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: Common Test Conditions on Subjective Image Quality Assessment

SOURCE: WG1

EDITORS: Michela Testolina (michela.testolina@epfl.ch)
Evgeniy Upenik (evgeniy.upenik@huawei.com)
Nicola Giuliani (nicola.giuliani@huawei.com)
Dietmar Saupe (dietmar.saupe@uni-konstanz.de)

PROJECT: ISO/IEC 29170 (JPEG AIC)

STATUS: Approved

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
COMMON TEST CONDITIONS ON SUBJECTIVE IMAGE QUALITY ASSESSMENT
Table of Contents

1. Scope of this document ........................................................................................................................................ 4
2. Test material .................................................................................................................................................. 4
   2.1 Contents ................................................................................................................................................ 5
   2.2 Distorted images ...................................................................................................................................... 6
3. Contributions assessment ................................................................................................................................. 7
ANNEX A. Image coding information .................................................................................................................. 8
ANNEX B. Dataset generation procedure ........................................................................................................... 10
Bibliography ....................................................................................................................................................... 14
Common Test Conditions on Subjective Image Quality Assessment

1. Scope of this document

This document describes the Common Test Conditions (CTC) related to the JPEG AIC Call for Contributions on Subjective Image Quality Assessment [1]. The main objectives of this document are:

- Identify a dataset of images that is recommended, but not mandatory, to be used in the evaluation of the subjective quality assessment contributions
- Define a number of procedures that may be used in the assessment and comparison of the subjective quality assessment contributions

As the JPEG AIC Call for Contributions on Subjective Image Quality Assessment [1] aims at developing the standard by consensus among the JPEG experts following a collaborative process, contributors are welcome to revise this document.

2. Test material

This section describes the test materials that may be used in the evaluation of the contributions to the JPEG AIC CfC on Subjective Image Quality Assessment, namely the JPEG AIC-3 dataset. The JPEG AIC-3 dataset is available for download through FTP using the following credentials:

- Protocol: FTP
- FTP address: tremplin.epfl.ch
- Username: jpegaic@mmspgdata.epfl.ch
- Password: .L:p*0
- FTP port: 21

More information on the dataset and download procedure is available on the JPEG AIC-3 dataset's webpage: https://www.epfl.ch/labs/mmspg/downloads/jpeg-aic3-dataset/.

The pre-release of the dataset is scheduled for 2022-11-11, while the final release of the dataset is scheduled for 2022-12-15.

The dataset includes 10 different contents compressed with multiple codecs at multiple quality levels, in the range of interest to JPEG AIC [1]. Contributors to the JPEG AIC CfC may use the
JPEG AIC-3 dataset to evaluate the proposed subjective quality assessment contribution or may use a different dataset, which must be included as part of the contribution.

2.1 Contents

The JPEG AIC-3 dataset is composed of 10 uncompressed original images of different resolution and content:

1. Object
2. Human portrait
3. Food
4. Computer-generated content
5. Animal
6. Scene with water
7. Night scene
8. Fabric/fine texture
9. Landscape
10. Buildings

A preview of the images part of the JPEG AIC-3 dataset is provided in Figure 1.

![Figure 1. Original uncompressed images part of the JPEG AIC-3 dataset.](image-url)

The JPEG AIC-3 dataset has the following characteristics:

- CC0 licensing
- Format: PNG
- Color space: sRGB
- Spatial resolution: from 560x888 to 2592x1946 pixels

A summary of the characteristics of the original undistorted JPEG AIC-3 images is available in Table 1.
Table 1. Summary of the characteristics of the original undistorted images part of the JPEG AIC-3 dataset.

<table>
<thead>
<tr>
<th>IMAGE NUMBER</th>
<th>CONTENT</th>
<th>RESOLUTION (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00001</td>
<td>Object</td>
<td>1192x832</td>
</tr>
<tr>
<td>00002</td>
<td>Human face</td>
<td>853x945</td>
</tr>
<tr>
<td>00003</td>
<td>Food</td>
<td>945x840</td>
</tr>
<tr>
<td>00004</td>
<td>Computer-generated</td>
<td>2000x2496</td>
</tr>
<tr>
<td>00005</td>
<td>Animal</td>
<td>560x888</td>
</tr>
<tr>
<td>00006</td>
<td>Water</td>
<td>2048x1536</td>
</tr>
<tr>
<td>00007</td>
<td>Night scene</td>
<td>1600x1200</td>
</tr>
<tr>
<td>00008</td>
<td>Fabric/fine texture</td>
<td>1430x1834</td>
</tr>
<tr>
<td>00009</td>
<td>Landscape</td>
<td>2048x1536</td>
</tr>
<tr>
<td>00010</td>
<td>Buildings</td>
<td>2592x1946</td>
</tr>
</tbody>
</table>

