Rendering-dependent compression and quality evaluation for light field contents

Irene Viola[a], Keita Takahashi[b], Toshiaki Fujii[b], and Touradj Ebrahimi[a]

[a] Multimedia Signal Processing Group (MMSPG), École Polytechnique Fédérale de Lausanne (EPFL); Lausanne, Switzerland
[b]School of Engineering, Nagoya University; Nagoya, Japan

ABSTRACT

Light field rendering promises to overcome the limitations of stereoscopic representation by allowing for a more seamless transition between multiple point of views, thus giving a more faithful representation of 3D scenes. However, it is indisputable that there is a need for light field displays on which the data can be natively visualised, fuelled by the recent innovations in the realm of acquisition and compression of light field contents. Assessing the visual quality of light field contents on native light field display is of extreme importance in future development of both new rendering methods, as well as new compression solutions. However, the limited availability of light field displays restrict the possibility of using them to carry out subjective tests. Moreover, hardware limitations in prototype models may lessen considerably the perceptual quality of experience in consuming light field contents. In this paper, we compare three different compression approaches for multi-layer displays, through both objective quality metrics and subjective quality assessment. Furthermore, we analyze the results obtained through subjective tests conducted using a prototype multi-layer display, and a recently-proposed framework to conduct quality assessment of light field contents rendered through a tensor display simulator in 2D screens. Using statistical tools, we assess the correlation among the two settings and we draw useful conclusions for future design of compression solutions and subjective tests for light field contents with muti-layer rendering.

Keywords: Light field, multilayer display, tensor display, subjective quality evaluation, objective quality evaluation

1. INTRODUCTION

Light field photography has recently seen a surge of popularity due to the increased capabilities in acquiring and rendering real-life scenes in a more immersive way. In particular, In the past chapters, we have presented image-based rendering as a viable method to experience, and thus evaluate, light field contents on traditional 2D screens. Light field rendering promises to overcome the limitations of stereoscopic representation by allowing for a more seamless transition between multiple point of views, thus giving a more faithful representation of 3D scenes. However, it is undisputable that there is a need for light field displays on which the data can be natively visualized, fueled by the recent innovations in the realm of acquisition and compression of light field contents.

A promising solution for light field rendering uses a stack of programmable light-attenuating layers in front of a light-emitting source to provide depth cues without the need of glasses.[1–3] As only a few attenuating layers are required to render multiple points of view, the term "compressive display" has been used to define this type of rendering devices. The layer patterns to be displayed in each light-attenuating layers can be obtained from the multi-view light field data through Nonlinear Tensor Factorization (NTF).[3] Recently, a new method has been proposed to generate the layer patterns from a stack of focused images (focal stack), which greatly reduces the number of images that are needed as input for the tensor displays.[4] The method was tested in a prototype 3D display to prove its efficacy.[5]

Testing the visual quality of compressed and uncompressed light field contents on native light field display is of extreme importance in future development of both new rendering methods, as well as new compression

Further author information: (Send correspondence to first author)
E-mail: irene.viola@epfl.ch

(a) Back layer (b) Middle layer (c) Front layer

Figure 1: Depiction of layer patterns generated from the light field perspective views.

solutions. However, the limited availability of light field displays hinders the assessment of their visual quality. Moreover, hardware limitations in prototype models considerably lessen the perceptual Quality of Experience (QoE) in consuming light field contents. Being able to simulate light field multi-layer rendering in a virtual environment is thus helpful in conducting evaluation of visual quality for light field displays in an ideal scenario.

We have recently proposed a framework to conduct quality assessment of light field contents rendered through a tensor display simulator in 2D screens.[6] Through a GUI the layer patterns composing the multi-layer tensor displays are simulated in a 3D environment. By interacting with the mouse, users can experience the light field from different points of view.

Considering the possibilities that light field imaging offers in terms of rendering, optimizing a compression architecture in terms of visual quality can be proven challenging. The problem of assessing the effect of compression distortions on the rendered quality has been tackled in the past for image-based rendering. For example, Rizkallah et al. investigate the impact of compression of light field contents on extended focus and refocusing applications.[7] Similarly, Perra et al. report the results of applying HEVC-based compression on light field refocusing.[8] Adhikarla et al. analyse the effect of various distortions, including compression artifacts, on the visual quality of light field on 3D displays with motion parallax.[9]

Most of the compression solutions are designed and optimized for image-based rendering; in fact, the visual quality is conventionally assessed by measuring the level of distortion of the perspective views composing the 4D light field structure,[10–12] even when different rendering procedures were evaluated (e.g., depth-based rendering[13]). Nevertheless, the visual quality of light field contents under compression distortions may considerably vary under different types of rendering technologies. Moreover, rendering-agnostic solutions for light field compression may not be the most efficient solutions when specific displays are adopted. For example, multi-layer displays offer a multi-viewing experience while only rendering a few light-attenuating layers. Consequently, a large number of viewing angles can be experienced through the use of few, carefully created layer patterns. It is thus worth exploring whether new compression strategies can be specifically designed for multi-layer rendering, and how different approaches can affect the visual quality of the final rendered content.

In this paper we analyse three compression approaches modeled on the unique opportunities given by multi-layer displays, to provide a thorough study of the effects of employing different approaches on the visual quality of light field contents in a multi-layer renderer. We also explore the effect of using different double-stimulus variants to perform subjective quality assessment of light field contents, rendered through a simulated compressive display. Moreover, we perform the comparison between the results obtained with the simulator and a prototype multilayer display. To account for cross-cultural variance among different testing groups, we perform the same experiment in two different laboratory settings, and we analyse the correlation among the scores.

2. RENDERING-DEPENDENT CODING STRATEGIES

Considering the peculiarities of multi-layer rendering, three viable alternatives for light field compression are envisioned and defined:

- **Compression of light field perspective views.** The first, rendering-agnostic solution performs the compression on the perspective views, which will then be transmitted through the communication channel.

