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: Use Cases and Requirements for Light Field Quality Assessment v4.0

SOURCE: WG1

EDITORS: Carla Pagliari, Peter Schelkens, Saeed Mahmoudpour, Eduardo A. B. da Silva, and Mylène Farias

PROJECT: JPEG Pleno

STATUS: Approved

REQUESTED ACTION: Distribution

DISTRIBUTION: Public

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

This is a living document that goes through iterations. Proposals for revisions of the text can be delivered to Carla Pagliari and Saeed Mahmoudpour, by downloading this document, editing it using track changes and sending it to carla.pagliari@ieee.org and saeed.mahmoudpour@vub.be

If you have interest in JPEG Pleno Light Field Coding activities, please subscribe to the email reflector, via the following link: http://listregistration.jpeg.org/
1. Introduction and Scope

Visual quality is important to applications where humans are the consumers of visual information. A large number of Image Quality Assessment (IQA) algorithms has been developed to subjectively and objectively evaluate natural and synthetic 2D images (and video). Several image databases annotated with subjective ratings to support the objective quality assessment methods, are available. As different plenoptic modalities, such as light fields, point clouds and holography, present different use cases and requirements, specific annotated databases (ground-truth databases) as well as specific subjective and objective quality assessment methods need to be designed. In this context, the purpose of this document is to initiate a standardization effort initially focused on light fields quality assessment. The other plenoptic modalities will be dealt with in the near future.

Light field imaging is defined based on a representation where the amount of light (the “radiance”) at every point in space and in every direction is made available [1]. This radiance can be approximated and captured by either an array of cameras (resulting in a wide baseline light field data) or a single light field camera using microlenses to sample the individual rays of light that contribute to the final image (resulting in a narrow baseline light field data). Combinations of the two capture approaches are also possible. Furthermore, light field data can be synthetically generated from computer models.

During the acquisition/generation of light fields, two important properties are defined: the angular (given by the baseline) and the 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 visual experience. In contrast, specific 3D displays would be able to replicate the rays of light (light field), including the directional and colour components.

Therefore, acquisition/generation and visualization of light fields and content (datasets) are key components to meet the functionalities given various use case contexts. For example, use cases of medical and industrial imaging demand displays that provide the functionality of different viewing angles. Therefore, the standard shall provide means to assess the objective and subjective qualities of different static viewing perspectives' views rendered from the decoded light field.

In the context of the JPEG Pleno standardization process with respect to light field coding (Part 2) [2], various subjective visual quality assessment procedures have been designed and used; thus, significant knowledge has been built-up with respect to challenges, good practice guidelines and methodological aspects. To assist the scope of this standardization effort, this document targets the definition of use cases and requirements to support the development of a light field quality assessment standard both for subjective and objective assessment.

The scope of JPEG Pleno Quality Assessment is the creation of a quality assessment standard, defining a framework including subjective quality assessment protocols and objective quality assessment procedures for lossy decoded data of plenoptic modalities in the context of multiple use cases. The first phase of this effort will address the light field modality and should build on the light field quality assessment tools developed by JPEG in recent years.
2. 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 are simultaneously captured. However, non-static light field use cases may be addressed in the future. Use cases and functionalities are detailed in this section.

2.1. XR-based Applications

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, that provide an alternative high-performance imaging system to stereo and multi-view.

The key difference between an AR and a VR device is that while the former has to superimpose/combine virtual content into 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 [3], Samsung Odyssey+ [4], PlayStation VR [5], Oculus Quest [6], and Oculus Rift [6]) and standalone headsets (Samsung Gear VR [4], Google Daydream [7], and Google Cardboard [7]) aim to offer the users realistic visual experiences. Users create VR experiences by moving their heads so that their view perspective is shown by the headset. Figure 1 shows a child wearing an HTC Vive [3] headset.

Many factors play important roles 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 a large field of view (FOV), refreshing at a fast frame rate. Head mounted displays, or wearable displays, or near-eye displays create a virtual image in the FOV of one or both eyes, integrating one or multiple image sources in the device. A key 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 integrates its optics with the eye models, thus mitigating the vergence-accommodation conflict [9]. 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. Among the relevant functions that may be performed, refocusing and interaction are at the top of the list.

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 for dedicated 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 state-of-the-art coding technology will allow multiple viewers experience and not only single viewer experience with standalone devices.

2.2. Industrial Imaging

Metrology based on light field imaging may be useful for numerous types of applications. With more information available, better analysis, decision and control performance can be achieved, particularly increasing the 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 industrial imaging 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 technique 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.

Figure 2 illustrates 3D plant analysis and metrology using light fields.

![Figure 2. 3D plant analysis and metrology using light fields using Raytrix cameras [10].](image-url)
2.3. 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 3 illustrates medical imaging using light field cameras.

![Figure 3. Medical image](image)

2.4. Light Field Telepresence

The “sense of presence” is the ability to experience the virtual presence of another person. Light field technology aims at providing 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) light field 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 4 shows two “telepresence” light field displays.

![Figure 4. Telepresence light field displays](image)
2.5. Summary
Table 1 summarises the use cases and corresponding needed functionalities.

