TITLE: JPEG Pleno Light Field Quality Assessment Common Test Conditions v1.1

SOURCE: WG 1

EDITORS: Carla Pagliari, Saeed Mahmoudpour, Fernando Pereira, Eduardo A. Barros da Silva, António Pinheiro, Shengyang Zhao, Yanding Peng, Hadi Amirpour

PROJECT: JPEG Pleno

STATUS: Approved

REQUESTED ACTION: For 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
JPEG PLENO - LIGHT FIELD QUALITY ASSESSMENT
COMMON TEST CONDITIONS

Table of contents

1. Scope 4
2. Test materials 4
2.1. JPEG Pleno Light Field database and additional light fields 4
2.1.1. JPEG Pleno Light Field database 4
   Lenslets 5
   HDCA 6
   Synthetic 7
2.1.2. Additional datasets 7
3. Coding conditions 8
3.1. Target bitrates 8
3.2. JPEG Pleno Light Field Quality Assessment Relevant Codecs 8
3.2.1. HEVC Coding 8
3.2.2. Pseudo-temporal Sequence Generation 10
3.2.3. JPEG Pleno Light Field Coding (Part -2) 11
3.2.4. VVC Coding 13
4. Evaluation Criteria for Subjective Quality Assessment 14
5. References 15
1. Scope

This document describes the Common Test Conditions for the collaborative process of the JPEG Pleno light field quality assessment contributions [1].

The main objectives of this document are:

- Define test materials to be used in the evaluation of subjective quality assessment methodologies, representative of light field data commonly used in use cases specified in [2].
- Define coding conditions to generate the test materials.
- Define criteria to identify and assess the relevant properties of the different contributions.

As the JPEG Pleno Final Call for Contributions on Subjective Light Field Quality Assessment [1] aims at developing the standard by consensus among the JPEG experts following a collaborative process, proponents are welcome to revise, improve and complete this document.

2. Test materials

This section describes the test materials selected for the collaborative process of the JPEG Pleno Light Field Quality Assessment Contributions. All data is available on the JPEG Subjective Quality Assessment website [3].

The JPEG Pleno subjective light field test material is diverse in terms of

- Coding artefacts
- Acquisition/creation technology, notably
  - Lenslet Lytro Illum camera
  - High Density Camera Array (HDCA)
  - Synthetic creation
- Scene geometry
- Spatial resolution
- Spatial complexity
- Number of views/perspectives
- Bit-depth

2.1. JPEG Pleno Light Field database and additional light fields

Large-scale and rich-diversity datasets play an important role in developing subjective and objective quality assessment methods. This section presents the uncompressed light fields of the JPEG Pleno Light Field database [4] and a list of requirements for potential additional light fields.

2.1.1. JPEG Pleno Light Field database

Current JPEG Pleno Light Field database presents different scene geometries and spatial-view geometry diversity [4, 5], and is composed of the following data:
- **Lenslets**: *Bikes, Danger de Mort, Fountain&Vincent2, and Stone Pillars Outside*;
- **HDCA** (High Definition Camera Array): *Tarot and Laboratory1*; and
- **Synthetic**: *Greek and Sideboard*.

Table I includes a summary of the light fields from the JPEG Pleno Light Field database [4], with their example views depicted in Figures 1 to 3.

<table>
<thead>
<tr>
<th>Light field name</th>
<th>Number of views</th>
<th>Spatial resolution (pixels)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bikes</em></td>
<td>13×13</td>
<td>625×434</td>
<td>Lenslets</td>
</tr>
<tr>
<td><em>Danger de Mort</em></td>
<td>13×13</td>
<td>625×434</td>
<td>Lenslets</td>
</tr>
<tr>
<td><em>Fountain&amp;Vincent2</em></td>
<td>31×31</td>
<td>1936×1288</td>
<td>HDCA</td>
</tr>
<tr>
<td><em>Stone Pillars Outside</em></td>
<td>31×31</td>
<td>1936×1288</td>
<td>HDCA</td>
</tr>
<tr>
<td><em>Laboratory1</em></td>
<td>31×31</td>
<td>1936×1288</td>
<td>HDCA</td>
</tr>
<tr>
<td><em>Tarot</em></td>
<td>17×17</td>
<td>1024×1024</td>
<td>HDCA</td>
</tr>
<tr>
<td><em>Greek</em></td>
<td>9×9</td>
<td>512×512</td>
<td>Synthetic</td>
</tr>
<tr>
<td><em>Sideboard</em></td>
<td>9×9</td>
<td>512×512</td>
<td>Synthetic</td>
</tr>
</tbody>
</table>