(a) Bikes (b) Danger de mort (c) Stone Pillars Outside (d) Fountain & Vincent (e) Friends

Figure 2: Central perspective view from each content used in the test.

The layer patterns are created at the receiver side, after the compressed views have been decoded. This solution has the advantage of being adaptable to any rendering system, as no assumption on how the views will be rendered is made on the compression stage. However, as pointed out by Takahashi *et al.*,[4] a large number of perspective views is needed to create a few light-attenuating layers. Thus, it might not be the most efficient solution when multi-layer displays are involved.

- **Compression of layer patterns.** The second solution performs the compression on the layer patterns directly, which can then be transmitted and rendered at the receiver side. This approach has the obvious advantage of compressing and transmitting only a few light-attenuating layers, thus gaining in compression efficiency. However, apart from being a rendering-dependent solution, which would require multiple transmissions for different devices, this approach might require ad-hoc algorithms to operate efficiently on the synthetic layer patterns, which are remarkably different from natural scenes (see Figure 1). Traditional image and video compression standards such as JPEG or HEVC are optimized for natural scenes; thus, compressing the light-attenuating layer patterns may result in a sub-par performance.

- **Compression of focal stack.** It was shown by Takahashi *et al.*[4] that, in order to reconstruct the layer pattern from the focal stack, we only need a number of focused images equal to the number of layer patterns which need to be rendered. The focal stack can be easily compressed using conventional image and video standards, unlike layer patterns, and its limited amount of images should guarantee a better coding efficiency with respect to the first approach. However, the construction of layer patterns from focal stacks is riddled with additional errors, due to the approximation in the tensor factorization.

3. EXPERIMENT DESIGN

In this section we will describe the subjective quality experiment we conducted to perform the analysis on quality assessment for simulated multi-layer displays.. In particular, we first list the dataset and the coding conditions. We then describe the lab settings in which the tests were conducted, as well as the employed methodologies. Finally, we give an overview of the statistical analysis we conducted on the gathered data.

3.1 Dataset and coding conditions

Five light field contents were selected from a publicly available database.[14] The contents were acquired with a Lytro Illum camera and processed using the Light Field Matlab Toolbox[15,16] to obtain a stack of 15×15 perspective image, each having a resolution of 625×434 pixels. Color and gamma corrections were applied on each perspective image for the rendering. To avoid unwanted distortions caused by the lenslet structure of the Lytro Illum camera, only the 9×9 central perspective views were selected for the test. The central perspective view from each content is displayed in Figure 2.

The three coding strategies described in Section 2 are employed for the test. For all three solutions, the state-of-the-art video encoding standard HEVC was employed for the compression, to ensure a fair comparison. To perform the encoding, the reference software HM was used.[17]

The layer patterns were created using the software implementation from Takahashi *et al.*[18] To create the focal stack, the Light Field Matlab Toolbox was employed.[15,16] In our validating test, the number of layers was fixed to $L = 3$.

The compression solutions were evaluated at four bit-rates, namely $R1 = 537$ kB, $R2 = 134$ kB, $R3 = 67$ kB, and $R4 = 27$ kB, corresponding to 0.2, 0.05, 0.025 and 0.01 bpp, respectively. The bpp are computed with respect to the original size of the 9×9 perspective views. The bit-rates were carefully chosen to cover the visual quality space while providing reasonable and fair comparison among the listed compression solutions.

3.2 Objective quality evaluation

To evaluate the impact of the distorsions caused by the proposed algorithms, PSNR and SSIM were selected from the literature to objectively assess the visual quality of the contents. The layer patterns obtained from the uncompressed light field were used as reference for each content.

The metrics were applied separately to the luma channel Y and for each layer pattern image, as follows:

$$PSNR_Y(l) = 10 \log_{10} \frac{255^2}{MSE(l)}, \tag{1}$$

$$SSIM_Y(l) = \frac{(2\mu_I\mu_R + c_1)(2\sigma_{IR} + c_2)}{(\mu_I^2 + \mu_R^2 + c_1)(\sigma_I^2 + \sigma_R^2 + c_2)}, \tag{2}$$

in which l is the index of each layer pattern used in the rendering, $MSE(l)$ is the mean square error, μ_I and μ_R are the mean values, σ_I^2 and σ_R^2 are the variances, and σ_{IR} is the covariance of the two perspective views in channel Y. Please note that unlike previous chapters, we are now computing the metrics on the layer patterns. PSNR was computed for chrominance channels U, V following Equation 1, and a weighted average[19] was calculated as follows:

$$PSNR_{YUV}(l) = \frac{6PSNR_Y(l) + PSNR_U(l) + PSNR_V(l)}{8} \tag{3}$$

The average PSNR value for Y channel was then computed across the viewpoint images:

$$\widehat{PSNR}_Y = \frac{1}{L} \sum_{l=1}^{L} PSNR_Y(l), \tag{4}$$

in which $L = 3$ represent the number of layer patterns. $\widehat{PSNR}_{YUV}$ and $\widehat{SSIM}_Y$ were analogously computed following Equation 4.

3.3 Subjective quality evaluation

For our experiments, the DSIS with 5-point grading scale (*5-Imperceptible, 4-Perceptible but not annoying, 3-Slightly annoying, 2-Annoying, 1-Very annoying*) was selected, according to the ITU-R Recommendation BT.500-13.[20] We used a toggle-based variant, which presents a user-driven intermittent presentation of the impaired and reference stimuli. This variant allows to compare the same region of interest in both stimuli without head or eye movements, by switching between the two contents; as such, it is particularly suitable for immersive multimedia in which a split screen would create unnatural effects and head movements could warrant unwanted consequences, such as omnidirectional imaging.[21] It is also to be preferred when testing just noticeable differences which would not be captured by a side-by-side presentation, and when the specifications of the screen on which the test is performed do not allow to display the contents in full resolution.

