A new approach to subjectively assess quality of plenoptic content

Irene Viola, Martin Řeřábek, and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG), EPFL, Lausanne, Switzerland

ABSTRACT

Plenoptic content is becoming increasingly popular thanks to the availability of acquisition and display devices. Thanks to image-based rendering techniques, a plenoptic content can be rendered in real time in an interactive manner allowing virtual navigation through the captured scenes. This way of content consumption enables new experiences, and therefore introduces several challenges in terms of plenoptic data processing, transmission and consequently visual quality evaluation. In this paper, we propose a new methodology to subjectively assess the visual quality of plenoptic content. We also introduce a prototype software to perform subjective quality assessment according to the proposed methodology. The proposed methodology is further applied to assess the visual quality of a light field compression algorithm. Results show that this methodology can be successfully used to assess the visual quality of plenoptic content.

Keywords: plenoptic function, light field compression, subjective evaluation, image coding.

1. INTRODUCTION

A plenoptic function represents the intensity of light rays passing through every possible point in space (V_x, V_y, V_z) at every possible angle (θ, ϕ), wavelength λ, and time t. This function is commonly used to describe Light Fields (LF). The plenoptic function can be simplified considering only information taken in a region free of occlusions at a single time instance.[1] In this case, the plenoptic function can be parameterized as an intersection of two planes (u, v) and (x, y). This parametrization allows to consider the LF as a collection of perspective images (viewpoints) of the xy spatial plane, each perceived from a different observer position in a uv viewing plane. This simplification is commonly referred to as a 4D LF. More specifically, in relation to image-based rendering, 4D LF represents the parameterized plenoptic information as a stack of 2D images taken from different viewpoints for any given point in time. A sequence of those images in time forms a LF video content. Each LF video content can be seen as a function of six dimensions:

$$\mathcal{L} = \mathcal{L}(u, v, x, y, c, t), \tag{1}$$

in which x and y represent one point in the color channel c of the perspective image taken at viewpoint (u, v), at the time t. In the following discussion, we will only consider content at a fixed time t_0, so that the LF content is a function of five dimensions:

$$\mathcal{L} = \mathcal{L}(u, v, x, y, c), \tag{2}$$

and we refer to it as a collection or stack of 2D images. Extension to video content by explicitly taking into account time dimension is straightforward.

From the parametrization of 4D LF stems the main idea behind plenoptic acquisition, which is to acquire several images from different viewpoints to obtain a 4D LF. The same idea is used to acquire holographic stereograms.[2] In general, plenoptic content can be acquired through different acquisition technologies, such as multicamera arrays and single lens cameras, point-clouds, holograms, etc. Regardless of the acquisition technology, it is possible to represent the acquired content with a proper parametrization as a 4D LF. For this reason, in the following discussion we will focus on 4D LF as we defined in Equation 2. We will refer to the acquired content that has been parametrized according to Equation 2 as LF content.

Further author information: (Send correspondence to authors) E-mail: {firstname.lastname}@epfl.ch

In general, the acquisition of plenoptic content creates more data, when compared to traditional photography, triggering the need for efficient compression schemes. In 2014, the JPEG standardization committee launched a new activity called JPEG PLENO, which aims at creating a standard framework for efficient storage and delivery of LF, point-cloud and holographic content. In particular, the goal of JPEG PLENO is to find an as small as possible number of representation models for these types of content that when necessary also can offer inter-operability with existing solutions, such as legacy JPEG and JPEG 2000 formats.

Additionally, it is essential to evaluate the visual quality of plenoptic content within different rendering and representation solutions. Assessing the quality of plenoptic contents poses interesting challenges, not only in terms of objective measures but especially in terms of subjective evaluations. Plenoptic contents offer a wide range of possibilities in rendering. For example, it is possible to change the focal plane or the point of view in an interactive way. These peculiarities have to be addressed in the methodology used for subjective quality evaluation. In this paper, we propose a new methodology for visual quality assessment of a plenoptic content, which allows for interaction with the content and assessment of user experience. The new methodology is subsequently validated within an experiment assessing the visual quality of compressed LF content.

The remainder of the paper is organized as follows. Section 2 presents relevant work in LF content acquisition, compression and quality evaluation. Section 3 presents the new methodology in details, whereas Section 4 exposes the validation experiments. Results of the experiments are discussed in Section 5. Section 6 concludes the paper.

2. RELATED WORK

In this section, the state of the art in LF acquisition, compression, and evaluation is reviewed. First, the related work in LF acquisition and LF rendering is briefly discussed as well as the current approaches to LF content compression. Additionally, the objective and subjective quality evaluation methods to assess the LF content are presented.