**Lenslets**

- *Bikes, Danger de Mort, Fountain&Vincent2, and Stone Pillars Outside*:
  - These outdoor natural scene content light fields present dense angular sampling (narrow baselines), with the same spatial resolution, and have been acquired by the same camera model [2].
  - The light fields *Bikes* (I01) and *Stone Pillars Outside* (I04) belong to a category 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, where the latter presents low to medium degrees of spatial complexity [2].

---

1 Only the central 13×13 views are used to avoid dark views due to vignetting of the original 15×15 views.
Danger de Mort (I02) images are part of the category Grids that have grid patterns with close details with a wide depth of field range [2].

Fountain&Vincent2 (I09) is part of the category People, displaying one person and a fountain, which are very close to the camera presenting a high level of spatial complexity.

All present positive and negative disparities.

Fig 1. From left to right and from top to bottom: Bikes, Danger de Mort, Stone Pillars Outside, and Fountain&Vincent2 example views.

HDCA

- Laboratory1:
  - This light field presents a sparse angular sampling (wide baseline) and has been acquired by a high-definition camera array [2].
  - It presents a natural indoor scene content with repetitive patterns, different levels of spatial complexity, and challenging scene geometry [2].

- Tarot:
  - This light field presents a non-dense angular sampling (medium baseline) and has been acquired by a high-definition camera array [2].
  - It presents a natural indoor scene content with non-Lambertian surfaces, different levels of spatial complexity, and positive and negative disparities. [2].
Synthetic

- **Greek:**
  - This computer-generated light field presents a sparse angular sampling (wide baseline), low spatial complexity, and positive and negative disparities. [2].

- **Sideboard:**
  - This computer-generated light field presents a non-dense angular sampling (medium baseline), different levels of spatial complexity, repetitive patterns, positive and negative disparities, and a complex scene geometry [2].

2.1.2. Additional datasets

Additional datasets with different types of light fields acquisition might include:
- Computer-generated light fields, lenslets-based, and HDCA;
- Different angular samplings (from narrow to wide baseline);
- Different parallax levels;
- Different spatial resolutions;
- Different/complex scene geometries;
- Different color gamut;
– Different degrees of spatial complexity;
– Different bit-depths;
– Different number of views/perspectives;
– Test data with specularity and transparency;
– Content type diversity.


3. Coding conditions

This section describes the codecs and associated target bitrates for the experiments performed with the JPEG Pleno Light Field database [4]. The original (uncompressed) and decoded light fields, as well as the corresponding codestreams, are available to the proponents from the JPEG Subjective Quality Assessment website [3]. The selected codecs employ different coding tools, thus generating a diversity of coding artefacts. For example, depth-based codecs may introduce geometrical distortions, in addition to the usual spatial artefacts (e.g., blockiness, blurring). All the selected coding methods also produce spatial-angular quality degradation and interactions among different distortion types.

3.1. Target bitrates

Table II shows the bitrates defined by the JPEG Pleno Common Test Conditions Light Field Coding [5], but they can be revised and extended in the future.

<table>
<thead>
<tr>
<th>Light Fields</th>
<th>Target Bitrate (bpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenslets, Synthetic and Tarot</td>
<td>0.005</td>
</tr>
<tr>
<td>Laboratory1</td>
<td>0.001</td>
</tr>
</tbody>
</table>

3.2. JPEG Pleno Light Field Quality Assessment Relevant Codecs

This section describes the coding processes with several relevant codecs to create the decoded light fields with compression artifacts for various coding conditions. This will be the test material for the to be developed subjective assessment protocols. The original light fields, codestreams, decoded light fields, and auxiliary files (e.g. pseudo-temporal sequences, encoder configuration files) are available at [3].