Participants were asked to rate the quality of the test stimuli when compared to the uncompressed reference. They could access the reference content by pressing a specific key, and they could return to the test content by pressing another designated key. Participants were only allowed to give a score when the test contents was being rendered on the screen, and at least one full switch between test and reference stimulus was required to perform

the rating. In order to accustom the participants with what distortions to expect in the test images, a training session was organized before the experiment. Three training samples, created by compressing one additional content on the test bit-rates, were manually selected by expert viewers.

All the compressed stimuli were shown in one session. Additionally, two hidden references per content were added to the test: one consisted in the layer patterns generated from the uncompressed stack of perspective views, while the other were created from the uncompressed focal stack. Thus, a total of 70 stimuli were evaluated. The display order of the stimuli was randomized for each participant, and the same content was never displayed twice in a row.

3.4 Test environments

Two laboratory settings were used for our tests, in the facilities of the École Polytechnique Fédérale de Lausanne (EPFL) and Nagoya University (NU).

In EPFL, a laboratory for subjective video quality assessment, which was set up according to ITU-R Recommendation BT.500-13,[20] was used for the test. A 27-inch Apple Display with native resolution of 2560×1440 pixels was used. The monitor settings were adjusted according to the following profile: sRGB Gamut, D65 white point, 120 cd/m^2 brightness, and minimum black level of 0.2 cd/m^2. The controlled lighting system in the room consisted of adjustable neon lamps with 6500 K color temperature against mid-grey background walls. The illumination level measured on the screens was 18 lux. Conforming to requirements in ITU-R Recommendation BT.2022,[22] the distance of the subjects from the monitor was approximately equal to 7 times the height of the displayed content. However, subjects were allowed to move further or get closer to the screen. A total of 20 subjects (10 males and 10 females) participated in the tests, amounting to 20 scores per stimulus per variant. Subjects were between 18 and 35, with a mean age of 23.29 years old. Before starting the test, all subjects were examined for visual acuity and color vision using Snellen and Isihara charts, respectively.

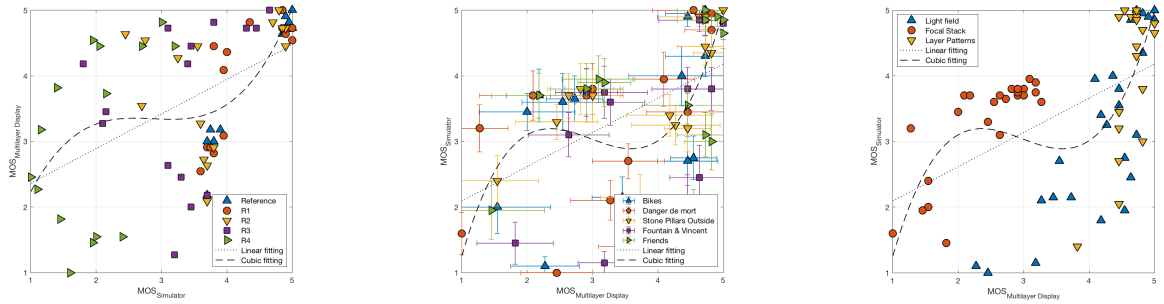
In NU, a controlled environment was selected to perform the experiment. However, no calibration on the lighting system for the room was conducted. A prototype multi-layer display was used to perform a pilot evaluation.[5] A total of 11 subjects (all males) took part in the test. Subjects were between 18 and 35, with a mean age of 23.28 years old. Before starting the test, all subjects were examined for visual acuity and color vision using Snellen and Isihara charts, respectively.

4. STATISTICAL ANALYSIS

Outlier detection and removal was performed on the results, independently for each test, according to the ITU Recommendations.[20] No outlier was detected in either batch of scores. After outlier removal, the Mean Opinion Score (MOS) was computed for each stimulus, independently for each methodology. The corresponding 95% Confidence Intervals (CIs) were computed assuming a Student's t-distribution.

In order to perform a benchmark of the objective quality metrics in predicting the subjective quality of light field contents rendered through the simulator, or the multi-layer display, several fittings were applied to the PSNR and SSIM values, following the Recommendation ITU-T J.149.[23] In particular, first order and third order fittings were used to compare the objective quality metrics to the MOS values, along with logistic fitting. Root Mean Square Error (RMSE), Pearson Correlation Coefficient (PCC), Spearman's Rank Correlation Coefficient (SRCC) and Outlier Ratio (OR) were computed for accuracy, linearity, monotonicity and consistency, respectively, according to Recommendation ITU-T P.1401.[24]

In order to draw a comparison among the different displays, no fitting, along with first and third order fittings, was also applied to the MOS values, and the performance indexes were computed. Moreover, multiple comparison tests were performed at a 5% significance level on the raw scores, to determine, for each stimulus, whether the MOS values obtained in different test settings, using different variants or different displays, were significantly different, and the percentage of correct estimation, underestimation and overestimation were computed. Additionally, the classification errors were computed using the same multiple comparison test to see if the results obtained with the tested conditions lead, for each pair of stimuli, to the same conclusions.[23]



(a) $MOS_{Simulator}$ as function of $MOS_{Multilayerdisplay}$.

(b) $MOS_{Multilayerdisplay}$ as function of $MOS_{Simulator}$.

(c) $MOS_{Multilayerdisplay}$ as function of $MOS_{Simulator}$.

Figure 3: Comparison of MOS values obtained with different displays, along with linear and cubic fittings. Points are differentiated by compression ratio (a), by content (b), and by compression solution (c).

Table 1: Performance indexes for the comparison among the multilayer display and the simulator.

