This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3096059, IEEE Transactions on Multimedia



# Quality evaluation of holographic images coded with standard codecs

Hadi Amirpour, *Student Member, IEEE*, Antonio M. G. Pinheiro, *Senior Member, IEEE,* Elsa Fonseca, Mohammad Ghanbari, *Life Fellow, IEEE* and Manuela Pereira

***Abstract*—Recently, more interest in the different plenoptic formats, including digital holograms, has emerged. Aside from other challenges that several steps of the holographic pipeline, from digital acquisition to display, have to face, visual quality assessment of compressed holograms is particularly demanding due to the distinct nature of this 3D image modality when compared to regular 2D imaging.**

**There are few studies on holographic data quality assessment, particularly with respect to the perceptual effects of lossy compression. This work aims to study the quality evaluation of digital hologram reconstructions presented on regular 2D displays in the presence of compression distortions. As there is no established or generally agreed on compression methodology for digital hologram compression on the hologram plane with available implementations, a set of state-of-the-art compression codecs, namely, HEVC, AV1, and JPEG2000, were used for compression of the digital holograms on the object plane. Both computer-generated and optically generated holograms were considered.**

**Two subjective tests were conducted to evaluate distortions caused by compression. The first subjective test was conducted on the reconstructed amplitude images of central views, while the second test was conducted on pseudovideos generated from the reconstructed amplitudes of different views. The subjective quality assessment was based on mean opinion scores. A selection of objective quality metrics was evaluated, and their correlations with mean opinion scores were computed. The VIFp metrics appeared to have the highest correlation.**

***Index Terms*—Digital holography, perceived quality, MOS, codecs**

## I. Introduction

DIGITAL holography (DH) is a three-dimensional imaging technique where the coherent superposition between the optical field scattered by an object and a suitable reference beam is digitally recorded [1], [2]. The resulting hologram is then illuminated with the same reference beam to yield the captured object field along with all its properties: light intensity, parallax, and depth. The interference pattern can be obtained by using either an optical or an electronic setup for generating and capturing the optical signal or by numerically simulating the object wavefield using computer-generated holography algorithms [3]. The first case yields the optically generated hologram (OGH), while the second case yields the well-known computer-generated hologram (CGH). A broad range of DH applications include data storage[4], security [5], medical imaging [6], microscopy [7], deformation/displacement measurements [8] and, inevitably, 3D displays. [9], [10]. Despite their limited resolution and viewing angle, current holographic displays can already convey full 3D holographic information.

The capability of rendering high-quality images enabling different perspectives of a three-dimensional scene within a wide range of viewing angles comes at the cost of recording and storing considerable quantities of data. Therefore, image compression techniques must be applied to enable future broadcasting of holographic-based images and videos. However, the compression of holographic data with current state-of-the-art methods for 2D contents has limited performance due to the distinct nature of the interferometric patterns that convey holographic information.

Different coding solutions have been proposed to encode holographic data [11], [12], [13], [14]. Hologram compression with the application of image and video coding standards was previously studied, both on the hologram plane [12] and object plane [15]. The quality evaluation of these coding standards was limited to the PSNR computation in both planes. In fact, there were no quality studies for the validation of any quality metric.

Although subjective quality assessments of compressed light fields [16] and point clouds [17] have been studied previously, the subjective quality assessment of compressed DHs is less exploited. However, subjective quality assessment is required since, ultimately, it provides the only trustworthy perceptual evaluation of a given coding solution. The effects of compression on the reconstruction quality of experimental holograms have been assessed in [18]. Ahar *et al* proposed a subjective quality assessment of 2D CGH reconstructions for comparing compression techniques [19]. In a recent study, it was shown that subjective tests on 2D, light field, and holographic displays have very high correlations [20]. In recent papers [21], [22]

Hadi Amirpour, Antonio M. G. Pinheiro and Manuela Pereira are with Instituto de Telecomunicações and Universidade da Beira Interior, Covilhã, Portugal. e-mail: (mpereira@di.ubi.pt).

Elsa Fonseca is with Fiber Materials and Environmental Technologies (FibEnTech) and Universidade da Beira Interior, Covilhã, Portugal. e-mail: (efonseca@ubi.pt).

Mohammad Ghanbari is a Professor at the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran (e mail: ghan@ut.ac.ir), as well as Emeritus Professor at the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, CO4 3SQ, (e-mail: ghan@essex.ac.uk).

This research was funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia and cofunded by FEDERPT2020 partnership agreement under the project PTDC/EEI-PRO/2849/ 2014 - POCI-01-0145-FEDER-016693, under the project UIDB/EEA/50008/2020, under the project UIDB/EEA/50008/2020, PLive X-0017-LX-20 and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de CompetÌncias em Cloud Computing, cofinanced by the European Regional Development Fund (ERDF) through the Programa Operacional Regional do Centro (Centro 2020), in the scope of the Sistema de Apoio  Investigao Cientfica e Tecnolgica - Programas Integrados de IC&DT.

Manuscript received x; revised y.

focused on the current status of the JPEG Pleno Holography standardization process in which the testing methodology for objective and subjective evaluation was analyzed. Recently, a study used subjective and objective evaluation methods for the performance analysis of several speckle noise removal techniques [23] applied to numerical reconstructions of OGHs. Although it would be of interest to assess the 3D properties of the reconstructed object fields, the abovementioned studies only considered the assessment of a single 2D perspective from each hologram, corresponding to the full field of view centered about the hologram normal. Aiming at shedding some light on this subject, a recent study on the objective evaluation of the performances of state-of-the-art compression techniques considered the reconstruction of different perspectives and focusing distances of digital holographic data [24].

Therefore, it is important to extend these studies to a subjective assessment of the three-dimensional properties of lossy compressed digital holograms. Ideally, it is preferable to have a complete 3D perceptual evaluation on a holographic display. However, current holographic display technology is still limited in terms of resolution and field of view and is quite expensive. Thus, we propose a compromise solution where a sequence of multiple views, in the form of a short pseudovideo, is presented to the observer. This procedure is intended to mimic the observer's behavior when looking at a real holographic display similarly to the light field [16] and point cloud [25] assessment techniques.

