/disparity information obtained from the coded representation.
  
  *Use cases from Table 8: depth or disparity estimation*

- The standard shall provide means to assess the accuracy of 3D reconstructions obtained from the coded representation.
  
  *Use cases from Table 8: depth or disparity estimation, metrological quality, diagnostic quality*

- The standard shall provide means to assess the objective and subjective qualities of the stereoscopic 3D rendering from different static viewing perspectives.
  
  *Use cases from Table 8: different viewing angles, VAC, visual fatigue, cognitive load, immersion, interaction*

- The standard shall provide means to assess the quality of experience provided by the dynamic interaction with the 3D scene perceived through stereoscopic rendering from reconstructed views.
  
  ○ The standard shall provide means to simulate different viewer trajectories when assessing quality of experience of 3D perception.
  
  *Use cases from Table 8: horizontal/full parallax, latency, VAC, visual fatigue, cognitive load,*
interaction, immersion

- The standard shall provide means to assess the objective and subjective qualities of the light field display 3D rendering from different static viewing perspectives.

*Use cases from Table 8: different viewing angles, VAC, visual fatigue, cognitive load, immersion, interaction*

- The standard shall provide means to assess the quality of experience provided by the dynamic interaction with the 3D scene perceived through light field display rendering from reconstructed views.
  - The standard shall provide means to simulate different viewer trajectories when assessing quality of experience of 3D perception

*Use cases from Table 8: horizontal/full parallax, latency, VAC, visual fatigue, cognitive load, interaction, immersion*

8. Other light field quality assessment initiatives

ISO/IEC JTC1/SC29/WG04 (MPEG-I) is conducting exploration experiments for coded representation of immersive media. In the context of MPEG-I visual activity, a diverse set of test materials has been collected and summarized [30]. Currently, three main groups of test material are available:

- Sparse Light Field: The view switching will not be smooth without view interpolation.
- Dense Light Field: The view switching will be smooth without view interpolation.
- Omnidirectional: The capturing device is an omnidirectional camera.

MPEG-I exploration on the dense representation of light field (LF) is focused on examining both lenslet and multi-view dense video format and investigates the application of multi-plenoptic video cameras for 3DoF+, 6DoF, and point cloud. The MPEG-I exploration covers lenslet-multiview conversion, calibration, depth estimation, compression, and quality assessment for dense LF. For conversion from lenslet to multi-view, two types of camera architectures Plenoptic camera 1.0 and Plenoptic camera 2.0 have been considered. In case of subjective quality assessment and visual evaluation of the technologies in the MPEG-I visual, LFs are currently visualized using 2D displays in which the participants sweep over the views [31-32].

The MPEG AG on Quality of Immersive Visual Media (ISO/IEC JTC 1/SC 29/AG 5) studies quality metrics and subjective methodologies for immersive media including 360-degree, point cloud, and MPEG Immersive Video (MIV). The subjective experiment for MIV is currently ongoing through an informal viewing session in which sequences are synthesized according to a set of pose traces and proposals are visualized side-by-side with the anchor codec. Participants then will make a binary decision whether the proposal improves over the anchor [33].
9. References

1. ISO/IEC 21794-2:2021 Information technology — Plenoptic image coding system (JPEG Pleno) — Part 2: Light field coding
2. from Loïc Baboulaz, EPFL, 2014.
4. Michael Broxton, John Flynn, Ryan S. Overbeck, Daniel Erickson, Peter Hedman, Matthew DuVall, Jason Dourgarian, Jay Busch, Matt Whalen, Paul E. Debevec
5. https://raytrix.de/products/
12. https://arvr.google.com/
15. https://holografika.com/
16. https://blog.google/technology/research/project-starline/


Annex A: Analysis of selected literature in subjective light field quality assessment

[34]: Perceptual Quality of Light Field Images and Impact of Visualization Techniques
As reference [18] presents a similar pipeline to [34], only the most recent reference is analyzed.

“Perceptual Quality of Light Field Images and Impact of Visualization Techniques”

• Dataset:
  – Lytro Illum: https://muse.uniroma3.it/light-field-image-quality-%820dataset/
  – Selected based on: Colorfulness (CF), Spatial perceptual Information (SI), Lambertian lighting, etc.

• Quality assessment categories:
  – Subjective quality assessment protocol: Absolute Category Rating – ACR.
    • Experiment setup: Graphical User Interface.
    • Training for each rendering technique.
    • Subjects: randomly selected among university students.
  – Objective assessment: 2D Image Quality Metrics
    • 20 metrics have been considered, including SSIM, M-SSIM and PSNR.
  – Independent tests:
    • Impact of Rendering on the Perceptual Quality of LF Data.
    • Effect of Typical Compression and Noise Artifacts on LF Image QoE.

• Rendering:
  – All-in-focused-view;
  – Pseudo-video;
  – Refocused image;
  – Refocused-pseudo-video.
  – Images and videos were displayed by using a computer monitor (DELL U2413 digital monitor, with resolution of 1920×1200 pixels), and the background of the desktop was set to gray color.