	$[\widehat{MOS}_{Simulator}, MOS_{Multilayerdisplay}]$										
	PCC	**SRCC**	**RMSE**	**OR**	**Correct Est.**	**Under Est.**	**Over Est.**	**Correct Decision**	**False Ranking**	**False Diff.**	**False Tie**
No fitting	0.5244	0.5817	1.1277	44.29%	72.86%	20.00%	7.14%	52.46%	3.73%	27.25%	16.56%
Linear fitting	0.5244	0.5817	0.9734	48.57%	61.43%	15.71%	22.86%	47.08%	3.11%	17.27%	32.55%
Cubic fitting	0.6008	0.5906	0.9139	42.86%	68.57%	15.71%	15.71%	55.78%	0.62%	18.47%	25.13%
	$[\widehat{MOS}_{Multilayerdisplay}, MOS_{Simulator}]$										
	PCC	**SRCC**	**RMSE**	**OR**	**Correct Est.**	**Under Est.**	**Over Est.**	**Correct Decision**	**False Ranking**	**False Diff.**	**False Tie**
No fitting	0.5244	0.5817	1.1277	44.29%	72.86%	7.14%	20.00%	52.46%	3.73%	16.56%	27.25%
Linear fitting	0.5244	0.5817	0.9679	50.00%	74.29%	4.29%	21.23%	46.71%	1.12%	4.60%	47.58%
Cubic fitting	0.6270	0.6023	0.8855	42.86%	77.14%	4.29%	18.57%	56.60%	1.53%	11.59%	30.27%

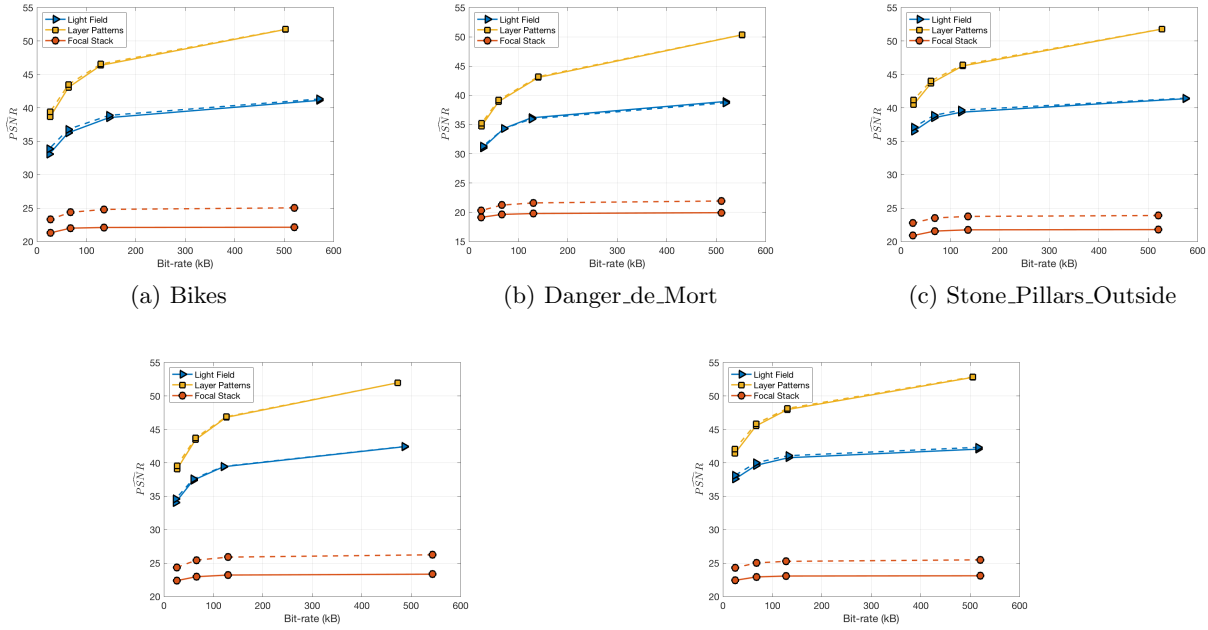
5. RESULTS AND DISCUSSION

In this section, results of objective and subjective evaluations are discussed, similarities and differences between prototype and simulated displays are presented, and a benchmark of objective quality metrics for visual quality prediction is given. First, in Section 5.1, we compare the subjective evaluation results obtained with the two displays. Then, in Section 5.2 we show the results of the objective evaluation, and we present and compare the results of the subjective assessment obtained with the two displays. Finally, in Section 5.3 we carry out a benchmarking of the objective quality metrics using both subjective assessment tests as the ground truth.

5.1 Comparison of different displays

Figure 3 shows the results of the comparison between the tests performed in the EPFL laboratory setting using the simulator, and in the NU laboratory using a prototype multilayer display. Points are shown as divided by compression ratio, content and compression solution, with linear and cubic fittings. Table 1 shows the results of the performance indexes for the sets of scores obtained with both displays.

Results of the comparison show that poor correlation is achieved between the results obtained using the simulator and the results associated to the multi-layer display, with PCC values as low as 0.5244. By observing Figure 3 (a) and (b), no visible trend can be observed regarding the compression ratios or contents employed in the tests; rather, it is easy to notice a clear pattern regarding the way scores are distributed in relation to the compression solution that was employed. In particular, when the layer patterns were directly encoded, no difference seems to be perceived in their quality, regardless of their compression ratio, when the multilayer display was employed for the test (Figure 3 (c), yellow points); in fact, the large majority of the points lay between MOS values of 4 and 5, whereas they span the entire axis when the simulator is used. A similar trend can be observed for part of the contents that employed traditional light field compression, although in some cases, values of MOS close to 2 were given, when the contents were compressed at the highest compression ratio (compare with Figure 3 (a)). It is interesting to see that, when the focal stack method was employed to compress the



(a) Bikes (b) Danger_de_Mort (c) Stone_Pillars_Outside

(d) Fountain_&_Vincent_2 (e) Friends_1

Figure 4: Results of $\widehat{PSNR}_Y$ (solid line) and $\widehat{PSNR}_{YUV}$ (dashed line) vs bitrate for different contents.

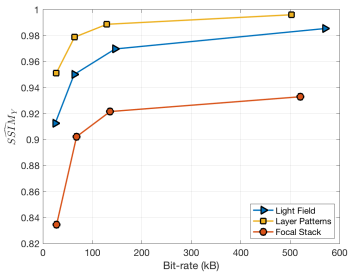
contents, the MOS scores seldom reached values higher than 3 when the multi-layer display was used. However, the method did not reach transparent quality either when the simulator was selected for the test: indeed, the MOS values always stop short of 4.

