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 DNA Common Test Conditions version 2.0** |
| **SOURCE:** | ICQ Subgroup |
| **EDITORS:** | Marc Antonini and Touradj Ebrahimi |
| **PROJECT:** | JPEG DNA Exploration |
| **STATUS:** | Final |
| **REQUESTED ACTION:** | Discussion |
| **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

**Table of contents**

# JPEG DNA COMMON TEST CONDITIONS
version 2.0

## 1 Scope

The scope of JPEG DNA is the creation of a standard for efficient coding of images that considers biochemical constraints and offers robustness to noise introduced by the different stages of the storage process that is based on DNA synthetic polymers.

This document describes the Common Test Conditions (CTC) for the JPEG DNA image coding experiments. The main objectives of this document are:

- Define the common dataset that should be used in the evaluation of image coding solutions for storage on DNA support.
- Define the anchors (for direct encoding and transcoding) that should be used to comparatively evaluate the performance of image coding solutions for storage on DNA support.
- Define the coding conditions, in terms of source coding, error correction and biochemical noise simulators, as well as the target rates and compression ratios that anchors or alternative image coding solutions for storage on DNA support should demonstrate.
- Define the performance metrics for quality assessment that can be used to reliably evaluate the decoded images obtained from image coding solutions that produce streams in ACTG.
- Define the subjective evaluation procedure to perceptually evaluate all decoded images quality, namely the anchors and image coding solutions that produce ACTG.

In the current form, these common test conditions should be used to evaluate different aspects of image coding for storage on DNA support. The CTC is defined according to the use cases and requirements identified [17] and should be followed in all the experiments carried out by participants.

## 2 JPEG DNA Dataset

The JPEG DNA dataset is used for the performance evaluation of image coding solutions for storage on DNA support. This JPEG DNA dataset is freely available to all JPEG DNA proponents which will be submitted in the framework of exploration experiments. The JPEG DNA dataset is organized according to:

- **Uncompressed dataset:** The uncompressed dataset provides a set of 10 images to be used during the coding and decoding as defined in relevant exploration experiments.
- **JPEG compressed dataset:** The JPEG compressed dataset provides a set of already compressed images with 10 different quality levels to be used during transcoding as defined in relevant exploration experiments.

The diversity of the images in the JPEG DNA dataset is high, namely in terms of their characteristics, such as content, color, and spatial resolution. A preview of the images part of the JPEG DNA dataset is provided in Figure 1.



Figure 1: JPEG DNA uncompressed dataset

The uncompressed datasets have the following characteristics:

- Contents - object, human portrait, food, computer-generated image, animal, a scene with water, a night scene, fabric/fine texture, landscape, and buildings.
- Format – PNG images (RGB color components, non-interlaced), JPEG 1 (including progressive and hierarchical modes), JPEG 2000, and JPEG XL compressed.
- Spatial resolution – from 560x888 to 2592x1946 pixels. Detailed information on the different content and resolutions is provided in Table 1.

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

Table 1: Summary of the characteristics of the JPEG DNA uncompressed dataset

The JPEG DNA dataset is available for download through FTP using the following credentials:

```
Protocol: FTP
FTP address: tremplin.epfl.ch
Username: jpegdna@mmspgdata.epfl.ch
Password: fT76Dl
FTP port: 21
Folder name: 04-2023
```

The uncompressed dataset is available in folder 'uncompressed', while the JPEG compressed dataset is available in folder 'JPEG compressed'.

For the JPEG compressed dataset, the following naming convention is used:

<CODEC>_<IMG NUMBER>_<WxH>_<QUALITY>

where <CODEC> can be 'JPEG-1' or 'JPEG-2000', and <QUALITY> is a number between 1 and 10, where 1 indicates the best quality and 10 the worst quality.

## 3 Evaluation Procedure

Objective and subjective quality evaluation will be performed by at least two independent organizations, following well-established procedures, and based on the decoded test images provided by each proponent. Proponents may perform encoding in any color space representation, but the input of the encoder and the output of the decoder must be in the PNG (in Y or in RGB color space) format as defined in the datasets. Objective image quality will be measured with luminance and color-based metrics and the RGB decoded images will be used for subjective quality evaluation.

## 4 Target Rates and Compression Ratios

The rate will be reported in the following way:
- The rate is expressed by the number of nucleotides (nts) per pixel (nts/pixel).

An implementation of these rates can be found at: https://gitlab.com/wg1/jpeg-dna/jpeg-dna-metrics.