3.2.1. HEVC Coding

This section describes the coding process as detailed in [4]. The following conditions apply

– The x.265 implementation [6] of HEVC, source version 2.3 (available from [7]) should be used. The source has to be compiled with MAIN10 (10-bit profile) profile support. Please read the Build Considerations at [8] for detailed information. The build instructions for Linux use the shell script
make-<code>Makefiles.bash</code>, while the one for Windows uses the <code>make-solutions.bat</code> file. Please note that the build configuration to 10-bit support sets the <code>HIGH_BIT_DEPTH</code> parameter. The <code>HIGH_BIT_DEPTH</code> parameter should be turned ON, during the build process, in order to create a binary file for the requested bit depth (10-bit). This parameter allows 10-bit encoding; however, the compressed version of the 10-bit files, once decoded, appear as if they were scaled so that their dynamic range is 16 bits. Therefore, after decoding the reconstructed images have to be scaled back to the original 10-bit dynamic range. (Note: Check the latest software version and make sure it works). This code requires re-compiling for 8-bit dynamic range.

- The input YUV 4:4:4 file is a pseudo-temporal sequence generated according to the command in Section 3.2.2. The 8-bit datasets should be converted to 10-bit using the ImageMagick [9] command <code>convert</code> as follows:

  ```bash
  Your_path/convert SSS_TTT.ppm -depth 10 SSS_TTT_10bit.ppm
  ```

- For encoding the lenses, <em>Greek</em>, <em>Sideboard</em> and <em>Tarot</em> light fields as an YUV 4:4:4 pseudo-temporal file, a two-step encoding process is used, which is based on the following command line arguments:

  **Step 1:**

  ```bash
  Your_path/x265 --input ${IN_PATH} --input-depth 10 --input-csp i444 --fps 30 --input-res "${I_W}x${I_H}" --output-depth 10 --profile main444-10 --output /dev/null --pass 1 --stats "stats_log"
  ```

  **Step 2:**

  ```bash
  Your_path/x265 --input ${IN_PATH} --input-depth 10 --input-csp i444 --fps 30 --input-res "${I_W}x${I_H}" --output-depth 10 --profile main444-10 --bitrate ${RATE} --output ${OUT_PATH} --pass 2 --stats "stats_log"
  ```

  To be specified: 〈RATE〉 values and I_W x I_H.

  The 8-bit light fields (Tarot, Greek, and Sideboard) should be converted to 10-bit representation prior to encoding and performance computation.

- For encoding the <em>Laboratory1</em> light field as an YUV 4:4:4 pseudo-temporal file, a two-step encoding process is used, which is based on the following command line arguments:

  ```bash
  Your_path/x265 --input ${IN_PATH} --input-depth 10 --input-csp i444 --fps 30 --input-res"${I_W}x${I_H}" --output-depth 10 --profile main444-10 --output ${OUT_PATH} --crf ${CRF}
  ```

  To be specified: I_W x I_H plus CRF.

  The 8-bit light field (Laboratory1) should be converted to 10-bit representation prior to encoding and performance computation.
In principle, any compiled decoder can be used. For convenience, an example is provided below on how the ffmpeg tool [10] can be used for decoding:

```
Your_path/ffmpeg -i HDCA_Set2_2K_sub_010_020_444_000250.265
results_directory/ Lab1_010_020_000250_%03d.ppm
```

Please note that the output of the ffmpeg command is a 16-bit PPM of 10-bit PPM reconstructed light fields. To convert from 16-bit to 10-bit, one can use any tool. For convenience, an example is provided below on how the ImageMagick [9] convert tool can be used for converting 16-bit PPM files to 10-bit PPM files:

```
Your_path/convert Lab1_010_020_444_000250_0001.ppm -depth 10
HDCA_Set2_2K_sub_010_020_444_000250_0001_10bit.ppm
```

### 3.2.2. Pseudo-temporal Sequence Generation

The input file for the encoding scheme is a YUV 4:4:4 10-bit file (in ppm file format), scanned and concatenated following a serpentine scanning order, as defined in the JPEG Pleno document [5], repeated here for convenience as Figure 4. Please note that the PPM to YUV conversion needs to use the ITU-R Recommendation BT. 709-6 [11].