Results of the multiple comparison show that the scores obtained in the two tests were statistically equivalent in $60-75\%$ of the cases, depending on what fitting has been applied to the scores. More importantly, the two tests seem to agree on the ranking to be given to the various stimuli on around 50% of the cases. The results are most likely caused by the hardware limitations of the prototype display, which do not allow to differentiate among compression artifacts when the layer patterns are directly compressed. In general, it appears that when the multilayer display is used, the choice of generating layer patterns from either the light field data or the focal stack has a greater impact on the visual quality, than the compression ratio chosen to encode the data. Using the simulator, on the other hand, allows to more easily differentiate among different levels of compression. This is likely due to the fact that the simulator offers an ideal scenario for multilayer rendering; the same cannot be said for the prototype display, whose LCD panels did not achieve the same level of transparency, thus affecting the quality of the rendered contents.

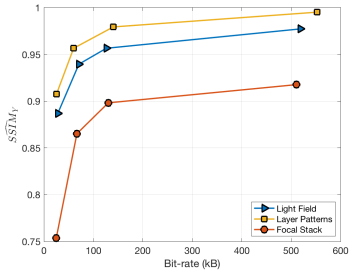
5.2 Objective and subjective results

Figures 4 and 5 show the results of the objective quality metrics, for all bitrates. Plots show that compressing the focal stack leads to a sharp decrease in performance when $\widehat{PSNR}_Y$ and $\widehat{PSNR}_{YUV}$ are used as metric, with a drop of $\sim$ 30dB for high bitrates, and of $\sim$ 15dB for low bitrates. Results for $\widehat{SSIM}_Y$ also indicate that the focal stack approach leads to reduced quality in the layer pattern data.

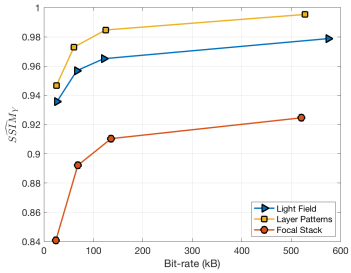
Figure 6 depicts the MOS scores obtained with the use of a simulator, whereas Figure 7 illustrates the MOS scores collected using a prototype multi-layer display. It has already been observed in the previous section that the scores collected with different displays do not exhibit very good correlation. It is evident by considering how the first two approaches (compression of light field and layer patterns, respectively) consistently achieve near-transparent quality, regardless of the bitrate, when the multi-layer display is employed in the evaluation (Figure 7). On the other hand, compression of focal stack data leads to strongly perceived distortions, as



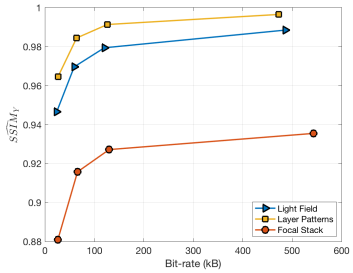
(a) Bikes



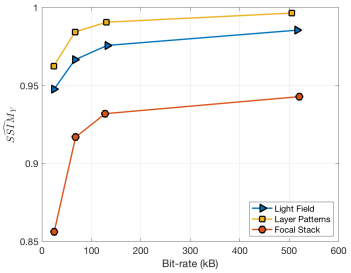
(b) Danger_de_Mort



(c) Stone_Pillars_Outside



(d) Fountain_&_Vincent_2



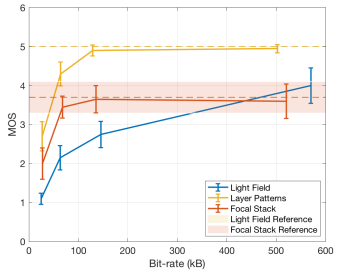
(e) Friends_1

Figure 5: Results of $\widehat{SSIM}_Y$ vs bitrate for different contents.

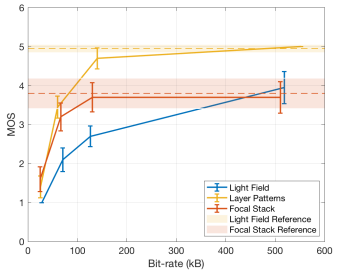
the scores associated with the uncompressed reference are consistently under the MOS score of 4. Although the compression of focal stack data never reaches transparent quality when using the simulator (Figure 6), it performs comparably to a direct compression of light field data for high bitrates, and in some cases outperforms the latter.

It is interesting to notice the difference in the CIs associated with the reference layer patters between the scores assigned using either display. When the simulator is adopted, the CIs are consistently small. This is to be expected, considering that the DSIS variant that is being used highlights subtle differences - or, in the case of the reference, the lack thereof. However, it is notable to see that the same does not happen when the multi-layer display is used. In this case, the large CIs would suggest that alterations and artifacts were perceived even when no difference between test and reference content was materially present. It would be worth exploring whether the uncertainty associated with the scores is a reflection of the hardware limitations of the prototype multi-layer display, which may interfere with the perception of distortions in subjective tests.

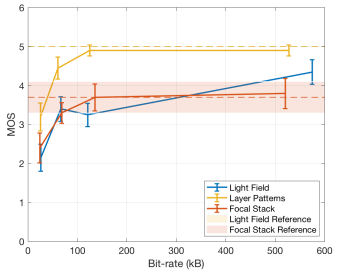
The graphs show that operating the compression on the layer patterns seems to be the preferable solution, as its performance is either statistically equivalent or better than the other solutions at all bitrates, for both displays. However, no definite answer can be given whether compressing the focal stack would be a preferrable solution with respect to encoding the entire light field, as contradicting results are obtained in the two tests. This is particularly evident when analyzing the results of the pairwise comparison among the scores, depicted in Figure 8 for the simulator test, and in Figure 9 for the prototype one. The boxes represent the number of contents for which the compression approach in each row is significantly better than the approach in each column. In the first case, encoding the layer patterns is the clear winning solution, as it outperforms the other two approaches in at least 4 out 5 contents; encoding the focal stack is the second preferred solution, as it fares better than encoding the light field data on more than half the contents for all bitrates, except the highest. The results, however, are overturned when considering the prototype multi-layer display (Figure 9): in this case, compression applied on the focal stack is never significantly better than the other two approaches, and it is nearly always outperformed. Layer pattern and light field encoding are statistically equivalent for high bitrates, but for lower bitrates the first approach leads to significantly better results.