This paper describes our studies for the quality assessment of DHs. First, an initial subjective test based on the central view of holograms, obtained from the amplitude of the reconstructed complex optical field, is described. Then, a second subjective test that pursues 3D information and is based on the evaluation of a video sequence composed of different views, similar to the model of light-field quality assessment [16], is reported. The outcomes from both subjective tests are expressed in terms of mean opinion scores (MOS) and were designed using a double-stimulus impairment scale (DSIS). Next, the evaluation of some well-known and commonly used objective metrics with subjective scores is computed to establish the metrics that provide the best quality representation for DHs. Finally, some concluding remarks are given.

## II. Quality assessment

As briefly referred to in the introduction, this work reports two subjective evaluations of DHs. The first one (test 1) intends to analyze the quality of the reconstructed holograms after compression in the object plane using conventional image codecs. The second (test 2) was designed to overcome the limitations of the first one by considering the 3D nature of the data.

As volumetric displays are still in their infant years, they do not offer the appropriate means for plenoptic data representation suitable for the appropriate evaluation of perceptual quality. JPEG Pleno [26] created subjective evaluation protocols for the different modalities, namely, light fields, point clouds and digital holography. In particular, light fields were evaluated using video sequences of the different views with a predefined order [16]. The mean opinion scores of the video visualization were considered to be the quality measure of the complete light field. This process was tested with different scenarios as part of the JPEG Pleno standardization initiative [26] and proved to provide an accurate representation of the perceptual quality. Considering the success of the light-field subjective quality evaluation protocol, the visualization of the central view is complemented by using a pseudovideo sequence with the reconstructed views. A set of hologram views is reconstructed using a sliding window across different areas of the original hologram. This approach enables the evaluation of the preservation of parallax quality in compressed holograms.

The DSIS test method [27] was adopted with a 5-level impairment scale (1 - *very annoying*, 2 - *annoying*, 3 - *slightly annoying*, 4 - *perceptible, but not annoying*, 5 - *imperceptible*) for the subjective tests. The test data was defined as explained in the following. 1) A bit rate that is evaluated with a quality level 5 or 4 as it almost shows any imperceptible difference with the reference, 2) a bit rate that causes very low quality and it is likely to be rated with the lower quality, and 3) two bit rates representing middle-quality levels. It should be noted that the best level is very unlikely to be rated 5 because of the speckle noise influence. Because of that, it does not make much sense to add another intermediate quality level that tends to complicate rating discrimination by the subjects.

The following sections describe the design of the subjective evaluation in detail. Figure 1 represents the general coding scheme and paths for both subjective tests. Details are discussed in the following.

### A. Datasets and coding conditions

The open access datasets selected for this work consist of four OGHs from the EmergImg-HoloGrail[1] database [28] and eight CGHs from the INTERFERE-II[2] database [29], [30]. The set of OGH is fixed for both subjective tests and was acquired with a digital holographic phase-shifting setup using a He-Ne laser at a wavelength of 632.8 nm. Three OGHs were used for the subjective test sessions, namely, *Astronaut*, *Dice1*, and *Skull*, while the *Car* hologram was used for the training each subject had prior to a test. The CGH dataset consists of two subsets of holograms, 4 specular and 4 diffuse holograms, that were generated from 3D point clouds using the multiple parallel wavefront recording planes (WRP) algorithm [30]. The first set of holograms (specular) is better tailored to the propagation of a full resolution central view. The second set of holograms (diffuse) enables the propagation of different views obtained from smaller windows centered about different points on the hologram plane. However, this multiview rendering capability comes at the cost of increased speckle noise [30]. The specular CGH subset used for the first subjective test (test 1) includes the *Ball8KS*, *Cat8KS*, and *Chess8KS* holograms plus the *Earth8KS* hologram for the training session. For the second subjective test (test 2), the diffuse hologram versions replaced the specular versions, namely, the *Ball8KD*, *Cat8KD*,

[1]http://emergimg.di.ubi.pt
[2]http://erc-interfere.eu

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3096059, IEEE Transactions on Multimedia



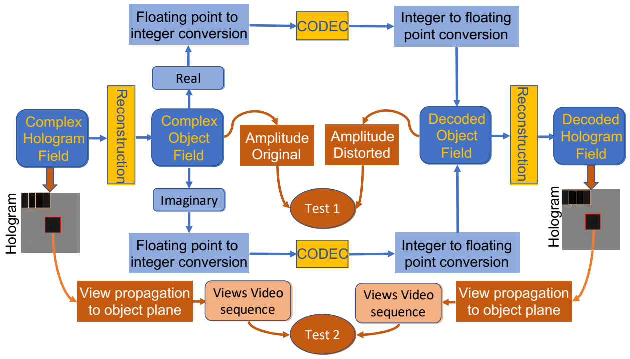
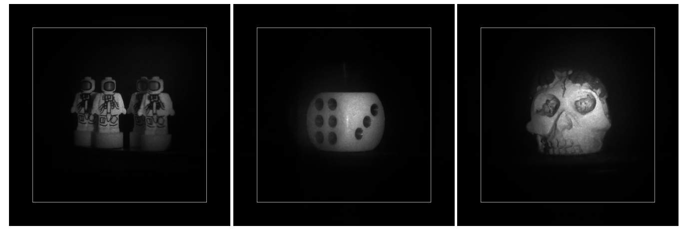
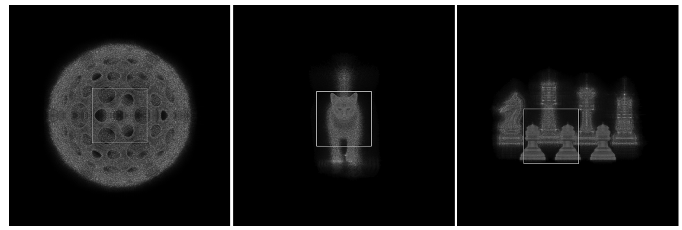



Fig. 1. Methodology used for subjective tests 1 and 2



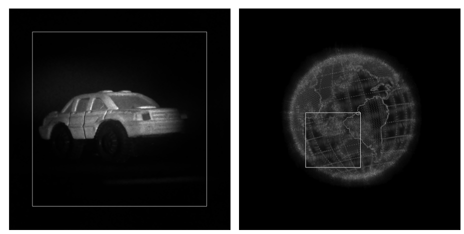
(a) *Astronaut* (b) *Dice1* (c) *Skull*

Fig. 2. Selected OGH from EmergImg-HoloGrail-v2.



(a) *Ball8KS* (b) *Cat8KS* (c) *Chess8KS*

Fig. 3. Selected CGH from the Interfere-II database (with the representation of the selected cropped areas).



