2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)

An in-depth analysis of single-image subjective quality assessment of light field contents

Irene Viola and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

***Abstract*—Quality assessment of light field images poses new questions and challenges, due to the enriched nature of the content and the possibilities it offers at the rendering step. Image-based rendering is conventionally used to showcase the increased capabilities of light field contents on traditional 2D screens. However, the range of possibilities for rendering parameters is virtually endless, which poses the problem of what rendered images should be used when performing visual quality assessments, as well as how to properly present them to subjects during quality evaluations. Single-image assessment has been used in the past to conduct subjective quality evaluations. Since this type of assessment generates a large number of stimuli to be evaluated, which increases the complexity, length, and cost of the test, it is fundamental to analyze whether the added strain on the evaluation procedure is compensated by statistically relevant results. In this paper, we analyze the results of a subjective evaluation campaign that used single-image assessment by means of statistical tools, to understand whether the advantages of evaluating light field contents through separately rendered images counterbalance the increase in complexity. In particular, we test whether different types of rendering lead to statistically different ratings, and if testing a variety of rendering parameters through single-image assessment is advisable. Results provide useful guidelines to designs more efficient subjective quality assessment for light field contents.**

***Index Terms*—light field, subjective evaluation, quality assessment, subjective methodologies**

I. Introduction

For any type of multimedia content, reliable quality assessment is of paramount importance in the design and validation of new compression solutions that aim at reducing the size of the original data without compromising its perceptual quality. While objective metrics have been developed in the last decades to effectively predict the perceptual quality of the contents under assessment, subjective evaluations remain the most reliable means to measure the quality of media contents. In particular, subjective evaluation of visual quality is of fundamental importance when deciding which compression solution should be used. In that matter, light field contents pose new questions and challenges, due to their enriched nature and the possibilities available for the rendering step.

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).

978-1-5386-8212-8/19/$31.00 ©2019 IEEE

One of the most natural and intuitive ways to consume light field contents would involve light field displays or simulators to create a multi-view, 3D rendering of the contents [1], [2]. Using this approach, the full potential of light field imaging is exploited to create a 3D representation of the scene in front of the user. However, such displays are not widely available to consumers, due to their cost and their requirements.

Another possible approach to render light field contents relies on image-based rendering [3]. Indeed, image-based rendering offers an impressive showcase of the rendering abilities made possible by light field technology. Among other possibilities, the point of view of the scene can be modified, digital refocusing can be applied to highlight a specific plane in the image, zooming can be performed to exclude some planes in the scene, and so on. Due to the nearly endless possibilities that are offered, the first challenge that is posed for any type of image-based light field evaluation is to select which rendering to take into account when evaluating the contents. Such challenge becomes particularly dire when the images obtained after the rendering procedure are evaluated one by one, as the length and complexity of the test grows exponentially. Equally challenging is obtaining a single score for each content under assessment, since several rendered images are presented and subsequently rated in a separate fashion. It is thus not clear what is the best method to obtain a unique rating for the entire light field content.

Single-image assessment was chosen as the subjective evaluation methodology for the ICME 2016 Grand Challenge on Light-Field Image Compression [4]. Five algorithms were received as response to the challenge [5]–[9], and were evaluated against the anchor of choice, namely JPEG, using both objective and subjective quality assessment methods. The number of possible renderings was purposely constrained to avoid an overly complex assessment scenario; even so, a total number of 720 stimuli was generated for the subjective evaluation. Hence, it is crucial to analyse the results of such a massive campaign, to assess whether the added complexity of the test was justified by a diversification in ratings among different rendering procedures. To this aim, in this paper we use statistical tools such as ANOVA to examine the similarities among the ratings, and in particular to determine whether there are statistically relevant differences among different rendered images. Results are decisive in selecting the best