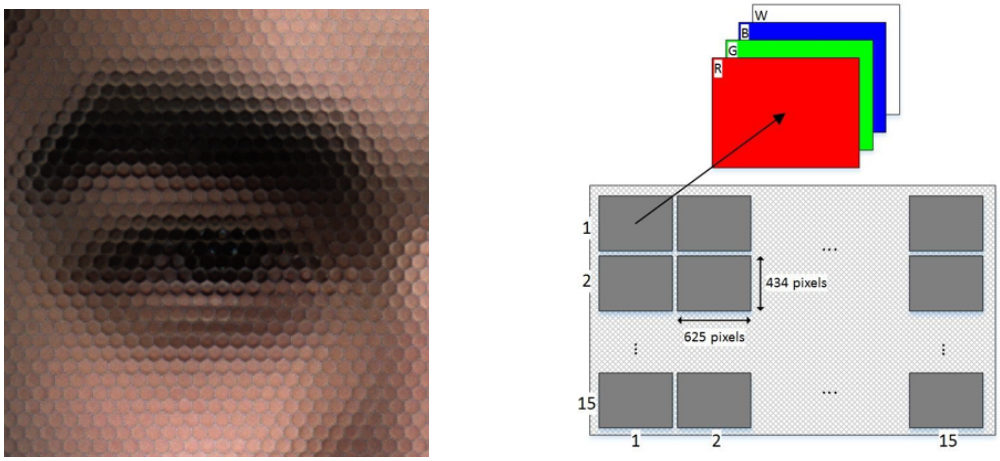
2.1 Light field acquisition

Several methods to acquire the LF image and video content have been proposed in the literature. For instance, the Stanford motorized gantry, which uses a single moving digital camera allowing to record a scene from different viewpoints with four degrees of freedom.[3] However, this solution works only for the static objects, where the LF information doesn't change during the capturing process. To capture a dynamic scene, a multi-camera arrays or special LF cameras need to be used. Multi-camera array assembled in a regular[4] or an irregular grid can be used to capture the LF image and video content, especially in applications where a good scene reconstruction is required.

On the other hand, a specific single LF camera solution to acquire LF content relies on the additional optical element placed in front of the camera sensor. Nowadays, most LF cameras use a microlens array solution. The main disadvantage of single sensor LF cameras is in their limitation in terms of captured camera baseline. In another words, the amount of viewpoints and the captured field of view is limited by the size of the optics and sensors. Referring to the 4D LF defined in Equation 2, the xy resolution depends on the number of microlenses, while the uv resolution depends on the number of pixels behind each microlens and other constraints in the camera optics. Hand-held cameras implementing this model[5] are already widely available[*,†]. The raw image obtained with this type of cameras closely resembles the array of lenses that has been used for the acquisition (see Figure 1(a)), and in this paper it will be further referred to as lenslet image. From lenslet images it is possible to obtain a stack of images, organized in a LF data structure (Figure 1(b)). Each image in the LF data structure represents one viewpoint formed by taking the corresponding samples (pixels) from each microlens element. Individual images in the LF data structure have an infinite depth of field, which means that every image plane in each image is in focus. The natural consequence of the transformation of lenslet image into LF data structure is that it can be rendered on conventional 2D displays.

[*]https://www.raytrix.de/

[†]https://www.lytro.com/



(a) Crop of an uncompressed lenslet image.

(b) LF data structure.

Figure 1: Raw data obtained from lenslet camera and corresponding data structure.

2.2 Light field compression

Acquisition of LF content creates a tremendous amount of data. Therefore, in order to transmit and store such content, efficient compression algorithms are required. Several approaches to compress LF content have been proposed, showing various degrees of performance.

Levoy and Hanrahan choose a two-stage pipeline, comprised of vector quantization and Lempel-Ziv coding.[1] Magnor et al. propose two approaches to LF compression.[6] The first approach adapts video compression techniques to LF data. A number of LF images are chosen to be compressed as intra frames, which will then serve as reference for the remaining images that will be predicted. The second approach relies on disparity compensation. A disparity map indicating the shifts occurring between one LF image and the other is computed for a certain number of LF images using summed squared error criterion. Then, the disparity map of the remaining images is estimated by compensating the disparity maps of already coded images. Girod et al. proposed a compression scheme that incorporates disparity compensation into 4D wavelet coding using disparity-compensated lifting.[7] Disparity compensation is proven more efficient with synthetic images, due to the fact that disparity compensation alone is in most of the cases sufficient to estimate the images correctly, without any error vector.[6] Zhu et al. focused on distributed compression for large camera arrays, and proposed a solution based on Wyner-Ziv coding which relies on geometry estimation.[8] Each camera is considered independently from the others, so each image can be accessed separately. Jagmohan et al. use disparity compensation vectors and disparity error vectors generated from the compensation, that are then transformed, quantized and entropy coded.[9]

Recently, several compression schemes have been proposed to directly compress the lenslet images. In general, two main approaches are proposed. The first approach compresses the lenslet image via intra prediction by exploiting existent redundancies. The second approach creates the LF data structure prior to coding, and then rearranges the sub-aperture images in a pseudo-temporal sequence to be coded with video coders. Li et al. incorporate a full inter-prediction scheme in HEVC Intra prediction. This scheme allows to exploit the redundancy in lenslet images, which is not commonly exploited by intra coders.[10] Vieira et al. compare five different HEVC-compatible coding of lenslet images with different data formats, including tiling of sub-aperture views and lenslet.[11] Chen et al. use a LF dictionary based on perspective shifting to code the sub-aperture images obtained from the lenslet.[12] Dai et al. extract sub-aperture images from the lenslet. Then they rearrange them and code them using JPEG and AVC.[13]

2.3 Light field quality evaluation

Evaluation of visual quality and user experience is of paramount importance in designing new compression solutions and new rendering techniques. Most solutions listed above provide a preliminary performance evaluation. However, only few publications are devoted to discussion about objective metrics and subjective methodologies for LF content quality assessment.