(a) *Car* (b) *Earth8KS*

Fig. 4. Holograms used in the training session. On the left, an OGH from EmergImg-HoloGrail-v2 and on the right, a CGH from the Interfere-II database with the representation of the selected cropped areas.

and *Chess8KD* holograms plus the *Earth8KD* hologram also used in the training session.

The specifications of the OGHs and CGHs selected from the two abovementioned databases are described in Table I. Both datasets were generated at a wavelength of 633 nm. The angular full field of view, computed from the pixel pitch and recording wavelength [10], is approximately 16.5° for the OGHs and 36.9° for the CGHs. However, the apparent parallax is narrower in the OGHs case since the recording distance is approximately 10× higher. The numerical reconstructions at the given reconstruction distances (table I) of the selected holograms are represented in Fig. 2 and Fig. 3 for the EmergImg-HoloGrail and INTERFERE-II databases, respectively. The two holograms used for the subjective evaluation training are represented in Fig. 4. Cropped areas of 2,038×2,038 pixels were selected from all holograms to allow a true resolution display of their original and distorted versions, side-by-side, in a 4K display. Those areas are represented by the box in the pictures.

For hologram coding, three state-of-the-art codecs were selected, namely, HEVC, AV1, and JPEG2000. The latest reference software HM-16.20, the AOMedia Project AV1 Encoder 1.0.0-1001-gf5c9213e7 and Kakadu implementation were selected for HEVC, AV1 and JPEG2000, respectively. Selected configuration parameters for the encoders are summarized in Table II. The holograms were compressed after propagation to the object plane. Compression with these codecs on the hologram plane was not considered because it would result in loss of high frequencies, thus leading to significant errors in the reconstruction domain [28], [22]. The quantization of high frequencies in the hologram plane creates visual artifacts in the object plane as these frequencies appear propagated in the complete spectrum. One example of these artifacts is the aliasing created by the subsampling mechanism of JPEG 2000 when applied in the hologram plane [10]. Furthermore, to overcome the limitations of AV1 for high resolution, the Interfere holograms were first divided into three slices and then encoded.

### B. Data generation for subjective test 1

The holograms were propagated to the object plane, where the complex amplitudes were represented as real and imagi-

TABLE I
HOLOGRAM SPECIFICATIONS FOR THE SELECTED HOLOGRAM.

| Name | Reconstr. Distance (m) | Pixel Size ($\mu m$) | Resolution (pixel) |
|---|---|---|---|
| Astronaut | 0.1721 | 2.2 | $2,588 \times 2,588$ |
| Dice1 | 0.1721 | 2.2 | $2,588 \times 2,588$ |
| Skull | 0.1690 | 2.2 | $2,588 \times 2,588$ |
| Car | 0.1563 | 2.2 | $2,588 \times 2,588$ |
| Ball8KS and Ball8D | 0.0125 | 1.0 | $8,192 \times 8,192$ |
| Cat8KS and Cat8KD | 0.0142 | 1.0 | $8,192 \times 8,192$ |
| Chess8KS and Chess8KD | 0.0140 | 1.0 | $8,192 \times 8,192$ |
| Earth8KS and Earth8KD | 0.0118 | 1.0 | $8,192 \times 8,192$ |

TABLE II
SELECTED SETTINGS FOR THE ENCODERS

| Encoder | Configuration parameters |
|---|---|
| HEVC | -i <input> -wdt <width> -hgt <height><br>--InputBitDepth=10 --InternaBitDepth=10<br>--OutputBitDepth=10 --InputChromaFormat=400<br>-q <QP> -b <bitstream> -o <output> |
| AV1 | <input> --good -w <width>-h <height>--codec=av1<br>--end-usage=q -o <output> --tune=psnr --aq-mode=0<br>--threads=8 --cpu-used=0 --cq-level=<QP><br>--monochrome --input-bit-depth=10 --bit-depth=10 |
| JPEG2000 | -i <input> -o <output> -rate <rate> |

nary parts. Holographic data cannot be directly handled by the standard codecs since these codecs operate only on nonnegative real-valued numbers with integer precision. Henceforth, complex-valued data were split into real/imaginary parts, and floating-point data were mapped to a 10-bit integer representation before encoding. The inverse process was immediately performed after decoding. Here, floating-point-to-integer conversion was performed using a uniform quantizer.

The real and imaginary parts of each hologram in the object plane were encoded independently. Then, the amplitudes of the reference and the decoded hologram, both in the object plane, were used for testing. The quality assessment structure is represented in the scheme of Fig. 1 relative to test 1.

The quality of the HEVC and AV1 encoding was controlled using the q- and cq-level parameters, respectively. For HEVC, the q parameter was set to {22, 27, 29, 32} for EmergImg-HoloGrail database holograms and to {22, 32, 42, 51} for Interfere database holograms. For AV1, the cq-levels that better matched the bitrate obtained with HEVC were selected. This resulted in cq values of {22, 32, 42, 51} for the EmergImg-HoloGrail database holograms. For the Interfere database holograms, the cq values were {16, 35, 52, 63} for *Ball8KS*, {14, 33, 50, 61} for *Cat8KS*, and {15, 34, 51, 63} for *Chess8KS*. To set the quality of JPEG2000, the set of bitrate values resulting from HEVC coding was used.

Since four different quality levels were considered, a total of 72 stimuli were generated for the subjective test session.

### C. Data generation for subjective test 2

As described in the introduction, the second subjective test was motivated by the need to understand how the 3D hologram information is preserved after compression in the object plane. Holograms are reconstructed using a sliding synthetic aperture. This procedure allows the evaluation of the parallax quality of the hologram.

In test 2, the same OGHs from EmergImg-HoloGrail and the diffuse versions of CGH from the Interfere database were selected. The real and imaginary parts of the full resolution holograms were independently compressed in the object plane. Then, to obtain several views, the following two-step procedure is performed: first, the full resolution reconstruction is uncompressed and backpropagated to the hologram plane. Then, the windows corresponding to several perspectives are selected and propagated to the object plane to render a multiview sequence (see Fig. 1, relative to test 2).

For evaluation, several views were obtained for each hologram, and a pseudovideo was created with all views disposed in sequence from the top-left view to the bottom right view using a serpentine scan.