An example on how to generate such a pseudo-temporal for a 3×3 light field ABCD with dimensions 400×300 is given below:

```
Your_path/ffmpeg -r 30 -f concat -safe 0 -I list.txt -s 400x300 -framerate 30 -c:v rawvideo -pix_fmt yuv444p10le2 ABCD_3x3.yuv
```

where the file list.txt should be as in Figure 5:

```
file 'ABCD_000_000.ppm'
duration 1
file 'ABCD_001_000.ppm'
duration 1
file 'ABCD_002_000.ppm'
duration 1
file 'ABCD_002_001.ppm'
duration 1
file 'ABCD_001_001.ppm'
duration 1
file 'ABCD_000_001.ppm'
duration 1
file 'ABCD_000_002.ppm'
duration 1
file 'ABCD_001_002.ppm'
duration 1
file 'ABCD_002_002.ppm'
duration 1
```

![Fig. 4. Pseudo-temporal sequence scan [4].](image1)

**Fig. 4. Pseudo-temporal sequence scan [4].**

**Fig. 5. list.txt file [4].**

---

2 yuv444p10le : planar YUV 4:4:4, 30bpp, (1 Cr & Cb sample per 1x1 Y samples), little-endian [10].
An example on how to play the pseudo-temporal sequence is given below:

```
Your_path/ffplay -f rawvideo -pixel_format yuv444p10le -video_size 625x434 -i Bikes.yuv
```

### 3.2.3. JPEG Pleno Light Field Coding (Part -2)

This section describes the generation process for the two coding modes of the JPEG Pleno Part-2 Light field coding standard [12].

The following conditions apply

- The input light fields should be 10-bit PPM (Portable Pixmap Format) files.
- The JPEG Pleno Part-2 has two encoding modes: 4D-Prediction Mode (4D-PM) and 4D-Transform Mode (4D-TM).
- The Verification Model 2.1 (VM2.1) software available at [13] is used to encode and decode the light fields.
- The Kakadu software is required [14] for the JPEG Pleno Part-2 4D-PM.

- The 4D-TM general encoding command line is:

```
Your_path/VM2.1/jpl-vm-encoder-bin --type 0 \ 
   --input <Input directory containing a set of uncompressed light field images (SSS_TTT.ppm)> \ 
   --output <Output directory containing temporary light field data and the compressed bitstream> \ 
   --config <Path to config file (.json for 4D-TM)> \ 
   --mule <Directory containing the 4D-TM codec (by default, it is the same directory of jpl-vm-encoder-bin)>
```

- The 4D-TM encoding command lines for Bikes and Greek light fields are:

```
Your_path/VM2.1/jpl-vm-encoder-bin --type 0 --input /input_PPM/Bikes/ --output \ 
   /4DTM_encoder_output/Bikes/ --config /configuration_files/bikes.json \ 
   --mule /VM2.1/
```

```
Your_path/VM2.1/jpl-vm-encoder-bin --type 0 --input /input_PPM/Greek/ --output \ 
   /4DTM_encoder_output/Greek/ --config /configuration_files/greek.json \ 
   --mule /VM2.1/
```

- The 4D-TM general decoding command line is:

```
Your_path/VM2.1/jpl-vm-decoder-bin --type 0 \ 
   --input <Input coded light field bitstream (output.LF)> \ 
   --output <Output directory the decoded light field in format
```
SSS_TTT.ppm> \ 
    --mule <Directory containing the 4D-TM codec (by default, it is the same directory of jpl-vm-encoder-bin)> \ 
    --config <Path to config file (.json for 4D-TM)>