Ramanathan et al. create a framework for analysing the impact of various parameters on the rate-distorsion curve of LF coding based on disparity compensation.[14] The parameters include correlation within an image and between images, geometry accuracy and prediction dependency structure. The paper focuses on compression schemes based on disparity compensation performed with an explicit geometric model. As such, it cannot be used to evaluate compression schemes using other methods. Shidanshidi et al. also introduce a new objective metric to perform an evaluation of LF rendering.[15,16] They present a geometric measurement, called effective sampling density, and they perform a comparison on existing techniques for interpolation of new views from LFs. However, they do not perform an assessment on compression solutions, but only on the algorithms used to render new views from existing ones. Fu et al. analyse the effects of LF photography in image quality.[17] They first develop a simulation approach to test visual resolution and other image quality evaluation metrics for LF photography. Then they compare the results with conventional cameras to discuss improvements and shortcomings of LF photography. Although, it performs visual quality analysis, the paper does not focus on evaluating compression solutions, but only on assessing the visual quality of images acquired with LF techniques when compared to images acquired in a traditional way.

Vieira et al. compare five different HEVC-compatible coding of lenslet images with different data formats.[11] The objective quality evaluation is performed using PSNR as a full reference metric. However, authors don't address subjective evaluation of the content. Finally, a grand challenge was recently organized at ICME 2016 under a collaboration between Qualinet and JPEG standardization committee to collect new compression solutions for LF images, and to evaluate them using both objective and subjective quality assessment methodologies.[18] The grand challenge was focused on compression schemes for raw LF images acquired with a lenslet based Lytro Illum camera. The subjective assessment was carried out on five rendered views from each content, which were evaluated separately by test subjects. The conducted evaluations, however, did not reflect the feature rich LF content. Since the assessment was conducted on predefined rendered views separately, it did not address the issue of evaluating global quality of experience offered by a compressed LF content.

3. PROPOSED EVALUATION METHODOLOGY

In many applications where LF content is used, LF rendering as any image-based rendering can offer the possibility either to visualize acquired image data directly or to produce new visual effects (e.g. change of perspective, refocus, change in depth of field, relighting, zoom, etc.) of the captured scene prior to display. Assessing the visual quality of LF content in such applications poses several challenges. For instance, it is not clear which and even how many variations of rendered views should be selected in evaluations and how to present them to subjects in an optimal way.

One approach to assess quality of LF content is to separately evaluate the quality of a limited number of rendered views, carefully selected. A single quality score for the entire LF content can be obtained by averaging the scores obtained from each rendered view. However, averaging the quality of the rendered views does not always give a good indication of the quality of the LF content as a whole, nor adequately reflects all potential in new experiences offered with such content.

Unfortunately, most current methodologies for evaluation of multimedia content are designed considering a passive relationship between the content and observers subjected to rendered content on conventional devices. Beside limitations of such approaches which assign quality scores for single rendered views as opposed to a global quality score for the LF as a whole, they require a large number of stimuli to be tested, thus increasing the length of the test and causing fatigue on test subjects. The decision about which of the rendered views to display, is another challenge that should be addressed when considering current evaluation methodologies.

In subjective evaluations of LF content, it is more efficient to design methodologies that enable global assessment of quality of experience in a flexible and interactive way.

To avoid interaction when it is difficult to implement, a compromise approach could be to proceed with an automatic presentation of rendered views in form of an animation that shows different rendered views of the LF content under test. This, in fact, reduces the number of stimuli to be tested, and allows evaluation of the LF content as a whole. However, there are a number of parameters to determine, such as the optimal rate of change in the automatic animation and the order of rendered views.

The methodology we propose, introduces interaction with LF content allowing to reflect its rendering possibilities and to explore in a better way the user experience while consuming the LF content. The interaction is enabled through an interface featuring a real time response. In particular, depending on modality in which the content is represented, it is possible to change the perspective in a narrow or wide mode, to modify the depth of field, to refocus on different objects in the scene and even to freely navigate within the scene and around objects, among many other possible effects such as relighting, zoom, change of dynamic range, and combinations of these. In general, it is possible to fully explore the various ways in which LF content can be rendered.

The LF test content can be rendered in a temporal or a side-by-side fashion, depending on the selected presentation method and scale. The LF content is rendered in real time according to the parameters (e.g. perspective, refocus) selected by the participants. For methodologies in which two LF contents are to be presented side-by-side, the two contents are synchronized, so that they are rendered with the same parameters simultaneously. This allows to effectively compare both contents with the same rendering conditions.

Several stimuli-comparison methods can be adopted to obtain scores for the test material, depending on the type of the test conducted:

- **Paired Comparison (PC)**: participants are asked to judge which content has the best overall visual quality ('left' or 'right', depending on the position on the screen). The option "similar" can be included to avoid random preference selections.
- **Double Stimulus Impairment Scale (DSIS)**: participants are asked to rate the level of impairment of the test material with respect to a reference, on a scale from 1 (Very annoying) to 5 (Imperceptible). They are informed of the position of the reference in the screen.
- **Double Stimulus Continuous Quality Scale (DSCQS)**; participants are asked to rate the visual quality of both the test material and the reference, on a scale from 0 (Bad) to 100 (Excellent). They are informed of the presence of a reference, but they are not informed on its position on the screen.