To avoid jumps between views, $7 \times 7$ overlapping views were generated. Each presentation consists of a pseudovideo comprising 49 frames at a frame rate of 10 fps, thus running in 4.9 seconds. Examples of the corner views generated for the *Cat8KD* hologram can be seen in Fig. 5. In the 6 original holograms, considering the 3 codecs and the 4 coding parameters, a total of 72 stimuli were generated, corresponding to 72 pseudovideos.

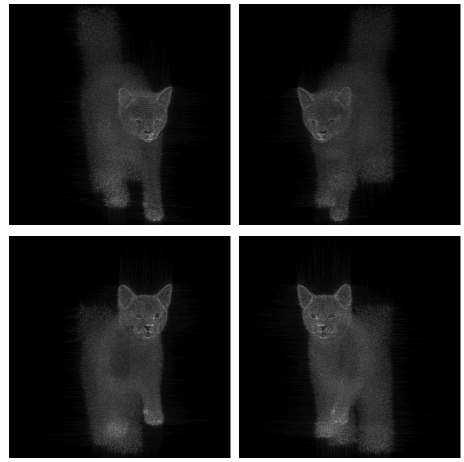


Fig. 5. Corner views generated for the Cat8KD hologram

The codec parameters were set the same as in test 1 for the OGHs. For the CGHs, the HEVC q and AV1 cq parameters were set as: for *Ball8KD* q∈{21, 26, 34, 38} and cq∈{33, 38, 45, 50}; for *Cat8KD* q∈{21, 26, 30, 34} and cq∈{27, 32, 37, 43}; and for *Chess8KD* q∈{21, 26, 34, 38} and cq∈{31, 35, 43, 50}.

JPEG 2000 quality was adjusted to result in the bit rate that resulted in HEVC, as in test 1.

### D. Subjective Assessment Methodology

The DSIS test method [27] was adopted with a 5-level impairment scale. Both the reference and the degraded stimulus were simultaneously shown to the observer, side-by-side. Every subject rated the visual quality of the processed stimulus with respect to the reference stimulus. To avoid bias, half of the individual evaluations had the reference placed on the right side and the degraded content on the left side of the screen, while their positions were swapped for the remaining half.

The sessions were conducted in the UBI test laboratory, which fulfills the recommendation for subjective evaluation of visual data issued by ITU-R BT.500 [27]. The test laboratory room was equipped with a dim lighting system of 5,500 K of correlated color temperature. The test stimuli were displayed on a 31.1" EIZO CG318 4K monitor with full 4k resolution (4,096 × 2,160 pixels) and maximum luminance 350 $\mathrm{cd\,m^{-2}}$. Prior to the tests, an informed consent form was handed to the participants for signature. Oral instructions were provided to explain the evaluation task. All subjects were screened for corrected visual acuity using a Snellen acuity chart.

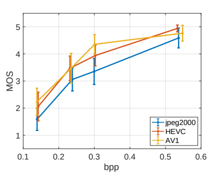
The participants were seated in front of the monitor perpendicular to the center at a distance of approximately 3 times the stimulus height. However, every subject was free to move closer or further away from the screen during the evaluation. To familiarize the subjects with the artifacts under assessment, each evaluation began with a short training session using 2 degraded versions (one with a very annoying distortion and another a slightly annoying distortion) and the original versions of the two holograms in Fig. 4. This session allowed subjects to become familiar with speckle noise and the visual appearance of different types of holograms prior to the subjective evaluation. The complete individual subjective test 1 took no longer than 10 minutes, while subjective test 2 took no longer than 15 minutes.

## III. Results and Discussion

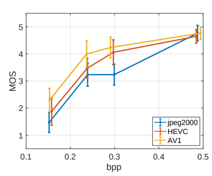
The average MOS ratings and the corresponding 95% confidence intervals (assuming a Gaussian distribution) for the OGH and CGH datasets were computed for both subjective tests. No outliers were found according to the recommendation ITU-R BT.500. In the following, the results of the two subjective tests, as well as a comparison between the two tests were reported.

### A. Subjective results of test 1

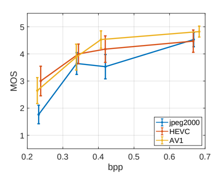
The results of subjective test 1 are shown in Fig. 6. A total of 17 individuals, 9 males and 8 females with a mean age of 24 years old, participated in the test. It can be observed that the different encoders reveal a more similar and monotonic behavior for the CGHs than for the OGHs. Furthermore, JPEG2000 presents the lowest performance for the OGHs, while it shows competitive performance with HEVC and AV1 for CGHs.



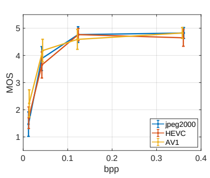

(a) *Astronaut*



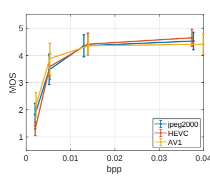

(b) *Dice1*



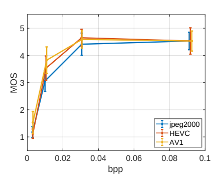

(c) *Skull*




(d) *Ball8KS*

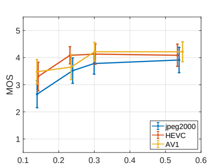



(e) *cat8KS*

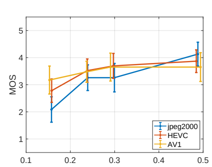



(f) *Chess8KS*

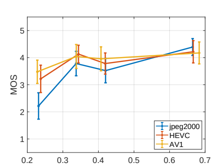
Fig. 6. MOS vs. bit rate (test 1) for OGH (top) and CGH (bottom).



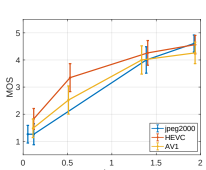

(a) *Astronaut*



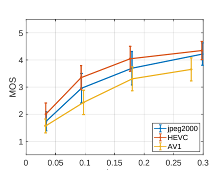

(b) *Dice1*



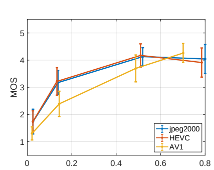

(c) *Skull*




(d) *Ball8KD*




(e) *cat8KD*




(f) *Chess8KD*

Fig. 7. MOS vs. bit rate (test 2) for OGH (top) and CGH (bottom).