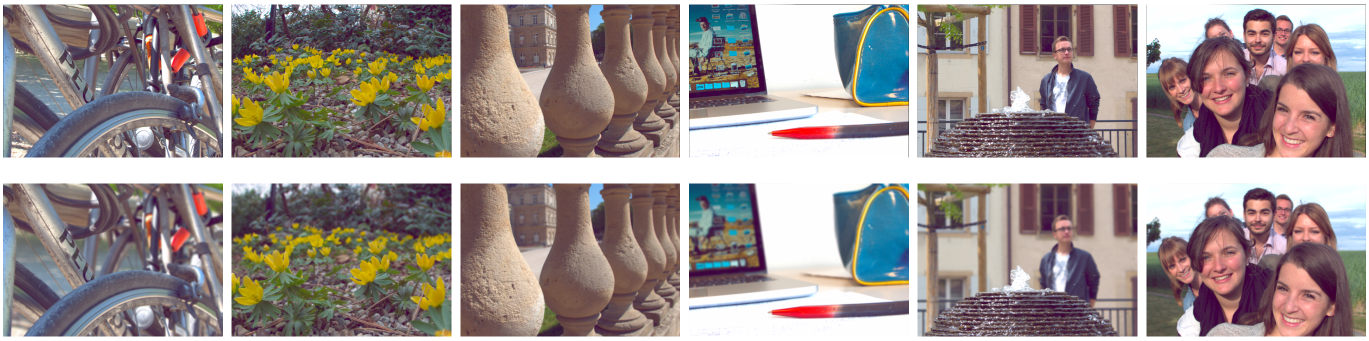

Bikes Flowers Stone_Pillars_Outside Desktop Fountain_&_Vincent_2 Friends_1

Fig. 1: Examples of central all-in-focus perspective view (first row) and view refocused on frontal plane (second row) from each content used in the experiments.

evaluation methodology for visual quality assessment of light field contents.

Our previous paper provides a presentation of the grand challenge results, including comparisons among the codecs and guidelines on different approaches [4]. However, the shortcomings and advantages of using single-image assessment to evaluate light field contents were not discussed. We aim at bridging the gap by providing a comprehensive analysis of the results obtained in the subjective evaluations. In the following sections, we will give a brief summary of relevant work; then, we will describe the subjective evaluation methodology, as well as the statistical tools used to perform the analysis, and we will present the results of our inquiry.

II. Related work

Several approaches have been adopted in literature to subjectively and objectively evaluate the quality of light field contents using image-based rendering. In [10], the visual quality of compressed light field contents was objectively evaluated using 9 perspective views and 3 refocused views, in order to test different rendering modalities. Filipe et al. [11] assessed the performance of several state-of-the-art focus metrics in the evaluation of extended depth of field (all-in-focus) images acquired by a focused plenoptic camera. In particular, they performed a subjective assessment in which they compared extended depth of field images obtained using optimal patch sizes, against images obtained with slightly larger patch sizes. In [12], the authors used the central, all-in-focus perspective view to subjectively assess the visual quality of light field contents affected by compression artifacts, using Paired Comparison (PC) as protocol. Paudyal et al. [13] analyzed the impact of different visualization techniques, including image-based assessment of all-in-focus perspective views and refocused views using Absolute Category Rating (ACR), concluding that there is high correlation between the scores obtained by image-based evaluation when compared to the corresponding animation-based passive evaluation. Battisti et al. [14] used an animation-based passive approach to study the effect of different trajectories on the perceived visual quality of light field contents. Perra [15] assessed the quality of rendered light field contents using a uniform circular animation, and proposed a new SSIM-based objective quality metric to predict its visual quality. In our previous work, we performed a comparison of different coding approaches using both objective and subjective assessment [16]. For the latter, both an interactive setup and an animation-based presentation were used to collect the results.

III. Data preparation and coding conditions

The coding conditions were defined in the context of the aforementioned ICME 2016 Grand Challenge [4]. Light field images created with a Lytro Illum plenoptic camera were selected as input. In particular, a lenslet image, which was created from the raw 10-bit sensor data by applying devignetting, demosaicing, clipping to 8-bit and color space conversion from RGB444 to YUV420, was to be compressed by the participants to the challenge. A total of six light field images were selected from a publicly available dataset to be used in the subjective quality assessment [17]. Thumbnails of rendered views for each content are indicatively depicted in Figure 1. Four fixed compression ratios were selected to evaluate the performance of the proposed compression algorithms, namely $R1 = 10:1$ (1 bpp), $R2 = 20:1$ (0.5 bpp), $R3 = 40:1$ (0.25 bpp), $R4 = 100:1$ (0.1 bpp). The ratios were computed with respect to the size of the raw data obtained from the camera ($5368 \times 7728 \times 10 = 414'839'040$ bit).