For the purposes of experimental validation of the proposed methodology, we implemented an interface allowing interaction with LF content. The MATLAB Light Field toolbox v0.4[19,20] was used as framework to build a Graphical User Interface (GUI) in MATLAB. The implemented interface allows interaction with LF image content by enabling the change of point of view (perspective) and the choice of different focal points (refocus) from a predefined set. Two stimuli-comparison methods can be chosen, PC or DSIS. Figure 2 shows an example screen for DSIS.

The interface takes as input two LF image contents, organized in 5-D arrays of dimension $U \times V \times H \times W \times C$, in which $U \times V$ is the resolution of the uv plane (i.e, the number of different viewpoints, which is the number of images composing the stack), $H \times W$ is the resolution of the xy plane (i.e, the resolution of each rendered view) and C is the number of color channels (1 for greyscale, 3 for RGB, 4 if a weighting component is added). Additionally, the interface requires a set of images rendered from the LF content at different focal points, which we will refer to as refocused views. Test and reference material must have the same resolution and the same number of rendered views. The central viewpoint image from the 5-D array taken as input is displayed as default for both reference and test material. By click-and-drag inside the rendered images, the user can change the viewpoint image, which is rendered in real time. A slider on the bottom of each rendered image allows access to the refocused views. Labels on the bottom of the slider indicate if the content will be refocused on the foreground or on the background. A button placed right below the slider allows to render the central viewpoint image from the 5-D array, which has every plane in focus. The two contents are rendered simultaneously and they are perfectly synchronized, so the displayed views are rendered with the same parameters. A panel on the bottom

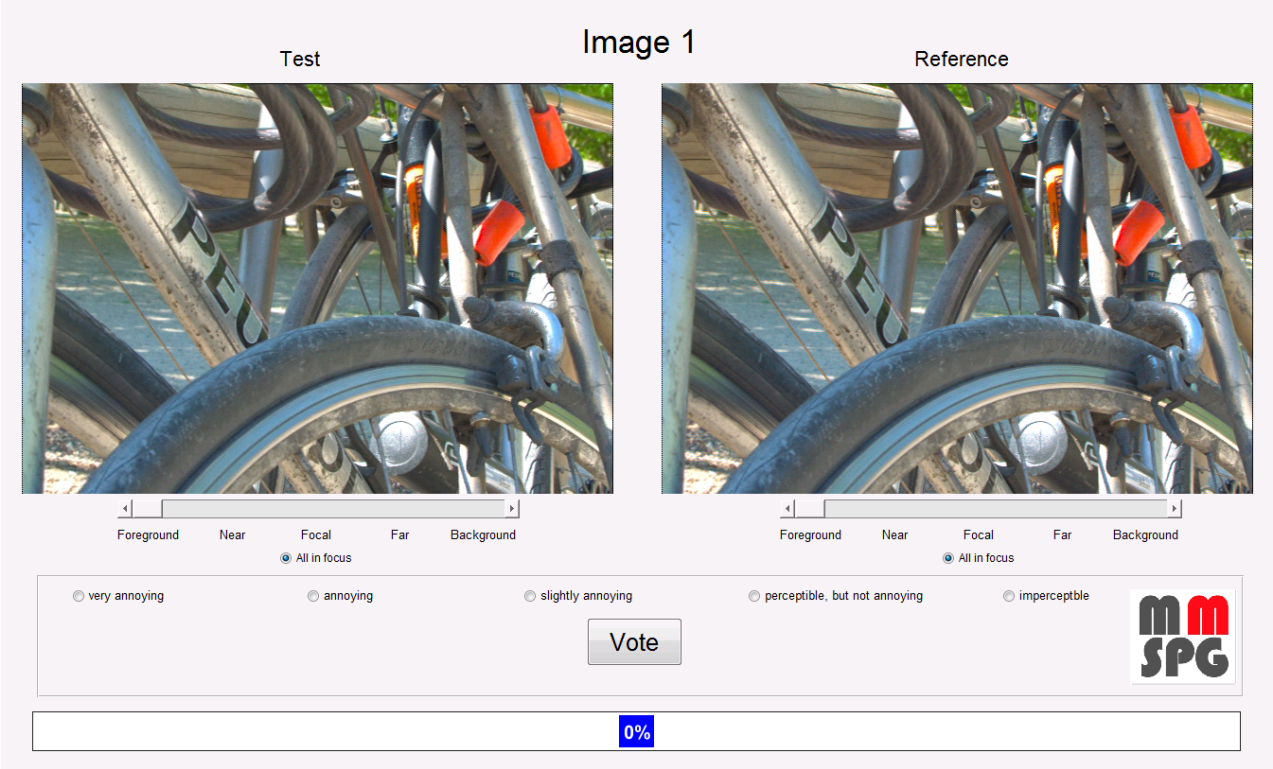


Figure 2: Example of evaluation interface screen for DSIS methodology.

of the screen shows the possible scores for the test material. After one option is selected, the user can vote and move to the next test material. A progress bar displays the percentage of test material successfully scored.

4. VALIDATING EXPERIMENT

This section describes the validating experiment in details. More specifically, the creation of the test material and its description is presented as well as the description of the testing environment. Then, a delineation of the test methodology and test plan is provided together with statistical analysis of collected subjective scores.

4.1 Dataset preparation and description

For the experiments, four LF contents, acquired by a Lytro Illum camera, were chosen from a publicly available LF image dataset.[21] More specifically, *Bikes*, *Stone_Pillars_Outside*, *Fountain_&_Vincent_2* and *Friends_1* contents were selected for our experiments.