### B. Subjective results of test 2

The results for subjective test 2 are shown in Fig. 7. A total of 23 individuals, 17 males and 6 females with a mean age of 22 years old, participated in the test. AV1 and HEVC encoders offer similar compression performance, while JPEG2000 reveals a lower performance for OGHs. For the higher bit rate, all the encoders result in a similar performance. In the case of CGHs, HEVC reveals the best results. JPEG2000 presents similar performance to HEVC for *Chess8KD*. The performance of AV1 is in general lower for CGHs.

### C. Discussion of subjective results

Subjective test 2 results in lower MOS values than subjective test 1. In particular, the CGHs require bit rates per pixel multiplied by a factor of 10 to obtain similar quality values in test 2. This is explained because test 2 used diffuse CGHs instead of the specular CGHs used in test 1, as explained in section II-A. As the diffuse CGHs used in test 2 cause larger speckle noise after reconstruction, the efficiency of the decoder is reduced, requiring a larger bit rate per pixel for similar qualities.

Moreover, other differences can be explained by: 1) For higher bitrates, it is difficult to judge whether two videos visualized simultaneously are identical in contrast to static images, which can explain the larger MOS observed in test 1. 2) For lower bitrates, artifacts due to coding are in general less visible in a video sequence than in static images, which explains why MOS tends to be larger in test 2.

Finally, for OGHs, it was observed that the two middle bit rates reveal a certain erratic behavior where, occasionally, a larger bit rate may yield a smaller MOS. This is caused by the prevalence of speckle noise, which tends to be removed for lower bit rates by the compression process. As the speckle is further removed for lower bitrates, then the details of those images are perceived with higher quality (see Fig. 8 for test 1 Skull OGH compressed with different bitrates). This effect is better observed in test 2, where the comparison between reference and distorted videos is more difficult to quantify than between the static images of test 1.

### D. Comparison between test 1 and test 2

To compare the methodologies used in the two tests, test 1 MOS is taken as ground truth and compared to MOS results of test 2 and the correlation between both was computed.

The Pearson linear correlation coefficient (PLCC) is used to measure accuracy, and Spearman rank-order correlation (SROCC) is used to measure monotonicity.

Table III shows the performance for the EmergImg-HoloGrail and Interfere datasets and the complete dataset.

In Fig. 9, the MOS scores of test 1 vs. the MOS scores of test 2 are plotted. A concave logistic curve is observed for the OGHs, while CGH's logistic curve is convex, which reveals the difference between the two types of holograms. In (a), all results are represented, while in (b), only the OGH results are shown. The OGHs are represented separately because they

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3096059, IEEE Transactions on Multimedia



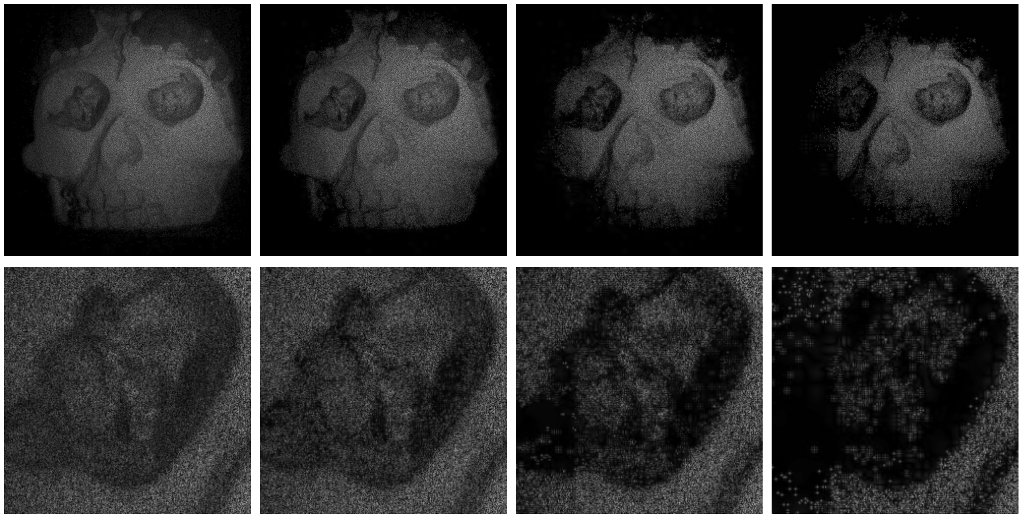
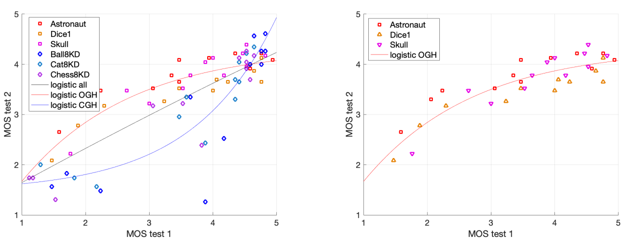


Fig. 8. Example of decoded reconstructed central view of the Skull hologram; from left to right, a higher to lower bitrate. Top: reconstructed holograms; Bottom: cropped view of the left eye.




(a) All (b) OGH

Fig. 9. Subjective scores of test 1 vs subjective scores of test 2.

use the same holograms, while for CGH, the specular versions were used for test 1, and the diffuse versions were used for test 2. Hence, it is possible that the comparison between test 1 and test 2 is influenced by that difference. However, the authors believe that the specific properties of the CGH (generated from point clouds) caused the difference in the behavior between the OGH and CGH. It is important to emphasize the point nature of the CGH that can be observed when the holograms are displayed without any resolution reduction. In both tests, the CGHs were cropped due to the large dimensions.

From the correlation results (which were higher than 80%), it is concluded that although the central image quality analysis is a good indicator of quality caused by the compression mechanism, a complete and more reliable analysis is also required for the assessment of several image views.

### *E. Performance of objective metrics*

Five objective metrics were selected, namely, PSNR, SSIM [31], MS-SSIM [32], FSIM [33], and VIFp [34]. The first two were used in similar studies [19], while the latter presented good performance in several quality studies [35].

To evaluate the performance of objective metrics in estimating the perceived quality, the MOS values were taken as the ground truth and compared to each objective metric results [35], [36].