• Types of distortion assessed:
  – Compression artifacts:
    • JPEG;
    • JPEG2000;
    • HEVC intra encoding.
  – Gaussian noise.
[35]: A Study on the Impact of Visualization Techniques on Light Field Perception

- **Datasets:**
  - Selected LFs from the EPFL dataset:
    - Selection based on Spatial perceptual Information (SI) and Colorfulness (CF) values.
- **Quality assessment categories:**
  - Subjective quality assessment protocol:
    - Absolute Category Rating – ACR.
  - Pseudo-videos created using 6 different scanning patterns (serpentine scan has been selected by the subjective assessment tests)
    - Played at 10, 15 and 20 fps on a 2D display.
- **Types of distortion assessed:**
  - No distortions introduced – original LFs.

[36]: Quality of Experience in a Stereoscopic Multiview Environment

- **Datasets:**
  - Online computer generated;
  - Offline computer generated using Physically-Based Ray Tracing – PBRT;
  - HDCA using a linear translating gantry.
- **Quality assessment categories:**
  - Subjective assessment:
    - The paper investigates how visualization factors, such as disparity, mobility, angular resolution, and viewpoint interpolation, influence the quality of experience (QoE) in a stereoscopic multiview environment.
    - The QoE aspects investigated are represented by items
      - (Q1) Visual comfort: Visual comfort refers to symptoms such as eyes tiredness, headache, nausea and dizziness; the higher the grade, the greater the comfort;
      - (Q2) Sense of immersion: The sense of immersion refers to the sensation of immersion with the environment; the higher the grade, the higher the sense of immersion;
      - (Q3) Difficulty to complete the task: Difficulty to execute the given task, as described in Table II; the higher the grades the smaller the difficulty;
      - (Q4) Experience as a whole: Overall experience of the performed test; the higher the grade, the better the experience.
- **Rendering:**
  - Stereoscopic displays: 23.6” Acer GD235HZ and a 46” JVC 463D10U display, with active and circular polarization technologies, respectively.
  - Interactive stereoscopic system implemented using 3D models and an eye tracking system.
- **Types of distortion assessed:**
  - No distortion introduced.
Annex B: Analysis of selected literature in objective light field quality assessment

[22]: LFDD: Light field image dataset for performance evaluation of objective quality metrics

- **Datasets:**
  - Synthetic: HCI 4D Light Field Dataset.
    - Spatial perceptual Information measurement (SI) was used to select LFs.
- **Quality assessment categories:**
  - Subjective assessment:
    - Modified version of the Double Stimulus Impairment Scale (DSIS) methodology, ITU-R Rec. BT.500-14.
  - Objective assessment: 2D Image Quality Metrics
    - PSNR; SSIM; MS-SSIM; FSIM and FSIMc; UIQI; VIF.
- **Rendering:**
  - Pseudo-video: serpentine scan order;
  - No display description has been provided.

- **Types of distortion assessed:**
  - Added distortion:
    - Gaussian and impulse noise;
    - Pincushion distortion.
  - Compression artifacts:
    - JPEG; JPEG 2000; HEVC; VP9; AV1; H.264; Better Portable Graphics (BPG).

[37]: On the performance of objective quality metrics for lightfields

- **Datasets:**
  - MPI-lightfield data set
    - Synthetic and real scenery, each represented by 101 perspective views of resolution 960×720 with horizontal parallax, coded by 3D-HEVC.
  - SMART data set (lenslet)
  - VALID data set (lenslet)
    - 625×434×15×15, coded by H.265/HEVC, VP9 and and by codecs proposed in three different papers.
  - All with annotated data.
• Quality assessment categories:
  – Objective assessment: 2D Image Quality Metrics:
    – FR 2D quality assessment methods: PSNR; SSIM; MS-SSIM; IFC; UQI; VSNR; VIF; IW-PSNR; IW-SSIM; GMSD; HDR-VDP; MAD; SSRM; FSIM;
    – NR 2D quality assessment methods: NIQE; BRISQUE; IL-NIQE; dipIQ; MEON; HOSA;
    – FR 3D quality assessment methods: MJ3DFR; FI-PSNR;
    – Video quality assessment methods: STMAD; Vis3.

• Rendering:
  – MPI-lightfield data set
    • The LFs were displayed on an ASUS VG278 27” Full-HD 120 Hz LCD desktop monitor with a pair of 3D active glasses to provide stereoscopic viewing.

  – SMART data set (lenslet)
    • The test LFs were displayed on a DELL U2413f screen with 1920×1200 pixels resolution;
    • The LF contents were rendered in 2D view and refocused at different points to produce an all-in-focused view for subjective assessment. Similar to the MPI-Lightfield, a pairwise comparison (PC) method was used for the quality rating in which the user selected the image with higher quality.

  – VALID data set
    • The LFs were displayed on an Eizo ColorEdge CG318-4K.
    • Pseudo-video in serpentine order displayed at 10 frames per second.

• Types of distortion assessed:
  – 3D-HEVC. Interpolation (optical flow); Display crosstalk; JPEG; JPEG 2000; H.265/HEVC Intra; VP9; three light field codecs.

[38]: An objective benchmark of light field view-interpolation methods

• Datasets:
  - A new dataset covering 5 interpolation methods
    - synthetic dataset of 11 scenes 512x512p per view and 9x9 views per scene.
    - A real dataset of 4 scenes 3984x2440 pixels per view, 50x50 views per scene. These images are cropped to several 512x512 patches of interesting regions.

• Quality assessment categories:
  - Objective assessment:
    - 2D metrics: PSNR, SSIM, VSI, GMSD and HDR-VDP-2
    - DIBR metric: 3DsvIM, MP-PSNR and MW-PSNR
    - Video quality metric: VQM

• Types of distortions
  - Five interpolation methods