The 10bit raw lenslet image was preprocessed according to the end-to-end chain depicted in Figure 3. The raw data was demosaiced, devignetted and clipped to 8 bits (Point B). Then the color space was converted from RGB to YCbCr and downsampled from 444 to 420 (Point A). At this point, the data was compressed and decompressed using two different codecs (Point A'). The color space was upsampled from 420 to 444 and converted from YCbCr to RGB (Point B'). The uncompressed reference was obtained by omitting the steps from Point B to Point B'.

HEVC Intra profile was chosen to compress the raw lenslet images. In order to compress the lenslet image, the reference software was used [‡]. The full command line used can be found in Table 1. The Quantization

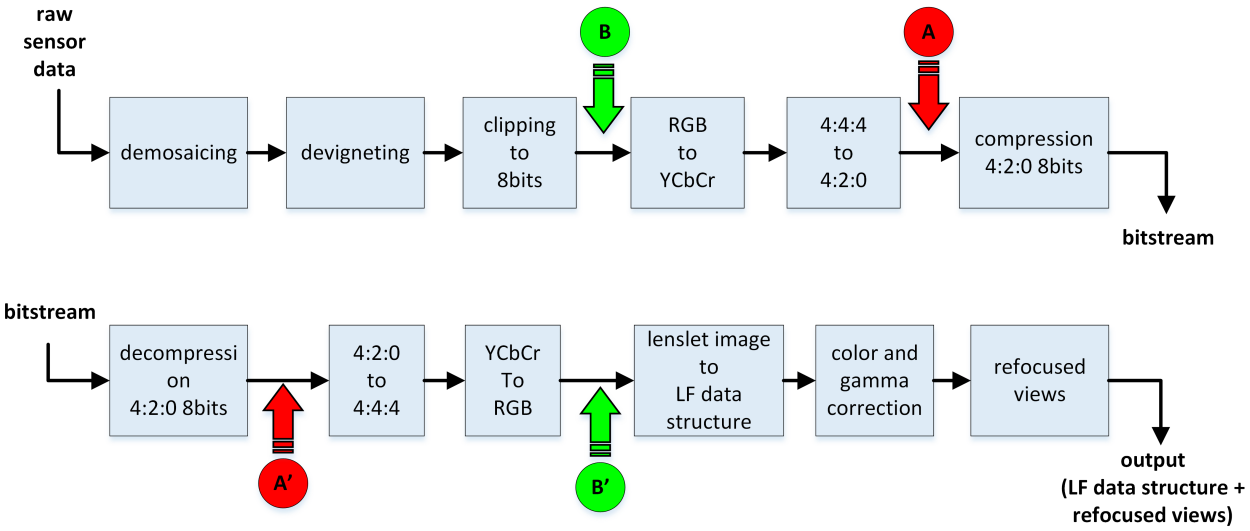


Figure 3: End-to-end chain used for creating the test material.

Table 1: Selected settings for HEVC Intra coder.

--input-res 7728x5368 --fps 1
--tune psnr --profile main10
--level-idc 85 --output-depth 8
--crf < *QP* > < *Input* > -o < *Output* >

Table 2: QP chosen to encode all contents with HEVC Intra.

Content	**R1**	**R2**	**R3**	**R4**
Bikes	27	33	38	45
Stone_Pillars_Outside	26	31	37	43
Fountain_&_Vincent_2	30	36	41	47
Friends_1	24	29	33	39

Parameters (QP) were chosen to match the corresponding compression ratios listed above. Table 2 summarizes the values of different QP used in the test.

The lenslet image obtained at Point B' cannot be rendered as it is. An additional step is needed, which creates the LF data structure from the lenslet image. The lenslet images are transformed into a LF data structure of viewpoint images using the MATLAB implementation of the Light Field toolbox v0.4.[19,20]The number of viewpoint images corresponds to the number of pixels behind each microlens, in this case 15×15. The size of each viewpoint image corresponds to the number of microlens in the camera, in this case 625×434. The viewpoint images are stacked in a LF data structure of size $15 \times 15 \times 625 \times 434 \times 4$, in which the last dimension corresponds to the RGB channels plus a weighting component. Color and gamma correction is applied on the LF data structure prior to rendering. The color correction parameters are obtained from the lenslet in the decoding process, while $\gamma = 1/2.2$. Figure 4 shows the central viewpoint image from the corresponding LF data structure for each content.

The test LF content was displayed together with the uncompressed reference in a side-by-side fashion, using the implemented interface. For each stimulus, the central viewpoint image from the LF data structure was displayed. By clicking inside the displayed image and dragging the mouse, the other viewpoints from the

[‡]https://www.videolan.org/developers/x265.html

(a) Bikes (b) Stone_Pillars_Outside (c) Fountain_&_Vincent_2 (d) Friends_1

Figure 4: Central viewpoint image from each content used in the test.

Table 3: Values of refocusing slope for each content.