To assess the accuracy of each objective metric, first, a logistic function was used to provide a nonlinear mapping between the objective and subjective scores according to ITU-R BT.500-13 [27]. Thereafter, PLCC and root mean square error (RMSE) were used to measure the accuracy of objective metrics. The SROCC was used to measure monotonicity, and the outlier ratio (OR) was used to estimate the consistency of objective metrics. As OGH and CGH have contents with different natures, objective metrics show different performances for each category. The correlations were computed separately for the two types of holograms and for the complete dataset.

*1) Subjective test 1:* Table IV shows the performance of the five selected metrics for the OGH and CGH and for the entire dataset. The metric VIFp shows a higher correlation with MOS values for both the OGH and CGH databases. The PSNR also reveals a good representation for the CGH. However, it shows a bad representation for the OGH, which can be explained by the predominance of speckle noise on the reconstructed images. No metric reveals a good perceptual quality representation when both types of holograms were considered, as they have specific properties that influence the perceived quality.

In Fig. 10, the MOS scores vs. objective metrics are plotted. The logistic curve used for the correlation computation is also shown.

*2) Subjective test 2:* The performance of the five selected metrics for the OGH and CGH datasets are shown in Table V. The results for the entire dataset are also presented.

As in test 1, objective metrics show different performances for each content. VIFp shows a higher correlation with MOS values for both the OGH and CGH databases. Once again, no metric reveals a good perceptual quality representation for both types of holograms.

In Fig. 11, the MOS scores vs. objective metrics are plotted. The good performance of the VIFp metric is confirmed when different types of holograms are considered separately.

*3) VIFp metric models:* The VIFp metric provides the best representation of the subjective results for both studies.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3096059, IEEE Transactions on Multimedia



However, it produces different models for the two types of studied sources, OGH and CGH. The different nature of the holograms is the cause of this behavior. The OGHs have a continuous nature, while the CGHs have a discontinuous nature, as they were generated from point clouds. Moreover, the two types of holograms have different pixel pitches and parallaxes.

## IV. Conclusion

Quality assessment methodologies of numerically reconstructed holograms compressed on the object plane were presented. The HEVC, AV1, and JPEG2000 codecs were used to compress the complex field of the holograms in the object plane at similar bitrates for both OGH and CGH. The subjective testing methodologies indicated that although evaluating the reconstructed central image reveals an approximation of the hologram quality, a detailed quality analysis requires the generation of multiple views. This analysis allows the quality evaluation of the 3D information representation provided by the digital hologram after compression on the object plane. It was also concluded that the VIFp metric is more appropriate for the quality evaluation of compressed digital holograms than the commonly used PSNR and SSIM. Furthermore, the specific differences between OGH and CGH lead to the conclusion that different acquisition processes require separate quality models.

TABLE III
Statistical analysis of correlation between Subjective scores of test 1 and test 2.

| | OGH | CGH | All |
|---|---|---|---|
| PLCC | 0.8602 | 0.8629 | 0.8044 |
| SROCC | 0.8187 | 0.8828 | 0.8183 |

TABLE IV
Statistical analysis of objective metrics (test 1).

| | | PLCC | SRCC | RMSE | OR |
|---|---|---|---|---|---|
| OGH | PSNR | 0.9296 | 0.9384 | 0.3756 | 0.1389 |
| | SSIM | 0.7179 | 0.7397 | 0.7161 | 0.5278 |
| | MS-SIM | 0.7255 | 0.7208 | 0.7052 | 0.5556 |
| | FSIM | 0.5987 | 0.2492 | 0.8197 | 0.7222 |
| | VIFp | **0.9347** | **0.9304** | 0.3641 | 0.1944 |
| CGH | PSNR | 0.6579 | 0.5654 | 0.9340 | 0.5278 |
| | SSIM | 0.8159 | 0.7302 | 0.7174 | 0.3333 |
| | MS-SIM | 0.8669 | 0.7788 | 0.6196 | 0.2500 |
| | FSIM | 0.7856 | 0.7597 | 0.7641 | 0.4167 |
| | VIFp | **0.9772** | **0.9305** | 0.2643 | 0.0556 |
| All | PSNR | 0.6696 | 0.6551 | 0.8532 | 0.4306 |
| | SSIM | 0.5390 | 0.3056 | 0.9640 | 0.7500 |
| | MS-SIM | 0.5722 | 0.3090 | 0.9374 | 0.7500 |
| | FSIM | 0.5040 | 0.3781 | 0.9839 | 0.8056 |
| | VIFp | **0.7098** | **0.7358** | 0.7988 | 0.5972 |

TABLE V
Statistical analysis of objective metrics (test 2).

| | | PLCC | SRCC | RMSE | OR |
|---|---|---|---|---|---|
| OGH | PSNR | 0.7500 | 0.7235 | 0.3571 | 0.6389 |
| | SSIM | 0.4687 | 0.3571 | 0.4786 | 0.8333 |
| | MS-SIM | 0.6636 | 0.4664 | 0.4046 | 0.7500 |
| | FSIM | 0.1489 | 0.1685 | 0.5368 | 0.8889 |
| | VIFp | **0.8268** | **0.8323** | 0.3071 | 0.5278 |
| CGH | PSNR | 0.7182 | 0.6908 | 0.7550 | 0.6667 |
| | SSIM | 0.6420 | 0.5743 | 0.8342 | 0.6667 |
| | MS-SIM | 0.6689 | 0.6709 | 0.8082 | 0.6667 |
| | FSIM | 0.8788 | 0.8826 | 0.5186 | 0.4167 |
| | VIFp | **0.9785** | **0.9488** | 0.2238 | 0.0278 |
| All | PSNR | 0.7232 | 0.6656 | 0.6175 | 0.5694 |
| | SSIM | 0.2189 | 0.2237 | 0.8760 | 0.9722 |
| | MS-SIM | 0.5912 | 0.5012 | 0.7300 | 0.6528 |
| | FSIM | 0.2603 | 0.0387 | 0.8682 | 0.9167 |
| | VIFp | **0.7947** | **0.7475** | 0.5443 | 0.5694 |