## 5 Objective Quality Evaluation

Objective quality testing shall be performed by computing several quality metrics, including $PSNR_Y$, $PSNR_{YUV}$, MS-SSIM, IW-SSIM, VMAF, VIFP, PSNR-HVS-M, NLPD and FSIM, between compressed and original images, at the target rates mentioned in the previous section. This section defines the objective image quality metrics that will be used for the assessment of image coding solutions. The reference implementation of $PSNR_Y$, $PSNR_{YUV}$ can be found at https://gitlab.com/wg1/jpeg-dna/jpeg-dna-metrics, and all other objective quality assessment metrics are available at: https://gitlab.com/wg1/jpeg-ai/jpeg-ai-qaf.

### 5.1 MS-SSIM Definition and Computation

Multi-Scale Structural SIMilarity (MS-SSIM) [1] is one of the most well-known image quality evaluation algorithms and computes relative quality scores between the reference and distorted images by comparing details across resolutions, providing high performance for learning-based image codecs. The MS-SSIM [1] is more flexible than single-scale methods such as SSIM by including variations of image resolution and viewing conditions. Also, the MS-SSIM metric introduces an image synthesis-based approach to calibrate the parameters that weight the relative importance between different scales. A high score expresses better image quality.

### 5.2 IW-SSIM Definition and Computation

Information Content Weighted Structural Similarity Measure (IW-SSIM) [2] is an extension of the structural similarity index based on information content weighted pooling. This metric assumes that when natural images are viewed, pooling should be made using perceptual weights proportional to the local information content. Moreover, advanced statistical models of the natural images are employed to derive the optimal weights combined with multiscale structural similarity measures to achieve the best correlation performance with subjective scores from well-known databases.

### 5.3 VMAF Definition and Computation

The Video Multimethod Assessment Fusion (VMAF) metric [3] developed by Netflix is focused on artifacts created by compression and rescaling and estimates the quality score by computing scores from several quality assessment algorithms and fusing them with a support vector machine (SVM). Even if this metric is specific to videos, it can also be used to evaluate the quality of single images and has been proven to perform reasonably well for learning-based image codecs. Since the metric takes as input raw images in the YUV color space format, the PNG (RGB color space) images are converted to the YUV 4:4:4 format using FFMPEG (BT.709 primaries). A higher score of this metric indicates better image quality.

### 5.4 VIF Definition and Computation

Visual Information Fidelity (VIF) [4] measures the loss of human-perceived information in some degradation processes, e.g. image compression. VIF exploits the natural scene statistics to evaluate information fidelity and is related to the Shannon mutual information between the degraded and original pristine image. The VIF metric operates in the wavelet domain and many experiments found that the metric values agree well with the human response, which also occurs for learning-based image codecs. A high score expresses better image quality.

## 5.5 PSNR-HVS-M Definition and Computation

The PSNR-HVS-M [5] is a simple and effective quality model which uses DCT basis functions and is based on the human visual system (HVS). The model operates with an 8x8 pixel block of an image and calculates the maximum distortion that is not visible due to the between-coefficient masking. The proposed metric, PSNR-HVS-M, considers the proposed model and the contrast sensitivity function (CSF).

## 5.6. PSNR-Y and PSNR-YUV Definition and Computation

The *PSNR* between the original component, *I*, and the reconstructed component, *I'*, (both n-bit) is computed as follows:

$$PSNR = 10 \ log_{10} \frac{(2^n - 1)}{MSE}$$

where the *MSE* between the two M×N images, *I* and *I'*, is given by:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i,j) - I'(i,j))^2$$

Once the *PSNR-Y*, *PSNR-U* and *PSNR-V* are individually computed, *PSNR-YUV* is computed using:

$$PSNR - YUV = \frac{6 \ PSNR-Y + PSNR-U + PSNR-V}{8}$$

## 5.7 NLPD Definition and Computation

The Normalized Laplacian Pyramid Distance (NLPD) is an image quality metric [14] based on two different aspects associated with the human visual system: local luminance subtraction and local contrast gain control. NLPD exploits a Laplacian pyramid decomposition and a local normalization factor. The metric value is computed in the normalized Laplacian domain, this means that the quality of the distorted image relative to its reference is the root mean squared error in some weight-normalized Laplacian domain. A lower score expresses better image quality.