2.2 Distorted images

This section describes the codecs employed to compress/decompress the original uncompressed images described in Section 2.1. Notably, the pre-release of the dataset will include images distorted using the following codecs:

- JPEG
- JPEG 2000
- HEVC Intra
- VVC Intra
- JPEG XL

The final release of the dataset will additionally include images distorted using the following codecs:

- JPEG AI
- AVIF

Image 00004 of the dataset (computer-generated content) is not available in HEVC Intra and VVC intra encoded form in the pre-release of the dataset, and will be made available in the final release of the dataset.

More information on the encoding of the images and on the selection of the distorted images is available in Annex A and B respectively.
3. Contributions assessment

Statistical analysis of the results of the different contributions will be performed, although not with the goal of ranking the contributions, but rather to obtain useful information which may be reviewed by JPEG experts during the development of the standard.

The different contributions may be compared to each other through statistical analysis which may include, but not limited to:

- **Correlation**: the correlation between the subjective quality scores obtained in different contributions may be computed, e.g. through Pearson linear correlation coefficient (PLCC) or Spearman rank order correlation coefficient (SROCC). Additionally, root-mean-square error (RMSE) and outlier ratio (OR) may be computed.

- **CD, FR, FD, FT**: correct decision (CD), false ranking (FR), false differentiation (FD), or false tie (FT), according to ITU-T Rec. J.149 [2], may also be computed.

- **Significance test**: a Student's t-test or Welch's test may be performed between subjective quality scores of multiple experiments to assess whether the means of the collected scores differ significantly.

- **CE, UE, OE**: the correct estimation (CE), over-estimation (OE), and under-estimation (UE) may be computed as in [3] to assess if the subjective quality scores are statistically distinguishable across the different contributions.

- **ANOVA**: a multi-way ANOVA may be performed on the subjective quality scores obtained from different contributions to verify whether the obtained subjective quality scores are statistically equivalent.

- **Size of confidence intervals (CI)**: The size of CI of the subjective quality scores of each stimulus may be computed and may be compared across multiple experiments.

- **Minimum number of subjects**: the minimum number of participants required to achieve statistically comparable outcomes to the same test conducted with an extensive number of participants may be assessed.
ANNEX A. Image coding information

1. JPEG

The following setup was adopted for the encoding/decoding of the JPEG AIC-3 dataset with JPEG:

- Implementation: [https://github.com/thorfdbg/libjpeg](https://github.com/thorfdbg/libjpeg)
- Version: commit 54ec64374a4991c7eadf05107389692267a459e1
- Command line example:

```
jpeg -h -qt 3 -s 1x1,2x2,2x2 -q [QUALITY_PARAMETER] [INPUT].ppm [ENCODED].bit
jpeg [ENCODED].bit [DECODED].ppm
```

2. JPEG 2000

The following setup was adopted for the encoding/decoding of the JPEG AIC-3 dataset with JPEG 2000:

- Implementation: [https://kakadusoftware.com/documentation-downloads/downloads/](https://kakadusoftware.com/documentation-downloads/downloads/)
- Version: v8.0.5
- Command line example:

```
kdu_compress -i [INPUT].ppm -o [ENCODED].bit Qfactor=[QUALITY_PARAMETER] -no_weights -tolerance 0 -full -precise -num_threads 1
kdu_expand -i [ENCODED].bit -o [DECODED].ppm -precise -num_threads 1
```

3. HEVC

The following setup was adopted for the encoding/decoding of the JPEG AIC-3 dataset with HEVC:

- Version: 16.20_SCM8.8
- Command line example:
4. VVC

The following setup has been adopted for the encoding/decoding of the JPEG AIC-3 dataset with VVC:

- Version: 11.1
- Command line example:

```
```

```
TAppDecoderStatic -d 10 -b [ENCODED].bit -o [DECODED].yuv
```

5. JPEG XL

The following setup has been adopted for the encoding/decoding of the JPEG AIC-3 dataset with JPEG XL:

- Implementation: [https://github.com/libjxl/libjxl](https://github.com/libjxl/libjxl)
- Version: v0.7.0
- Command line example:

```
cjxl [INPUT].png -q [QUALITY_PARAMETER] [ENCODED].jxl
```

```
djxl [ENCODED].jxl [DECODED].png
```
ANNEX B. Dataset generation procedure

This annex aims at reviewing the procedure adopted for the creation of the JPEG AIC-3 dataset.