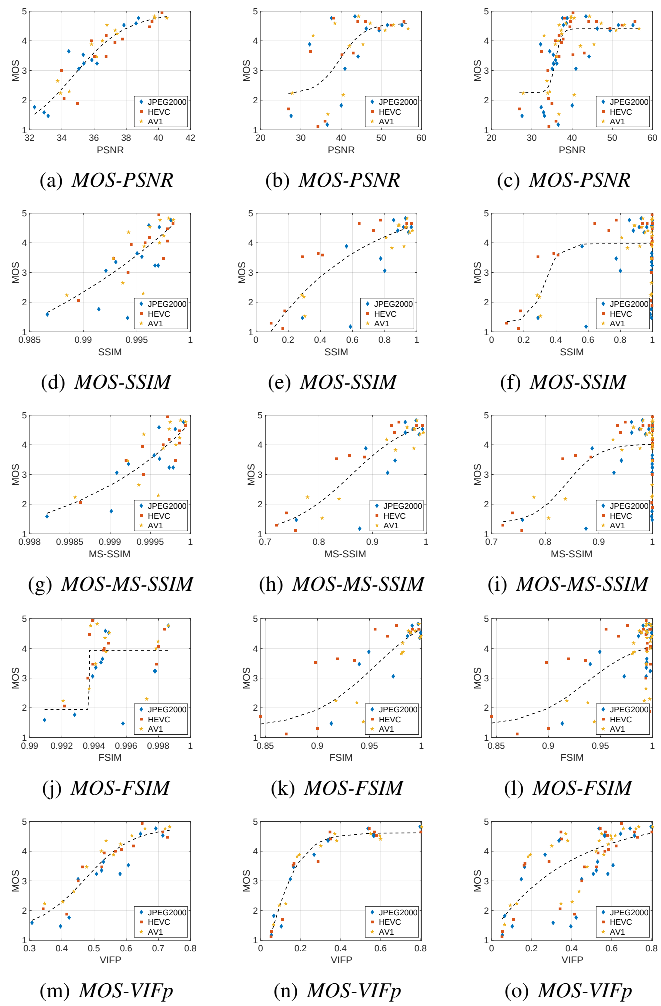



Fig. 10. Subjective scores vs objective metric (test 1); (left) OGH, (middle) CGH, (right) all.

## References


[1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, 1948.

[2] Ulf Schnars and Werner PO Jüptner, "Digital recording and numerical reconstruction of holograms," *Measurement science and technology*, vol. 13, no. 9, pp. R85, 2002.

[3] Tomoyoshi Shimobaba, Takashi Kakue, and Tomoyoshi Ito, "Review of fast algorithms and hardware implementations on computer holography," *IEEE Trans. on Ind. Informatics*, vol. 12, no. 4, pp. 1611–1622, 2015.

[4] Hans J Coufal, Demetri Psaltis, Glenn T Sincerbox, et al., *Holographic data storage*, vol. 8, Springer, 2000.

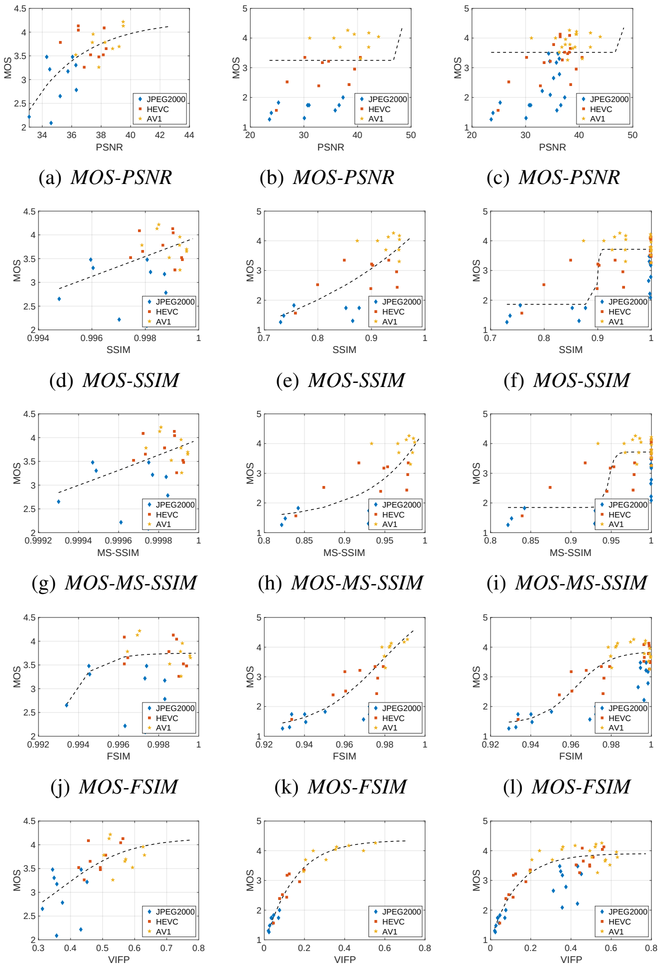
[5] Min-Tzung Shiu, Yang-Kun Chew, Huang-Tian Chan, Xin-Yu Wong, and Chi-Ching Chang, "Three-dimensional information encryption and anticounterfeiting using digital holography," *Appl. Opt.*, vol. 54, no. 1, pp. A84–A88, Jan 2015.

[6] Elchanan Bruckheimer, Carmel Rotschild, Tamir Dagan, Gabriel Amir, Aviad Kaufman, Shaul Gelman, and Einat Birk, "Computer-generated real-time digital holography: first time use in clinical medical imaging," *Europ. Heart Journal - Card. Imag.*, vol. 17, no. 8, pp. 845–849, 06 2016.

[7] Myung K Kim, "Principles and techniques of digital holographic microscopy," *SPIE reviews*, vol. 1, no. 1, pp. 018005, 2010.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2021.3096059, IEEE Transactions on Multimedia






(a) *MOS-PSNR* (b) *MOS-PSNR* (c) *MOS-PSNR*

(d) *MOS-SSIM* (e) *MOS-SSIM* (f) *MOS-SSIM*

(g) *MOS-MS-SSIM* (h) *MOS-MS-SSIM* (i) *MOS-MS-SSIM*

(j) *MOS-FSIM* (k) *MOS-FSIM* (l) *MOS-FSIM*

(m) *MOS-VIFp* (n) *MOS-VIFp* (o) *MOS-VIFp*

Fig. 11. Subjective scores vs objective metric (test 2) (left) OGH, (liddle) CGH, (right) all.


[8] B Gombkötő, J Kornis, and Z Füzessy, "Difference displacement measurement using digital holography," *Optics communications*, vol. 214, no. 1-6, pp. 115–121, 2002.

[9] L. Onural, F. Yara, and H. Kang, "Digital holographic three-dimensional video displays," *Proceedings of the IEEE*, , no. 99, pp. 1–14, 2011.