	Slopes										
Content	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Bikes	-10	-8	-6	-4	-2	0	2	4	6	8	10
Stone_Pillars_Outside	-10	-8	-6	-4	-2	0	2	4	6	8	10
Fountain_&_Vincent_2	-10	-8	-6	-4	-2	0	2	4	6	8	10
Friends_1	-5	-4	-3	-2	-1	0	1	2	3	4	5

data structure were accessed and displayed. Each image was displayed in its native resolution of 625×434 pixels. Eleven refocused images were created for each content, using a modified version of the toolbox function *LFFiltShiftSum*. The function shifts all the images in the stack according to a parameter, called slope, and performs a sum of the shifted images to obtain a single image that is refocused on a specific plane, which depends on the value of the slope. The number of images to be shifted and consequently summed defines the depth of field. Summing all 15×15 images creates the smallest depth of field, in which only one specific plane in the image is in focus. On the other hand, taking just the central image, which is equivalent to summing just 1×1 images, brings all the objects in focus (largest depth of field). For the test, it was chosen to sum images from index 3 to index 13 (11×11 images) to have a larger depth of field that still shows the effects of refocusing. The values of the slopes are summarized in Table 3. The refocused images were accessible through a slider shown at the bottom of each stimulus. The slopes were selected so as to assure gradual transition between refocusing on the foreground and on the background.

The compression algorithm was evaluated on four bitrates, namely $R1 = 1$ bpp $(10:1)$, $R2 = 0.5$ bpp $20:1)$, $R3 = 0.25$ bpp $(40:1)$, $R4 = 0.1$ bpp $(100:1)$. The compression ratios are computed as ratios between the size of the uncompressed raw images in 10bit precision ($5368 \times 7728 \times 10$ bits $= 414839040$ bits $= 10$ bpp) and the size of the compressed bitstream.

4.2 Testing environment

To avoid the involuntary influence of external factors and to ensure the reproducibility of results, the laboratory for subjective video quality assessment was set up according to ITU recommendation BT.500-13.[22] Professional Eizo ColorEdge CG301W 30-inch monitors with native resolution of 2560×1600 pixels were used for the test. The monitors were calibrated using an i1Display Pro color calibration device according to the following profile: sRGB Gamut, D65 white point, 120cd/m^2 brightness, and minimum black level of 0.2 cd/m^2. The room was equipped with a controlled lighting system that consisted of neon lamps with 6500 K color temperature, while the color of all the background walls and curtains present in the test area was mid grey. The illumination level measured on the screens was 15 lux. The distance of the subjects from the monitor was approximately equal to 7 times the height of the displayed content, conforming to requirements in ITU Recommendation ITU-R BT.2022.[23] A picture of the MMSPG test laboratory serving as an evaluation environment is shown in Figure 5.

Figure 5: Testing environment.

4.3 Test methodology and planning

The selected methodology was based on DSIS.[22] The participants were asked to interact with the LF contents and to rate the level of impairment of the test LF content with respect to the reference, on a scale from 1 (Very annoying) to 5 (Imperceptible). Each LF content was presented together with the uncompressed reference in a side-by-side fashion, in its native resolution of 625×434 pixels. The position of the reference was fixed for each experiment, and the participants were made aware of its location on the screen (either left or right).

Before the experiments, a training session was organized to allow participants to get familiar with artifacts and distorsions in the test images. Five training samples were manually selected by expert viewers. In order not to influence the results, the training samples were created by compressing other contents on various bitrates. The content used for the training was chosen from the same LF database used for the test images.[21] The training samples were presented along with the uncompressed reference, exactly as they were shown in the test.

The test samples were randomly distributed for each participant. The same content was never shown consecutively. Before the test, one dummy sample was inserted to ease the participants into the task. The resulting scores from dummy stimuli were not included in the results.

A total of 24 subjects (19 males and 5 females) participated in the experiment, for a total of 24 scores per stimulus. Subjects were between 18 and 35 years old, with an average of 24.79 and a median of 25 years of age. All subjects were screened for correct visual acuity with Snellen charts, and color vision using Ishihara charts.

4.4 Data processing

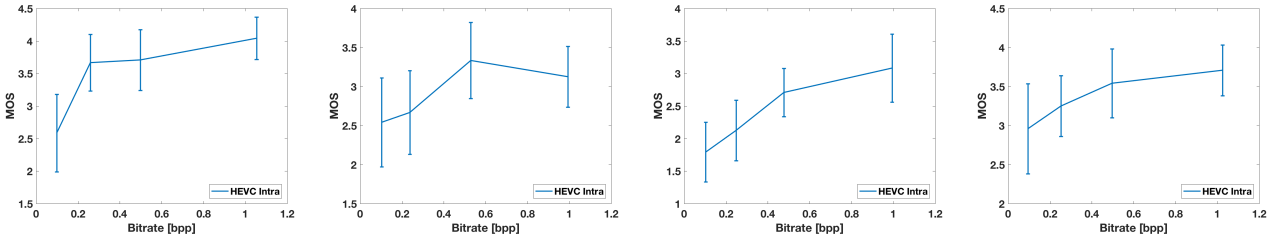
Outlier detection was performed according to the guidelines defined in ITU recommendation BT.500-13.[22] One outlier was detected and the relative scores were discarded, thus leading to 23 scores per stimulus. The Mean Opinion Score (MOS) was computed for each coding condition j (i.e., each content, codec and compression ratio) as follows:

$$MOS_j = \frac{1}{N} \sum_{i=1}^{N} m_{ij}, \tag{3}$$

where N is the number of participants and m_{ij} is the score for stimulus j by participant i. The corresponding 95% confidence intervals were computed.