- The 4D-TM decoding command lines for Bikes and Greek light fields are:

```bash
Your_path/VM2.1/jpl-vm-decoder-bin --type 0 \ 
    --input /4DTM_encoder_output/Bikes/output.LF \ 
    --output /4DTM_decoder_output/Bikes/ --mule /VM2.1/ \ 
    --config /configuration_files/bikes.json
```

```bash
Your_path/VM2.1/jpl-vm-decoder-bin --type 0 \ 
    --input /4DTM_encoder_output/Greek/output.LF \ 
    --output /4DTM_decoder_output/Greek/ --mule /VM2.1/ \ 
    --config /configuration_files/greek.json
```

- The 4D-PM general encoding command line is:

```bash
Your_path/VM2.1/jpl-vm-encoder-bin --type 1 \ 
    --input <Input directory containing a set of uncompressed light field images (SSS_TTT.ppm)> \ 
    --output <Output directory containing temporary light field data and the compressed bitstream> \ 
    --kakadu <Directory containing the Kakadu codec (kdu_compress)> \ 
    --config <Path to config file (.conf for 4D-PM)>
```

- The 4D-PM encoding command line for Greek light field is:

```bash
Your_path/VM2.1/jpl-vm-encoder-bin --type 1 --input /input_PPM/Greek/ \ 
    --output /4DPM_encoder_output/Greek/ \ 
    --kakadu /VM2.1/kakadu_compress \ 
    --config /configuration_files/greek-81_46.2945_0.11016.conf
```

- The 4D-PM general decoding command line is:

```bash
Your_path/VM2.1/jpl-vm-decoder-bin --type 1 \ 
    --input <Input coded light field bitstream (output.LF)> \ 
    --output <Output directory the decoded light field in format SSS_TTT.ppm> \ 
    --kakadu <Directory containing the Kakadu codec (kdu_expand)>
```

- The 4D-PM decoding command line for Greek light field is:

```bash
Your_path/VM2.1/jpl-vm-decoder-bin --type 1 \ 
    --input /4DPM_encoder_output/Greek/output.LF \ 
    --output /4DPM_decoder_output/Greek/ \ 
    --kakadu /VM2.1/kakadu_expand
```
3.2.4. VVC Coding

This section describes the coding process with the VVC codec. The following conditions apply:

- The input light field image should be 10-bit PPM format and converted to YUV 4:4:4 file format with the command described in section 3.2.2.

- The VTM 17.2 available at [15] is used for encoding and decoding. The encoder is modified to remove the GOP size constraint. The Random Access configuration is used. Four coding files (.cfg) are provided to implement the coding structure as in [16]: Lenslet.cfg, Synthetic.cfg, Lab1.cfg, and Tarot.cfg.

- As the pseudo-temporal sequence generated in section 3.2.2 is in serpentine order, it is needed to be reordered. To encode the light field with coding structure as in [16], a python script (OrderTransform.py) is provided to re-arrange the order of the views:

```python
Your_path/python OrderTransform.py --origin Bikes.yuv --target Bikes_reorder.yuv
   --origin_order serpentine --target_order hadi
   --order_file VVCScanorder.json
```

Here the VVCScanorder.json is a json file containing the target coding structure.

- The encoding command line for Bikes is:

```bash
Your_path/VVC/bin/EncoderAppStatic -c Lenslet.cfg
   -i Bikes_reorder.yuv -q 27
   -b Bikes_qp_27.bin -o Bikes_qp_27.yuv
```

For Bikes, Pillars, Fountain and Danger the Lenslet.cfg is used. The Synthetic.cfg is for Greek and Sideboard. The Lab1.cfg is for the Lab1 light field and Tarot.cfg is for the Tarot light field.