[10] David Blinder, Ayyoub Ahar, Stijn Bettens, Tobias Birnbaum, Athanasia Symeonidou, Heidi Ottevaere, Colas Schretter, and Peter Schelkens, "Signal processing challenges for digital holographic video display systems," *Signal Proc.: Image Com.*, vol. 70, pp. 114 – 130, 2019.

[11] Peter Schelkens, Zahir Y. Alpaslan, Touradj Ebrahimi, Kwan-Jung Oh, Fernando Pereira, António M. G. Pinheiro, Ioan Tabus, and Zhibo Chen, "Jpeg pleno: a standard framework for representing and signaling plenoptic modalities," in *Optical Engineering + Applications*, 2018.

[12] J. P. Peixeiro, C. Brites, J. Ascenso, and F. Pereira, "Holographic data coding: Benchmarking and extending hevc with adapted transforms," *IEEE Trans. on Multimedia*, vol. 20, no. 2, pp. 282–297, Feb 2018.

[13] A. El Rhammad, P. Gioia, A. Gilles, M. Cagnazzo, and B. Pesquet-Popescu, "Color digital hologram compression based on matching pursuit," *Appl. Opt.*, vol. 57, no. 17, pp. 4930–4942, Jun 2018.

[14] Kartik Viswanathan, Patrick Gioia, and Luce Morin, "Wavelet compression of digital holograms: Towards a view-dependent framework," in *Applications of Digital Image Processing XXXVI*, Andrew G. Tescher, Ed. International Society for Optics and Photonics, 2013, vol. 8856, pp. 539 – 548, SPIE.

[15] Marco Bernardo, Antonio Pinheiro, and Manuela Pereira, "Benchmarking coding standards for digital holography represented on the object plane," in *Proc. SPIE*. April 2018, pp. 6–8, Spie.

[16] I. Viola, M. Rerabek, and T. Ebrahimi, "Comparison and evaluation of light field image coding approaches," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1092–1106, Oct 2017.

[17] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *19th Int. Workshop on Multimedia Sign. Proc. (MMSP)*, Oct 2017, pp. 1–6.

[18] E. Darakis, M. Kowiel, R. Nasanen, and J. Naughton, "Visually lossless compression of digital hologram sequences," *System*, vol. 7529, pp. 752912–752912–8, 2010.

[19] Ayyoub Ahar, David Blinder, Tim Bruylants, Colas Schretter, Adrian Munteanu, and Peter Schelkens, "Subjective quality assessment of numerically reconstructed compressed holograms," in *Appl. of Digital Image Proc. XXXVIII*. Int. Soc. for Opt. and Phot., 2015, vol. 9599.

[20] Ayyoub Ahar, Maksymilian Chlipala, Tobias Birnbaum, Weronika Zaperty, Athanasia Symeonidou, Tomasz Kozacki, Malgorzata Kujawinska, and Peter Schelkens, "Suitability Analysis of Holographic vs Light Field and 2D Displays for Subjective Quality Assessment of Fourier Holograms," *arXiv e-prints*, p. arXiv:1909.12594, Sep 2019.

[21] Peter Schelkens, Antonin Gilles, Saeed Mahmoudpour, Kwan-Jung Oh, Cristian Perra, and Antonio Pinheiro, "Standardization of holographic compression: Jpeg pleno," in *Imaging and Applied Optics Congress*. 2020, p. HF1D.1, Optical Society of America.

[22] Raees Muhamad, Tobias Birnbaum, Antonin Gilles, Saeed Mahmoudpour, Kwan-Jung Oh, Manuela Pereira, Cristian Perra, Antonio Pinheiro, and Peter Schelkens, "Jpeg pleno holography: Scope and technology validation procedures," *Applied Optics*, 12 2020.

[23] Elsa Fonseca, Paulo T Fiadeiro, Marco V Bernardo, António Pinheiro, and Manuela Pereira, "Assessment of speckle denoising filters for digital holography using subjective and objective evaluation models," *Applied Optics*, vol. 58, no. 34, pp. G282–G292, 2019.

[24] R. Corda and C. Perra, "Hologram domain data compression: Performance of standard codecs and image quality assessment at different distances and perspectives," *IEEE Trans. on Broadcast.*, pp. 1–18, 2019.

[25] L. A. da Silva Cruz, E. Dumi, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh Int. Conf. on Quality of Multimedia Experience*, June 2019, pp. 1–6.

[26] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, and Peter Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[27] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep., International Telecommunication Union, January 2012.

[28] Marco Bernardo, Pedro Fernandes, Angelo Arrifano, Marc Antonini, Elsa Fonseca, Paulo Fiadeiro, António Pinheiro, and Manuela Pereira, "Holographic representation: Hologram plane vs. object plane," *Signal Processing: Image Communication*, vol. 68, pp. 193–206, 2018.

[29] D. Blinder, A. Ahar, At. Symeonidou, Y. Xing, T. Bruylants, C. Schretites, B. Pesquet-Popescu, F. Dufaux, A. Munteanu, and P. Schelkens, "Open access database for experimental validations of holographic compression engines," in *Seventh International Workshop on Quality of Multimedia Experience*, May 2015, pp. 1–6.

[30] Athanasia Symeonidou, David Blinder, Ayyoub Ahar, Colas Schretter, Adrian Munteanu, and Peter Schelkens, "Speckle noise reduction for computer generated holograms of objects with diffuse surfaces," in *Optics, Photonics and Digital Technologies for Imaging Applications IV*, Peter Schelkens, Touradj Ebrahimi, Gabriel Cristbal, Frdric Truchetet, and Pasi Saarikko, Eds. International Society for Optics and Photonics, 2016, vol. 9896, pp. 67 – 76, SPIE.

[31] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Proc.*, vol. 13, no. 4, pp. 600–612, April 2004.

[32] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, Nov 2003, vol. 2, pp. 1398–1402 Vol.2.

[33] 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, Aug 2011.

[34] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. on Image Proc.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[35] Philippe Hanhart, Marco Bernardo, Manuela Pereira, António Pinheiro, and Touradj Ebrahimi, "Benchmarking of objective quality metrics for hdr image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 39, Dec 2015.

[36] H. Amirpour, A. M. G. Pinheiro, M. Pereira, and M. Ghanbari, "Reliability of the most common objective metrics for light field quality assessment," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 2402–2406.