## 5.8 FSIM Definition and Computation

The feature similarity (FSIM) metric [6] is based on the computation of two low-level features that play complementary roles in the characterization of the image quality and reflect different aspects of the human visual system: 1) the phase congruency (PC), which is a dimensionless feature that accounts for the importance of the local structure and the image gradient magnitude (GM) feature to account for contrast information. The color version of the FSIM metric will be used. A high metric value expresses better image quality.

## 6 Subjective Quality Evaluation

To evaluate the selected coding solutions, a subjective quality assessment methodology should be used. Subjective quality evaluation of the compressed images will be performed on the test dataset.

The Double Stimulus Continuous Quality Scale (DSCQS) methodology will be used, where subjects watch side by side the original image and the impaired decoded image, and both are scored on a continuous scale. This scale is divided into five equal lengths, which correspond to the normal ITU-R five-point quality scale, notably Excellent, Good, Fair, Poor, and Bad. This method requires the assessment of both original and impaired versions of each test image. The observers are not told which one is the reference image, and the position of the reference image is changed in pseudo-random order. The subjects assess the overall quality of the original and decoded images by inserting a mark on a vertical scale. The vertical scales are printed in pairs to accommodate the double presentation of each test picture.

The subjective test methodology will follow BT500.13 [7], and a randomized presentation order for the stimuli, as described in ITU-T P.910 [8], will be used; the same content is never displayed consecutively. There is no presentation or voting time limit. A training session should be organized before the experiment to familiarize participants with artifacts and distortions in the test images. At least, three training images will be used before actual scoring.

To perform the tests, a semi-controlled crowdsourcing setup framework and/or a more controlled lab environment procedure can be used to show the images according to the DSCQS methodology. The semi-controlled crowdsourcing setup has been proven in the past to be reliable, i.e. maintains a low variance of the scores [9]. The QualityCrowd2 [10] software and Amazon Mechanical Turk (or other similar platforms) will be used for crowdsourcing. The number of subjects will be large enough in order to conclude in a statistically meaningful fashion.

## 7 Biochemical Coding Constraints

DNA data storage is a very error-prone process. The different components of the biochemical process for DNA data storage, especially sequencing, generate a lot of errors. The error rates of these processes depend on the different technologies adopted by each sequencing machine but also largely depend on the DNA code that should be embedded in a molecule. Some patterns and other characteristics of the DNA codes have been identified as error-generating and should be avoided to make the entire data storage process functional and more reliable. This section describes those constraints that will be considered during this call for proposals. All codecs submitted to JPEG DNA call for proposals should comply with these constraints. Additional constraints will be considered after the conclusion of the call and in the framework of core experiments when a starting point has been decided.

To check codec compliance with the biochemical constraints, a series of compliance verification software should be run on the encoded streams. A detailed set of instructions on how to use such software and how to report the results will be provided to those proponents who will pre-register. Likewise, proponents will be provided with the necessary software and templates to use for their submissions.

The constraints that will be considered are listed below:

● **Strand length limitations**

- Definition

Since the handling of very long strands of DNA during synthesis can be challenging, the DNA codestreams need to be cut into shorter strands. A different DNA molecule will then be synthesized for each short strand which is also referred to as an oligo. The submitted codestream should therefore consist of a collection of strands of a maximum fixed length. Moreover, a strand index should be embedded in the strand's codestream to denote each strand's order to correctly reassemble the entire data stream for an image.

- Criterion

A satisfactory strand length should lie in an interval of 100 to 300 nucleotides. The maximum strand length is generally fixed for all the coded data for easy decoding.

For performance testing, a commonly accepted length for the encoded strands is fixed to a maximum of typically 200 nts. The final instructions provided to pre-registered proponents will inform them about the exact value used in the assessment of proposals.

- Scope

This criterion should be respected for every strand of the submitted codestream.

● **Homopolymer runs**

- Definition

A homopolymer run is the repetition of the same nucleotide several times in a strand. It is important to limit the number of such repetitions to produce stable molecules. Different sequencing technologies can tolerate different maximum lengths for homopolymers.

- Criteria

As a baseline, without any prior assumption on the sequencing technologies that will be used, the encoding should optimally respect the minimum threshold of tolerance so as to ensure the reliability of the decoding process. Therefore, all submitted encoders should not contain codewords with homopolymers of length greater than typically 3 and avoid generating homopolymers greater than typically 7 during encoding. The final instructions provided to pre-registered proponents will inform them about the exact values used in the assessment of proposals.