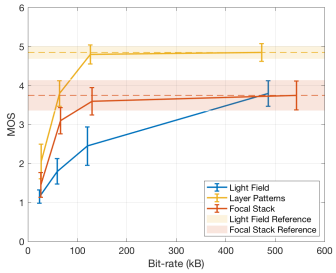
(a) Bikes



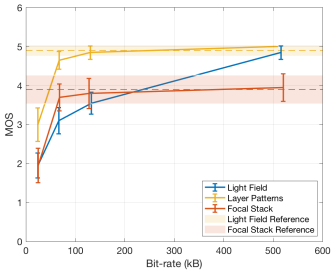
(b) Danger_de_Mort



(c) Stone_Pillars_Outside



(d) Fountain_&_Vincent_2



(e) Friends_1

Figure 6: MOS vs bitrate for different contents, with respective CIs. Results obtained using the simulator.[6]

Table 2: Performance indexes for the comparison among different objective quality metrics, fitted to the MOS scores obtained with the simulator.

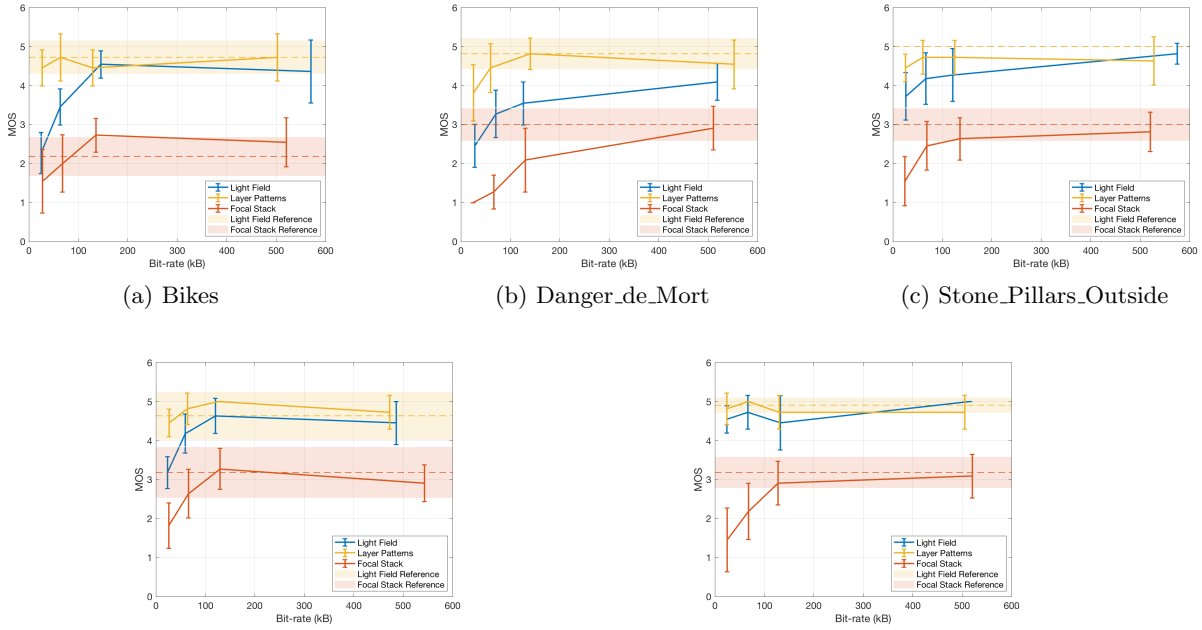
	$[\widehat{PSNR}_Y, MOS]$				$[\widehat{PSNR}_{YUV}, MOS]$				$[\widehat{SSIM}_Y, MOS]$			
	PCC	**SRCC**	**RMSE**	**OR**	**PCC**	**SRCC**	**RMSE**	**OR**	**PCC**	**SRCC**	**RMSE**	**OR**
Linear fitting	0.4563	0.5899	1.0164	86.67%	0.4809	0.5912	1.0014	88.33%	0.5982	0.6833	0.9153	80.00%
Cubic fitting	0.6502	0.5899	0.8678	83.33%	0.6490	0.5912	0.8690	85.00%	0.6823	0.6833	0.8351	85.00%
Logistic fitting	0.4226	0.5899	1.0371	86.67%	0.4479	0.5912	1.0222	85.00%	0.5826	0.6833	0.9289	80.00%

5.3 Benchmarking of objective quality metrics

Figures 10 and 11 depict the scatter plots between the objective quality metric results and the corresponding MOS scores, obtained using the simulator and the prototype multi-layer display, respectively. Tables 2 and 3 report the performance indexes for each metric, using linear, monotonic cubic and logistic fitting. Low values of PCC and SRCC confirm that the objective quality metrics are very poorly correlated with the subjective scores collected using the simulator. Among the three metrics, $\widehat{SSIM}_Y$ seems to perform slightly better (PCC = 0.6823 and SRCC = 0.6833 when cubic fitting is applied), although high levels of RMSE and OR indicate that accuracy and consistency are still lacking.

When considering the MOS scores collected using the multi-layer display, however, results show a strong correlation with all the objective quality metrics; again, $\widehat{SSIM}_Y$ is the best performing one, achieving PCC = 0.9348 and SRCC = 0.8904 with cubic fitting.