1. Expert viewing session

An expert viewing session was conducted in order to collect information on the visual quality of decoded images, and used for selecting the distorted images part of the JPEG AIC-3 dataset. The expert viewing session was conducted remotely through the platform “QualityCrowd 2” [4,5].

The protocol adopted for the experiment is a variant of pair comparison (PC), where the subjects were asked to select the stimulus with the highest visual quality among two options, displayed side-by-side. The position of each stimulus and the order of the pairs were chosen randomly and were different for every subject. The question presented to the subjects was the following:

![Select the image with the highest quality](image)

Each compressed image was compared to its original and to all the other images compressed with the same codec, without considering the comparisons among different codecs. Therefore, 15 comparisons per codec per image were performed. Following this approach, the total number of comparisons that was performed for the experiment is 750. To avoid fatigue, the test was divided into two parts (approximately 1h long), where the first part included images 1 to 5 and the second part included images 6 to 10. The interface of the experiment is shown in Figure 2.

![Experiment Interface](image)

Figure 2. Example of experiment interface.
Prior the beginning of the experiment, a short training session was conducted to get the subjects acquainted with the format of the experiment and with the grading scale. Additionally, the size of the screen of each subject was tested before the experiment, being the main requirements for the monitor the following:

- Minimum size of 1920x1080
- Retina mode turned off

If any of these requirements was not respected, the subject could not proceed with the experiment.

Table 2 summarize the statistics of the subjects participating in the expert viewing session:

| Experts completing the viewing | 31 |
| Experts completing part 1 | 15 |
| Experts completing part 2 | 16 |
| Average age | 35.7 |
| Median age | 31 |
| Minimum age | 22 |
| Maximum age | 68 |
| Females | 9 (29%) |
| Males | 21 (68%) |
| Prefer not to disclose gender | 1 (3%) |

Table 2. Statistics of the subjects participating in the expert viewing.

2. Results analysis

A similar procedure to [6] was adopted for the processing of the results of expert viewing. Standard reconstruction was applied by maximum likelihood estimation according to the Thurstonian probabilistic model (Case V), where the “eba” R-Package containing the “thurstone()” function was used. The results were shifted so that the source image is at zero. Moreover, the results were scaled to JND units by dividing all scale values by \( \Phi(0.75) = 0.6745 \), where \( \Phi \) is the normal CDF. Accordingly, if two images are 1 JND unit apart, then the model predicts a 50% difference detection probability of difference by a random observer.

In some of the 50 image sequences, the worst-quality image won hardly any of the comparisons, leading to the zero-frequency problem. In order to regularize the outcome, the pair comparison matrix was initialized with a small constant at each entry. In this experiment, an initialization value of 0.1 was used, i.e., for each pair comparison a virtual “not sure” vote was introduced, weighted with a factor of 0.2.
Figure 3. Result of the expert viewing. Each row corresponds to a different source image, numbered from 1 to 10. Each column corresponds to one codec. The bigger colorful dots are the reconstructed quality scale values from the expert viewing, in JND units. The small black dots on the fitting curve are the selected samples for the JPEG AIC-3 dataset.
3. Quality range selection

A parametric curve was fitted to the subjective quality scores collected during the expert viewing session. In particular, the parametric curve was estimated as the sum of a linear section with a slope parameter and a logistic section with two parameters. The fitting curves are shown in Figure 3.

The selected minimum scale value is -2.5 JND, and consequently the scale interval [-2.5,0] was subdivided into 10 subintervals of equal 0.25 JND length. The corresponding selected quality levels (reported in Figure 3 as small black dots) were then mapped back to the integer quality levels for each of the five codecs.

As an inconsistency in the results was detected for image 00003 distorted with JPEG-1, alternative quality levels were sampled for the lowest quality levels, chosen by visual inspection by JPEG experts.

The selected quality parameters for each codec are reported on a table provided as part of the JPEG AIC-3 dataset.
Bibliography


[5] https://github.com/mmspg/qualitycrowd2.1