- Scope

These criteria should be respected in the DNA code, in the combination of the different DNA codewords, and in the different headers, indexes, and identifiers. A specific analysis software will be provided for checking the compliance of submitted codecs with the homopolymer constraints used in the assessment of the proposals.

● **GC content balance**

- Definition

The GC content describes the usage of the G and C bases in the strand. More specifically, the GC content describes the percentage of all the bases in the code that are either a G or a C and is given by the following formula:

$$GC\_content = \sum_{i=0}^{l} \frac{\delta(c[i],\{G,C\})}{l}$$

with $c[i]$ denoting the nucleotide at index $i$ in the strand $c$, $\delta(c[i],\{G,C\}) = \{1\ if\ c[i]\ in\ \{G,C\}, 0\ otherwise\}$ and $l$ referring to the length of the strand.

- Criterion

The typically acceptable interval in which the GC content should fall is between 40% and 60%. When over 50%, the error rate for nanopore sequencers significantly increases in some methods. Therefore a better interval could be between 40% and 50% [15]. The final instructions provided to pre-registered proponents will inform them about the exact value used in the assessment of proposals.

- Scope

This criterion should be respected for every strand of the submitted codestream.

● **Repetition of patterns**

- Definition

A short pattern is a sequence of nucleotides with a minimum length of 3 and a maximum length of 5 which is repeated many times consecutively in a strand.

The desired strand should be free from repetitions of patterns. For example, the strand ATCATCATC, where pattern ATC is repeated several times is not acceptable. Patterns should not repeat in any DNA strand as sequencing technologies perform much better when increasing variability in the codestream, it is important that the submitted encoders do not generate repetitions of the same short patterns [16].

- Criterion

Any pattern of length between typically 3 to 5 nucleotides should not be repeated consecutively more than typically 4 times. The final instructions provided to pre-registered proponents will inform them about the exact values used in the assessment of proposals.

- Scope

This criterion should be respected for every strand of the submitted codestream.

## 8 Errorless Anchors Generation

### 8.1 Transcoder Anchor 1
Compressing using legacy JPEG and followed by Goldman DNA coding [12]

The anchor 1 describes a coding method that consists in segmenting an already compressed JPEG bitstream into bytes and encoding each byte separately. The byte-wise encoder is composed of two blocks:

- The first block is a ternary Huffman encoder that considers all the possible values of a byte as an alphabet. Each byte value has a ternary (written with 0s, 1s and 2s) codeword associated with it. The length of the codewords for each byte value will depend on the frequency of appearance of that value in the bitstream. The more frequent that value, the shorter its corresponding codeword and the less frequent that value, the longer its corresponding codeword.
- The second block is a Goldman encoder: it encodes the ternary bases (0s, 1s and 2s) of the ternary Huffman codewords with a rotating {A, T, C, G} alphabet which results in quaternary DNA codewords.

The software implementation of this anchor will be provided to pre-registered proponents.

### 8.2 JPEG DNA Benchmark Codec (JPEG DNA BC) Anchor 2
The JPEG DNA Benchmark Codec in Python is an implementation of the coding algorithm of the same name. The main purpose of the JPEG DNA Benchmark Codec is to encode the DCT coefficients into DNA quaternary codes instead of binary (see Figure 1) while maintaining a coding strategy as close as possible to the legacy JPEG standard. This work has been published in [11]. First, a quantization is applied, followed by a zigzag scan, resulting in a sequence of integers in which the first element represents the DC coefficient, and the following 63 coefficients correspond to AC coefficients.
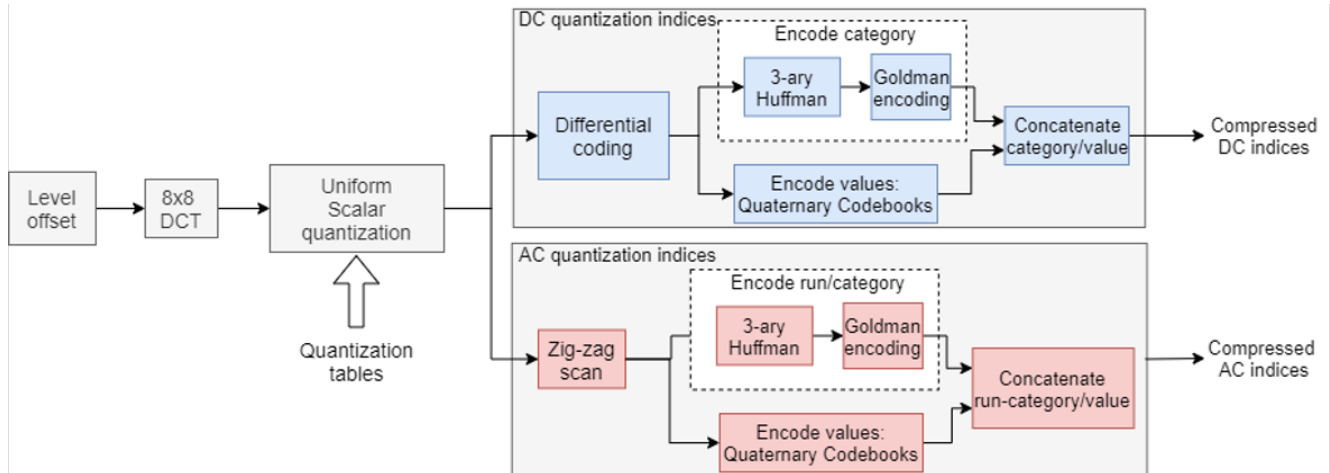
Figure 1: The JPEG DNA Benchmark codec.