5. RESULTS

Figure 6 shows the MOS against bitrate for all the contents under test, with their respective confidence intervals. The higher MOS obtained is between 3 and 4, so the level of impairment in the test content when compared to the reference was rarely considered “Imperceptible”. An expert visual comparison of the test LF contents



(a) Bikes (b) Stone_Pillars_Outside (c) Fountain_&_Vincent_2 (d) Friends_1

Figure 6: MOS vs bitrate for all contents.

(a) Uncompressed reference (b) HEVC Intra

Figure 7: Central image ($u = v = 8$) for content *Bikes* from LF data structure for reference and HEVC Intra for compression ratio R1.

revealed relevant changes in color fidelity when compared to the uncompressed references (Figures 8 and 7). This is mainly due to the lenslet based compression scenario, when a conversion from RGB 444 to YUV 420 is first performed on the lenslet image, thus before applying color and gamma correction. This can explain why MOS scores for subjective evaluation do not reach the "Imperceptible" level.

To further analyze the performance of evaluated coding scheme, especially in relation to how color fidelity affects objective performance, PSNR was selected as a full reference metric.

PSNR values are computed between test material and original reference images. The metric is computed on the 5-D array obtained by transforming the lenslet images in LF data structures (Figure 3), thus only on the viewpoint images. Nevertheless, the computation of PSNR metric was adapted to better suit properties of LF content. The PSNR value is computed on the Y channel as follows:

$$PSNR_Y(k,l) = 10\log_{10}\frac{255^2}{MSE(k,l)}, \tag{4}$$

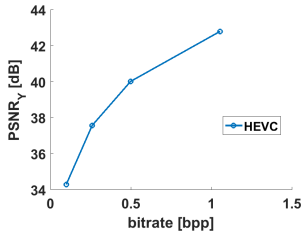
in which k and l are the indexes of the acquired views. The $MSE(k,l)$ for each image is computed as follows:

$$MSE(k,l) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - R(i,j)]^2, \tag{5}$$

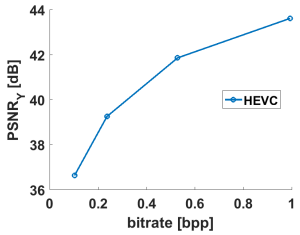
(a) Uncompressed reference

(b) HEVC Intra

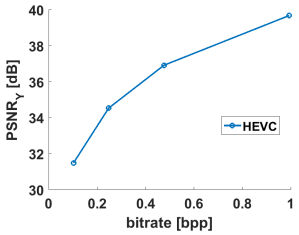
Figure 8: Central image ($u = v = 8$) for content *Friends_1* from LF data structure for reference and HEVC Intra for compression ratio R1.



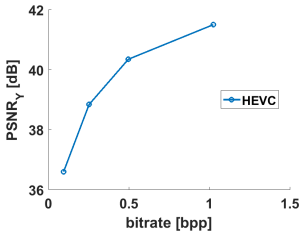
(a) Bikes



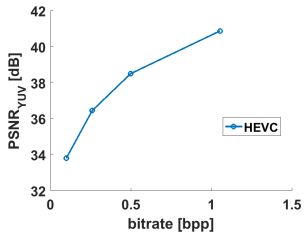
(b) Stone_Pillars_Outside



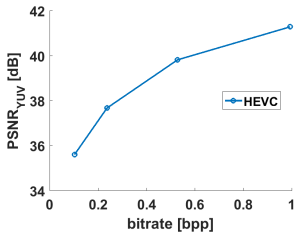
(c) Fountain_&_Vincent_2



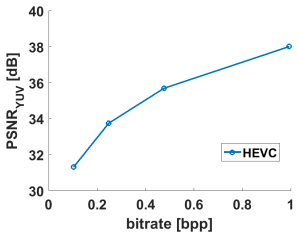
(d) Friends_1



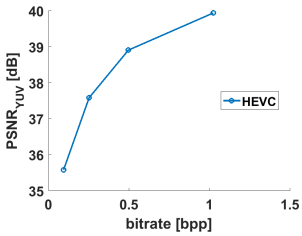
(e) Bikes



(f) Stone_Pillars_Outside



(g) Fountain_&_Vincent_2



(h) Friends_1

Figure 9: PSNR vs bitrate for Y channel ((a) to (d)) and for YUV channels ((e) to (h)).

where m and n are the dimensions of one image (i.e., $n = 625$, $m = 434$). $I(i, j)$ is the Y value for the selected acquired view in the evaluated LF data structure, whereas $R(i, j)$ is the corresponding value in the reference data structure. In the same way, the PSNR for the other two channels U and V is obtained. A weighted average[24] is then computed as follows:

$$PSNR_{YUV}(k,l) = \frac{6PSNR_Y(k,l) + PSNR_U(k,l) + PSNR_V(k,l)}{8} \tag{6}$$

The mean of all viewpoint images is subsequentially computed to have an average value for PSNR for Y channel and for YUV:

$$PSNR_{X_{mean}} = \frac{1}{(K-2)(L-2)} \sum_{k=2}^{K-1} \sum_{l=2}^{L-1} PSNR_X(k,l), \tag{7}$$

in which $K = 15$ and $L = 15$ represent the number acquired views, and $X = Y$ and $X = YUV$ for Y channel and for YUV channels, respectively. Figure 9 show the PSNR values computed on Y and YUV channels for all contents.