<table>
<thead>
<tr>
<th>Immersive user experience</th>
<th>Industrial imaging</th>
<th>3D modelling</th>
<th>Medical</th>
<th>Telepresence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal parallax</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full parallax</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different viewing angles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refocusing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View interpolation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vergence/accommodation conflict (VAC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual fatigue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrological quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth or disparity estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Requirements
The subjective and objective quality evaluation of light field images is mainly targeted to assess the artefacts induced by coding algorithms. This section presents the general quality assessment requirements, common to subjective and objective quality assessment solutions, the specific subjective quality assessment requirements, the specific objective quality assessment requirements and the functionality-related quality assessment requirements.

3.1. General Quality Assessment Requirements
The following general quality assessment requirements have been identified:

- Light field type diversity: The standard shall be capable of effectively assessing the quality of light fields with different angular samplings (from narrow to wide baseline), and with different spatial resolutions, generated by different camera models (e.g., plenoptic cameras, planar and non-planar camera arrays).
Content type diversity: The standard shall be capable of effectively assessing the quality for different types of scene content, including but not limited to natural and computer-generated stimuli, specularity and transparency and low to high dynamic range.

Viewing and display characteristics and configurations diversity: The standard should be adaptable to different viewing and display characteristics and configurations.

Coding artefacts diversity: The standard shall be capable of effectively assessing coding artefacts, including but not limited to geometrical distortions, occlusion-related artefacts, spatial artefacts (e.g., blockiness, blurring), spatial-angular quality degradation, and distortion types interactions.

Functionality diversity: The standard shall be capable of effectively assessing the quality of light fields under different functionality requirements.

3.2. Subjective Quality Assessment Requirements

The following subjective quality assessment requirements have been identified:

- **Reliability**: The standard shall consistently produce statistically equivalent results when using the same methods under the same circumstances (e.g., test stimuli to probe for the attention of test subjects).

- **Reproducibility**: The standard shall be capable of producing consistent results when the experiment is replicated.

- **Efficiency**: Given a targeted statistical sensitivity, the standard should strive to reach it with a minimum number of test subjects and conditions.

- **Test setup complexity**: Given a targeted statistical sensitivity, the standard should minimise the required test setup complexity.

- **Scalability**: The standard shall be scalable in terms of test features (e.g., viewpoints, test data, and compression artefacts).

- **Test subject screening**: The standard shall include screening procedures for test subjects with respect to visual acuity, colour vision deficiency and 3D perception capabilities (e.g., depth acuity, vergence facility).

- **Score screening**: The standard shall define procedures to increase the reliability of the score data (e.g., outlier removal).

- **Controllability**: The standard should be effective in scenarios with limited control over the experimental setup.

3.3. Objective Quality Assessment Requirements

Subjectively-annotated light field databases will support the evaluation of the objective quality assessment methods. The following objective quality assessment requirements have been identified:

- **Statistical validity**: The standard shall provide quality scores that are valid against the differential subjective scores, as quantified by appropriate statistical measures notably correlation measures.

- **Objective methods versatility**: The standard shall provide objective solutions that address an as wide range of artefacts as possible.
- Computational complexity: The standard should define objective quality assessment measures with reasonable computational complexity, allowing implementation on multiple platforms (in software and hardware).

- Reference-based: The standard shall provide a quality score by comparing a light field to be assessed to a reference (e.g., assessing the quality of the decoded views and/or depth information).

- No-reference: The standard should provide a quality score of a light field to be assessed without the need for any reference (e.g., assessing the quality of the decoded views).

3.4. Functionality-related Quality Assessment Requirements

The following functionality-related quality assessment requirements (and associated use cases and functionalities) have been identified:

A. Content-related functionality requirements

- View switching smoothness: The standard shall provide means to assess the perceived smoothness along a viewing trajectory.
  
  *Use cases from Table 1: horizontal/full parallax, immersion*

- View switching trajectories: The standard shall provide means to simulate different viewer trajectories.

- Stereoscopic quality: The standard shall provide means to assess the objective and subjective (auto) stereoscopic quality of the decoded light field.
  
  *Use cases from Table 1: different viewing angles, vergence-accommodation conflict, visual fatigue, cognitive load, immersion, interaction*

B. Rendering/display-related functionality requirements

- Refocused views: The standard should provide means to assess the objective and subjective qualities of refocused views rendered from the decoded light field.
  
  *Use cases from Table 1: refocusing, latency (due to codec random access when changing focus)*

- Interpolated views: The standard should provide means to assess the objective and subjective qualities of interpolated views rendered from the decoded light field.
  
  *Use cases from Table 1: view interpolation, interaction, immersion*

The requirements listed above are directly related to the properties of the decoded light field content and their impact on the potentially subsequent rendering mechanisms. Requirements that relate to viewpoint switching under viewer control and 3D reconstruction are facilitated, but potential latency issues are not included since they relate too much to the properties of the application. The impact of the codec is in these cases is often buried in the overall system behaviour and consequently hard to differentiate from the other system capabilities.
4. Royalty-free Goal

The royalty-free patent licensing commitments made by contributors to previous standards, e.g., JPEG 2000 Part 1 and JPEG XL, have arguably been instrumental to their success. JPEG expects that similar commitments would be helpful for the adoption of new standards.

5. References

[8] https://www.youtube.com/watch?v=mWVZoMniOYs
[10] https://raytrix.de/products/
[12] https://blog.google/technology/research/project-starline/