The software implementation of this anchor will be provided to pre-registered proponents.

**8.3 JPEG DNA BC Transcoder Anchor 3**
Existing JPEG files can be losslessly transcoded to JPEG DNA BC and vice-versa.

The software implementation of this anchor will be provided to pre-registered proponents.

# 9 Experimentation Workflow

This section defines a generic procedure to conduct experiments on any codec for DNA data storage, that goes beyond what is assessed in the evaluation of submissions and describes those elements in the workflow that will be evaluated. The general workflow of those experiments is proposed in figure 2 and is composed of different tools that deal with specific processes. The main components of this workflow are:
- The codec (that is able to both encode the input image into a pool of formatted DNA sequences called strands and decode a pool of formatted strands into an image),
- The noise model (a simulator that alters an input pool of formatted strands by introducing errors – insertions, deletions, substitutions – approximating the behavior of the real biochemical processes – synthesis, storage, amplification, sequencing –),
- The filtering system or strand selector, that, from the noised strands, discards those strands that contain too many errors,
- The consensus, that from a set of erroneous strands will generate a pool of strands most likely to be those encoded.

For each process in each component, it is important to identify the external parameters that need to be adjusted to conduct thorough experiments. They are:
- The coding rate, determined during the encoding process,

- The length of the formatted strands, and the primers used for the strands during formatting,
- The noise level for all the components of the noise model (synthesis, storage degradation, amplification, and sequencing).