After compression and decompression, each lenslet image was converted to a stack of all-in-focus perspective views using the MATLAB Light Field Toolbox v0.4 [18], [19]. The stack contained 15×15 perspective views, each of resolution 434×625 pixels. Color and gamma corrections were applied to each perspective view prior to visualization. Analogously, the reference light field structure of perspective views was created from the uncompressed YUV420 lenslet image.

For each content, three all-in-focus perspective views were directly extracted from the light field data structure to be subjectively assessed. Namely, from the 15×15 stack of perspective views, those at indexes $(8, i)$, where $i = 5, 8, 11$,

TABLE I: Values of slope for refocused views.

Image ID	Slope 1	Slope 2
Bikes	-0.65	0.22
Flowers	-0.3	0.3
Stone_Pillars_Outside	-0.5	0.2
Desktop	-0.5	0.5
Fountain_&_Vincent_2	-0.5	0.35
Friends_1	-0.15	0.2

were selected to represent different perspectives of each scene. Additionally, the MATLAB Light Field toolbox was used to create two refocused views for each light field content, using a modified version of the function *LFFiltShiftSum*. In particular, perspective views from index 5 to index 11 (7×7 views) were used for the computation, leading to a rather large depth of field that still showed the effects of refocusing. Two slopes were selected in order to focus the image on two semantically relevant planes in the scene, as listed in Table I. Figure 1 shows one example of refocused view for each content, using Slope 1. The three all-in-focus perspective views, along with the two refocused views, form the five views that were used to assess each compressed content.

In total, six algorithms (five proponents [5]–[9] and one anchor) were evaluated in the subjective assessment tests. From each of the six light field contents compressed at 4 different bitrates, five rendered views were created. Thus, a total of 720 stimuli were assessed by human subjects in the test.

IV. Subjective quality assessment methodology

The methodology selected to conduct the subjective tests was based on Double Stimulus Continuous Quality Scale (DSCQS). Two images in native resolution (625×434 pixels) were presented simultaneously in a side-by-side fashion. One of the two images was always the uncompressed reference, and its position on the screen was randomized. The other image was compressed by one of the evaluated algorithms, at one of the selected bitrates. Both the reference and the test images were visualized using the same rendering parameters. Subjects were asked to rate the quality of reference and test images separately, on a discrete scale from 5 (Excellent) to 1 (Bad). Although they were informed that one of the visualized images would always be the reference, they did not receive any indication on its relative position on the screen. Before the experiments, a training session was organized to help subjects to adjust to the peculiarities of light field rendering, and to help them detect various distortions and compression artifacts. In particular, great care was applied in instructing subjects not to consider refocusing blur as a distortion, to avoid bias towards certain rendered images. To do so, five training samples (one for each rendered view) were generated using an additional content from the light field database [17].

To perform the tests, the QualityCrowd 2 framework [20] was modified to suit the DSCQS methodology. The experiment was split into four sessions, each containing 180 pairs of images. Each session lasted for approximately 45 minutes. At the beginning of first session, one dummy example was shown to ease the subject into the task, and its corresponding scores were subsequently discarded. The display order of the stimuli was randomized, and the same content was never displayed twice in a row. Each subject took part in two sessions, and a break of ten minutes was enforced between the sessions to avoid fatigue. A MacBook Pro 15.4-inch display was used for the test, with a resolution of 2880×1800 pixels. Only one observer per viewing session was employed, and the environment conditions were not fixed.

Overall, 35 naïve subjects (24 males and 11 females) participated in the subjective experiments, each rating 360 stimuli over the course of two sessions. Subjects were between 18 and 33 years old, with an average and median age of 22.4 and 22 years old, respectively. All subjects were screened for correct visual acuity with Snellen charts, and color vision using Ishihara charts.