- The VVC encoder and ffmpeg tool are used in the decoding process. A python script is provided to help decode the codestreams into PPM images automatically. The decoding command lines for Bikes are as follows.

```bash
Your_path/python VVCDecode.py --decoder_path
your_path/VVC/bin/DecoderAppStatic
   --ffmpeg_path ffmpeg
   -b Bikes_qp_27.bin
   -s 625x434 -a 13x13
   --scan_order ./VVCScanorder.json
   -o output_folderYour_path/VVC/bin/EncoderAppStatic -c Lenslet.cfg -i Bikes.yuv -q 27 -b Bikes_qp_27.bin -o Bikes_qp_27.yuv
```

This python script will read the coding order from VVCScanorder.json and decode the .bin file to yuv444 format. Then it automatically builds a folder named Bikes_qp_27 in the output folder. Finally, it calls the ffmpeg tool to convert the YUV file to PPM files.
4. Evaluation Criteria for Subjective Quality Assessment

The methods submitted as a contribution to the CfC will be reviewed by testing them in the laboratories of the participating organizations. Statistical analysis will be employed to analyze and compare quantitatively the results of the contributed methodologies. The test may include, but is not limited to: root mean square error (RMSE), Pearson correlation coefficient, Spearman's rank-order correlation coefficient, and outlier ratio [17-18]. These will be computed for accuracy, linearity, monotonicity, and consistency, respectively. The quantitative evaluation criteria aim to assess the repeatability, reliability, and efficiency of the contributions according to the subjective quality assessment requirements identified in the "Use Cases and Requirements for Light Field Quality Assessment" document [2].

- **Size of confidence intervals (CI):** The size of CI [19-20] of subjective scores on each stimulus will be computed in each subjective test and may be compared across multiple experiments (reliability and repeatability).

- **Discriminability score:** A more accurate method should result in more pairs of stimuli whose quality can be said to be different under a statistical test [19-20]. Significance tests between scores of each possible pair of stimuli will be performed and the percentage of stimuli pairs with a significant difference will be obtained (reliability and repeatability).

The following criteria will be used to collect the general statistics of the behavior of the studied subjective methodologies:

- **Correlation:** the correlation between the scores obtained in different laboratories may be computed, for example using the Pearson linear correlation coefficient (PLCC) and the Spearman rank order correlation coefficient (SROCC). Additionally, the root-mean-square error (RMSE) and the outlier ratio (OR) may be computed [17-18].

- **Student's t-test or Welch’s t-test:** may be used to assess whether the means of the collected scores differ significantly [17-18].

- **ANOVA:** a multi-way ANOVA will be performed on the scores obtained in different laboratories with the goal of verifying whether the obtained scores are statistically equivalent [17-18].

- **CE, UE, OE:** the correct estimation (CE), over-estimation (OE), and under-estimation (UE) may be computed to see if the scores are statistically distinguishable across the different experiments proposed in the different laboratories [21].

- **CD, FR, FD, FT:** correct decision (CD), false ranking (FR), false differentiation (FD), or false tie (FT) may also be computed [22].

To assess the efficiency requirement [2], the above criteria will be used for various numbers of subjects. Subjective tests will be conducted with a sufficiently large number of subjects which can be compared with a subset of scores or a separate test with a smaller number of subjects:

- What is the minimum number of participants required to achieve statistically comparable outcomes to the test with all participants (using correlation and significance tests).
- Comparison of plots of the number of subjects versus the discriminability score.
- Comparison of plots of the number of subjects in method A versus method B to reach the same discriminability score.
5. References

[1] ISO/IEC JTC 1/SC29/WG1, "JPEG Pleno Final Call for Contributions on Subjective Light Field Quality Assessment", WG1N100307, Online, October 2022
[2] ISO/IEC JTC 1/SC 29/ WG 1, “Use Cases and Requirements for Light Field Quality Assessment v5.0”, WG1N100306, Online, October 2022
[6] https://x265.org
[7] https://bitbucket.org/multicoreware/x265
[10] https://ffmpeg.org
[14] https://kakadusoftware.com/
[17] Recommendation ITU-R BT.500-14 - Methodology for the Subjective Assessment of the Quality of Television Images