**In the context of this call for proposals, only the portion of the workflow with blue components is evaluated. It is, however, the intention of the JPEG standardization committee to set up various core experiments in order to evaluate the rest of the workflow after the selection of a starting point for a verification model.**
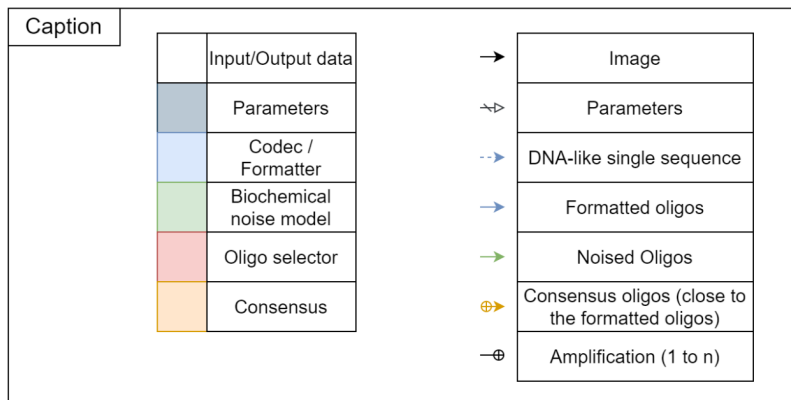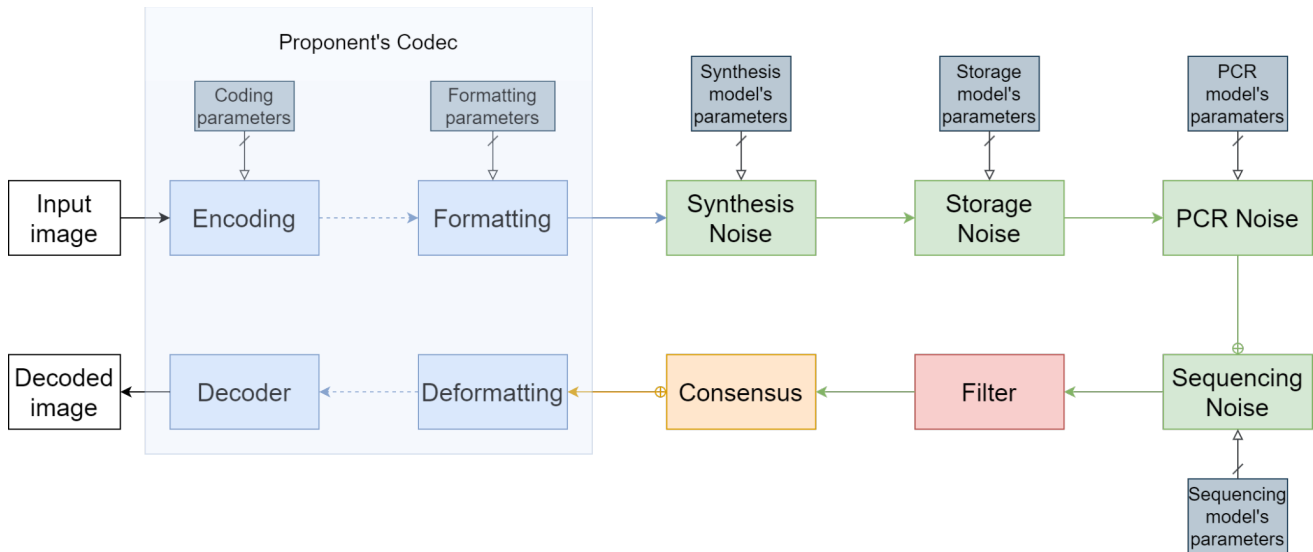


Figure 2: General workflow for image coding/decoding simulations.

# 10 References

[1]   Z. Wang, E. P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," ThirtySeventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, November 2003.

[2]   Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment", IEEE Transactions on Image Processing, Vol. 20, No. 5, May 2011.

[3]   Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy and M. Manohara, "Toward a practical perceptual video quality metric", [Online], Available here.

[4]   H.R. Sheikh and A. C. Bovik, "Image information and visual quality" IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, August 2004.

[5]   N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, "On between-coefficient contrast masking of DCT basis functions", Proceedings of the third international workshop on video processing and quality metrics (Vol. 4), 2007.

[6]   L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: a feature similarity index for image quality assessment," IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, August 2011.

[7]   ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Geneva, Switzerland, 2012.

[8]   ITU-T Recommendation P. 910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, 2008.

[9]   D. Saupe, F. Hahn, V. Hosu, I. Zingman, M. Rana, and S. Li, "Crowdworkers proven useful: a comparative study of subjective video quality assessment," International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, June 2016.

[10]  C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd: a framework for crowd-based quality evaluation," Picture Coding Symposium, Krakow, Poland, May 2012.

[11]  M. Dimopoulou, E. Gil San Antonio and M. Antonini, "A JPEG-based image coding solution for data storage on DNA", European Signal Processing Conference (EUSIPCO), Aug. 2021.

[12]  N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E.M. LeProust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA", Nature, vol. 494, no. 7435, pp. 77–80, 2013.

[13]  M. Dimopoulou, M. Antonini, P. Barbry and R. Appuswamy, "A biologically constrained encoding solution for long-term storage of images onto synthetic DNA", 27th European Signal Processing Conference (EUSIPCO), Sep. 2019.

[14]  V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, "Perceptual Image Quality Assessment using a Normalized Laplacian Pyramid", S&T Symposium on Electronic Imaging: Conf. on Human Vision and Electronic Imaging, San Francisco, CA, USA, February 2016.

[15]  C. Delahaye and J. Nicolas, "Sequencing DNA with nanopores: Troubles and biases", PLOS ONE 16(10): e0257521, 2021

[16]  T. J. Treangen and S. L. Salzberg, "Repetitive DNA and next-generation sequencing: computational challenges and solutions", Nature Reviews Genetics, vol. 13, no. 1, p. 36, 2012

[17]  ISO/IEC JTC 1/SC29/N100252, Use Cases and Requirements for DNA-based Media Storage version 1.0, 96th Meeting, Online, 25-29 July 2022.