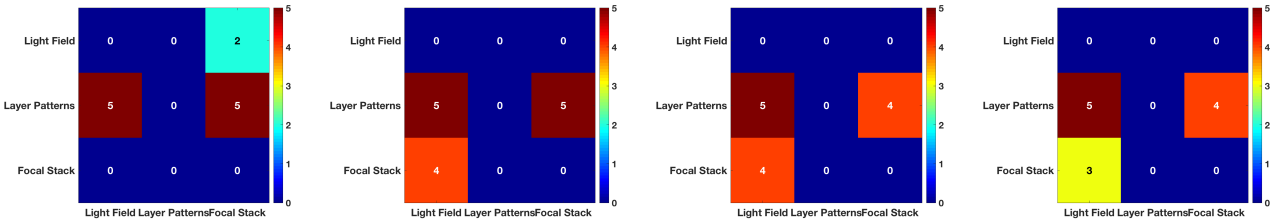
Results show that all objective quality metrics are good predictors for visual quality of light field contents when visualized through a prototype display. However, considering the level of uncertainty associated with the scores, as proven by the large CIs for the uncompressed reference stimuli, we are hesitant in recommending the use of said objective quality metrics as quality predictors for compression artifacts. As shown also in the previous sections, the method employed to generate the layer patterns seems to have a higher impact on the final subjective scores, at least when the prototype display is used. Thus, it is safe to assume that objective



(a) Bikes (b) Danger_de_Mort (c) Stone_Pillars_Outside

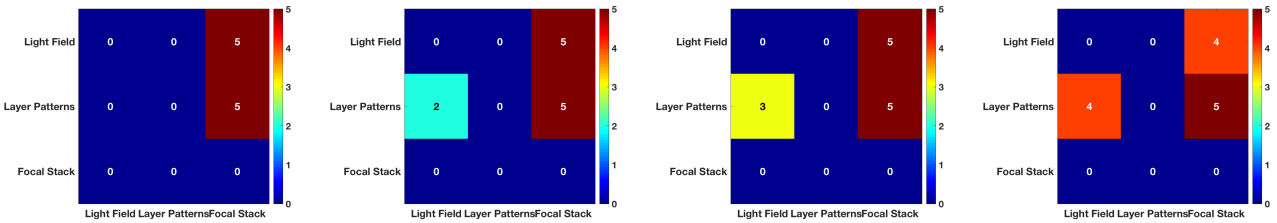
(d) Fountain_&_Vincent_2 (e) Friends_1

Figure 7: MOS vs bitrate for different contents, with respective CIs. Results obtained using the prototype display.



(a) R1 (b) R2 (c) R3 (d) R4

Figure 8: Pairwise comparison of codecs for different bitrates, for results obtained using the simulator.



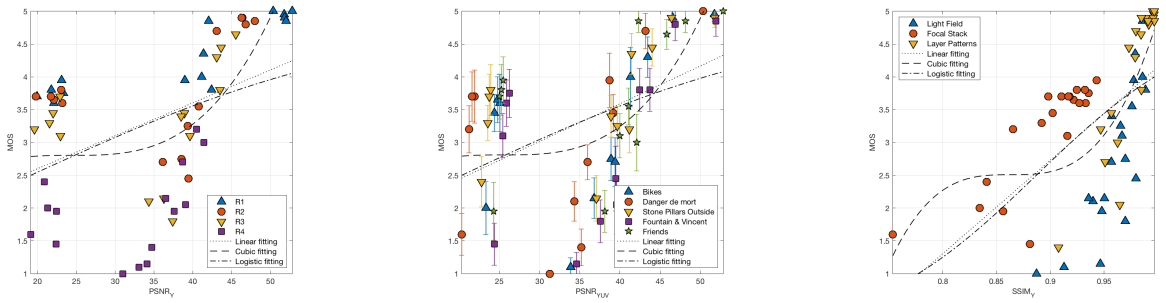
(a) R1 (b) R2 (c) R3 (d) R4

Figure 9: Pairwise comparison of codecs for different bitrates, for results obtained using the prototype display.

quality metrics would give a reliable estimation on which method for layer pattern generation would lead to the best visual quality. However, when compression artifacts need to be assessed and compared, new objective quality metrics should be developed. Moreover, particular care should be given in assessing whether the hardware limitations of the display in use restrict the perception of compression distortions; in that case, the use of an

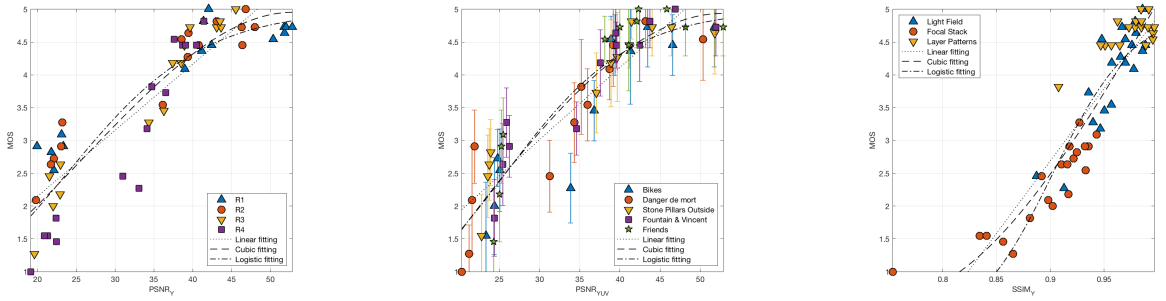
Table 3: Performance indexes for the comparison among different objective quality metrics, fitted to the MOS scores obtained with the multi-layer display.

	$[\widehat{PSNR}_Y, MOS]$				$[\widehat{PSNR}_{YUV}, MOS]$				$[\widehat{SSIM}_Y, MOS]$			
	PCC	**SRCC**	**RMSE**	**OR**	**PCC**	**SRCC**	**RMSE**	**OR**	**PCC**	**SRCC**	**RMSE**	**OR**
Linear fitting	0.8999	0.9110	0.5075	40.00%	0.9031	0.9135	0.4998	40.00%	0.9101	0.8904	0.4821	31.67%
Cubic fitting	0.9191	0.9110	0.4584	33.33%	0.9285	0.9135	0.4322	26.67%	0.9348	0.8904	0.4134	23.33%
Logistic fitting	0.9148	0.9110	0.4705	31.67%	0.9255	0.9135	0.4410	25.00%	0.9232	0.8904	0.4557	31.67%



(a) MOS as function of $\widehat{PSNR}_Y$. (b) MOS as function of $\widehat{PSNR}_{YUV}$. (c) MOS as function of $\widehat{SSIM}_Y$.