Objective results show that the codec achieves transparent quality level for higher bitrates. However, subjective results show that subjects rarely rated the impairment level as "Imperceptible", which can be explained by the loss in color fidelity. The results we obtained show how steps that are necessary to properly render LF content, such as color and gamma correction, heavily affect the visual quality of LF content. New evaluation methodologies addressing the complex nature of LF rendering are thus required to obtain reliable automatic assessment of visual quality of LF content.

6. CONCLUSION

In this paper we proposed a new methodology for subjective quality assessment of plenoptic content. Most state of the art assessment methodologies are designed for a passive evaluation of contents lacking interactions and thus they are not suitable for a complete evaluation of the possibilities that plenoptic contents offer. We presented a new methodology in details and created an implementation in form of a software interface. We also presented and discussed results obtained by using the proposed implementation on lenslet-based light field content. Results showed that the proposed methodology can be successfully used to assess the visual quality of plenoptic content.

More features can be added to the implementation to further enrich the user experience of plenoptic content and to evaluate a wider range of possibilities. Moreover, tracking could be added to assess and subsequently analyze how users interact with the content. This assessment can help to evaluate efficiency of various compression and processing solutions for plenoptic content.

ACKNOWLEDGMENTS

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021_159575) project Light field Image and Video coding and Evaluation (LIVE) and also in the framework of ImmersiaTV under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 688619) funded by Swiss State Secretariat for Education, Research and Innovation SERI. The authors would like to thank Fernando Pereira from IT, Portugal, for useful comments and fruitful discussions.

REFERENCES

[1] Levoy, M. and Hanrahan, P., "Light field rendering," in [*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*], 31–42, ACM (1996).

[2] Halle, M. W., "Holographic stereograms as discrete imaging systems," in [*IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*], 73–84, International Society for Optics and Photonics (1994).

[3] "The (new) stanford light field archive," (7 2016).

[4] Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antunez, E., Barth, A., Adams, A., Horowitz, M., and Levoy, M., "High performance imaging using large camera arrays," in [*ACM Transactions on Graphics (TOG)*], **24**(3), 765–776, ACM (2005).

[5] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P., "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR* **2**(11), 1–11 (2005).

[6] Magnor, M. and Girod, B., "Data compression for light-field rendering," *IEEE Transactions on Circuits and Systems for Video Technology* **10**(3), 338–343 (2000).

[7] Girod, B., Chang, C.-L., Ramanathan, P., and Zhu, X., "Light field compression using disparity-compensated lifting," in [*Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*], **1**, I–373, IEEE (2003).

[8] Zhu, X., Aaron, A., and Girod, B., "Distributed compression for large camera arrays," in [*Statistical Signal Processing, 2003 IEEE Workshop on*], 30–33, IEEE (2003).

[9] Jagmohan, A., Sehgal, A., and Ahuja, N., "Compression of lightfield rendered images using coset codes," in [*Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*], **1**, 830–834, IEEE (2003).

[10] Li, Y., Sjöström, M., Olsson, R., and Jennehag, U., "Efficient intra prediction scheme for light field image compression," in [*2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 539–543, IEEE (2014).

[11] Vieira, A., Duarte, H., Perra, C., Tavora, L., and Assuncao, P., "Data formats for high efficiency coding of lytro-illum light fields," in [*Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on*], 494–497, IEEE (2015).

[12] Chen, J. and Chau, L.-P., "Light field compressed sensing over a disparity-aware dictionary," (2015).

[13] Dai, F., Zhang, J., Ma, Y., and Zhang, Y., "Lenselet image compression scheme based on subaperture images streaming," in [*Image Processing (ICIP), 2015 IEEE International Conference on*], 4733–4737, IEEE (2015).

[14] Ramanathan, P. and Girod, B., "Rate-distortion analysis for light field coding and streaming," *Signal Processing: Image Communication* **21**(6), 462–475 (2006).

[15] Shidanshidi, H., Safaei, F., and Li, W., "Objective evaluation of light field rendering methods using effective sampling density," in [*Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*], 1–6, IEEE (2011).

[16] Shidanshidi, H., Safaei, F., and Li, W., "A quantitative approach for comparison and evaluation of light field rendering techniques," in [*2011 IEEE International Conference on Multimedia and Expo*], 1–4, IEEE (2011).

[17] Fu, Q., Zhou, Z., Yuan, Y., and Xiangli, B., "Image quality evaluation of light field photography," in [*IS&T/SPIE Electronic Imaging*], 78670F–78670F, International Society for Optics and Photonics (2011).

[18] ISO/IEC JTC 1/SC29/WG1 JPEG, "Grand challenge on light field image compression." Doc. M72022, Geneva, Switzerland (June 2016).

[19] Dansereau, D. G., Pizarro, O., and Williams, S. B., "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in [*Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*], IEEE (Jun 2013).

[20] Dansereau, D. G., Pizarro, O., and Williams, S. B., "Linear volumetric focus for light field cameras," *ACM Transactions on Graphics (TOG)* **34** (Feb. 2015).

[21] Rerabek, M. and Ebrahimi, T., "New light field image dataset," in [*8th International Conference on Quality of Multimedia Experience (QoMEX)*], (EPFL-CONF-218363) (2016).

[22] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union (January 2012).

[23] ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays." International Telecommunication Union (August 2012).

[24] Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T. K., and Wiegand, T., "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc)," *IEEE Transactions on Circuits and Systems for Video Technology* **22**(12), 1669–1684 (2012).