Figure 10: Comparison of performance of different objective quality metrics in predicting the MOS scores obtained with the simulator, along with linear, cubic and logistic fittings. Points are differentiated by compression ratio (a), by content (b), and by compression solution (c).



(a) MOS as function of $\widehat{PSNR}_Y$. (b) MOS as function of $\widehat{PSNR}_{YUV}$. (c) MOS as function of $\widehat{SSIM}_Y$.

Figure 11: Comparison of performance of different objective quality metrics in predicting the MOS scores obtained with the multi-layer display, along with linear, cubic and logistic fittings. Points are differentiated by compression ratio (a), by content (b), and by compression solution (c).

error-free, ideal rendering scenario should be preferred.

6. CONCLUSIONS

In this paper, we presented a comparison of different compression solutions and subjective quality assessment scenarios for light field contents on multi-layer tensor displays. We performed different sets of experiments in two separate laboratory settings, using both a prototype tensor display and a simulator for 2D screens.

Further work is needed to confirm whether a different DSIS variant could lead to different results in the comparison between displays. Moreover, it is worth analysing how the visual quality of light field rendering is affected by the hardware limitations of the available multi-layer displays, for example by doing a benchmarking of existing displays. New objective quality metrics should be designed to successfully estimate the impact of coding

solutions on the visual quality of light field contents. Furthermore, great care should be employed in designing subjective tests to assess the visual quality associated with multi-layer displays. In particular, the hardware limitations linked to prototype displays may lead to a poor quality of experience, which can be reflected on the distribution of the collected subjective scores.

ACKNOWLEDGMENTS

This work has been conducted in the framework of projects "Light field Image and Video coding and Evaluation" and "Advanced Visual Representation and Coding in Augmented and Virtual Reality" both funded by The Swiss National Foundation for Scientific Research under grant numbers 200021-159575 and 200021-178854.

REFERENCES

[1] Lanman, D., Hirsch, M., Kim, Y., and Raskar, R., "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," in [*ACM Transactions on Graphics (TOG)*], **29**(6), 163, ACM (2010).

[2] Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., and Raskar, R., "Beyond parallax barriers: applying formal optimization methods to multilayer automultiscopic displays," in [*Stereoscopic Displays and Applications XXIII*], **8288**, 82880A, International Society for Optics and Photonics (2012).

[3] Wetzstein, G., Lanman, D., Hirsch, M., and Raskar, R., "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," (2012).

[4] Takahashi, K., Kobayashi, Y., and Fujii, T., "From focal stack to tensor light-field display," *IEEE Transactions on Image Processing* **27**(9), 4571–4584 (2018).

[5] Kobayashi, Y., Kondo, S., Takahashi, K., and Fujii, T., "A 3-D display pipeline: Capture, factorize, and display the light field of a real 3-D scene," *ITE Transactions on Media Technology and Applications* **5**(3), 88–95 (2017).

[6] Viola, I., Takahashi, K., Fujii, T., and Ebrahimi, T., "A comprehensive framework for visual quality assessment of light field tensor displays," (2019).

[7] Rizkallah, M., Maugey, T., Yaacoub, C., and Guillemot, C., "Impact of light field compression on refocused and extended focus images," in [*2016 24th European Signal Processing Conference (EUSIPCO)*], (2016).

[8] Perra, C. and Giusto, D., "An analysis of hevc compression for light field image refocusing applications," in [*2018 IEEE Seventh International Conference on Communications and Electronics (ICCE)*], 273–277, IEEE (2018).

[9] Kiran Adhikarla, V., Vinkler, M., Sumin, D., Mantiuk, R. K., Myszkowski, K., Seidel, H.-P., and Didyk, P., "Towards a quality metric for dense light fields," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 58–67 (2017).

[10] Viola, I., Řeřábek, M., and Ebrahimi, T., "Comparison and evaluation of light field coding approaches," *IEEE Journal of selected topics in signal processing* (2017).

[11] Viola, I. and Ebrahimi, T., "Quality assessment of compression solutions for ICIP 2017 Grand Challenge on light field image coding," *2018 International Conference on Multimedia and Expo Workshops* (2018).

[12] Viola, I., Maretic, H. P., Frossard, P., and Ebrahimi, T., "A graph learning approach for light field image compression," in [*Applications of Digital Image Processing XLI*], **10752**, 107520E, International Society for Optics and Photonics (2018).

[13] Graziosi, D. B., Alpaslan, Z. Y., and El-Ghoroury, H. S., "Compression for full-parallax light field displays," in [*Stereoscopic Displays and Applications XXV*], **9011**, 90111A, International Society for Optics and Photonics (2014).

[14] Řeřábek, M. and Ebrahimi, T., "New light field image dataset," *8th International Conference on Quality of Multimedia Experience (QoMEX)* (2016).

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

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

[17] ITU-T Q.6/SG 16 and ISO/IEC JTC 1/SC 29/WG 11, "High Efficiency Video Coding (HEVC) reference software HM." [Online]. Available: https://hevc.hhi.fraunhofer.de/trac/hevc/browser/trunk.

[18] Takahashi, K., "Light field display project." Available at http://www.fujii.nuee.nagoya-u.ac.jp/~takahasi/Research/LFDisplay/.

[19] 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).

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

[21] Perrin, A.-F., Bist, C., Cozot, R., and Ebrahimi, T., "Measuring quality of omnidirectional high dynamic range content," in [*Applications of Digital Image Processing XL*], **10396**, 1039613, International Society for Optics and Photonics (2017).

[22] 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).

[23] ITU-T J.149, "Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)." ITU (Mar. 2004).

[24] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models." International Telecommunication Union (July 2012).