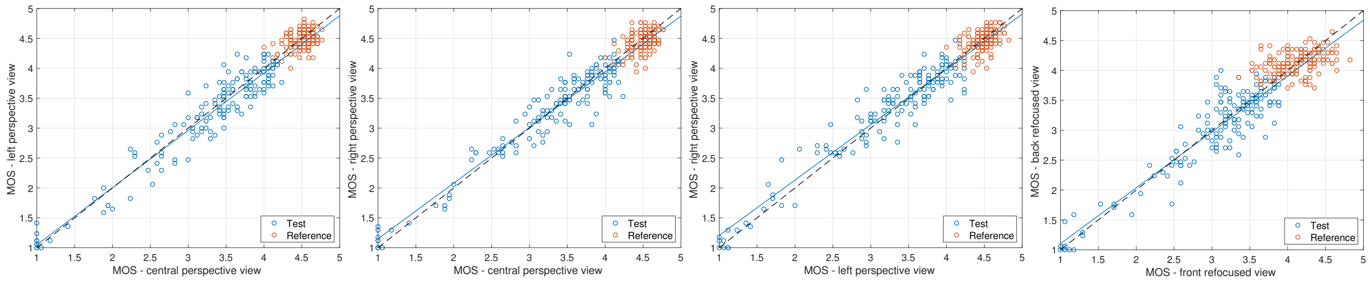
V. Data processing and statistical analysis

Outlier detection and removal was performed according to the ITU Recommendations [21]. One subject was found to be an outlier, and the corresponding scores were discarded, thus leading to 17 scores per stimulus. After outlier removal, the Mean Opinion Score (MOS) was computed for each coding condition j (i.e. for each content, view, proponent and bitrate). The corresponding 95% Confidence Intervals (CIs) were computed using Student's t-distribution [22]. Analogously, the MOS score was computed for each reference stimulus, separately for each rendering condition $\bar{j}$ (i.e., for each refocused and perspective view).

In order to determine whether statistically significant differences are present among the ratings given for differently rendered images, we performed analysis of variance (ANOVA) on the data. In particular, we first performed a one-way ANOVA on the MOS scores associated with each coding condition j, analyzing whether there was any statistical difference associated with each rendering parameters, using as null hypothesis the equivalence of all group means. We then performed the same analysis on the MOS scores associated with each rendering condition $\bar{j}$, to further understand whether any difference in performance could be due to coding artifacts. Finally, we performed a multi-way ANOVA on the full set of scores to gain insights on the statistical differences within the coding conditions, using as null-hypothesis the equivalence of each group means for the main effects, and the absence of non-zero interaction terms for the interaction factors.

VI. Results and discussion

Figure 2 compares the MOS scores given to each perspective and refocused view, for each test content and respective reference. As further showed by the linear fitting for test contents, the scores are evenly distributed along the $y = x$ line, proving that strong correlation can be found between the scores assigned to perspective and refocused views, within their rendering group. To further demonstrate that different perspective and refocused views were scored similarly within



(a) Central vs left perspective view (b) Central vs right perspective view (c) Right vs left perspective view (d) Front vs back refocused view

Fig. 2: Comparison between MOS values for different perspective views and different refocused views, for test contents (blue) and respective references (orange), with relative linear fitting. The dashed black line represents the $y = x$ function.

the same type of rendering, we perform a one-way ANOVA on the test scores, using as factor the corresponding perspective (central, left or right) or refocused (front or back) view. Tables II and III show the results of the analysis. As made clear by the large p-values (0.3071 and 0.341 for the perspective and refocused views, respectively), the scores assigned to test contents rendered through different perspective and refocused views were deemed statistically equivalent, as they failed to reject the null hypothesis of equal means at 1% significance level. Thus, only one representative of each group could have been used in the test, sensibly reducing the complexity and length of the evaluation, without causing any disruption in the collected scores. Results from one-way ANOVA applied on the reference data show similar trends ($p = 0.0166$ and $p = 0.772$ for perspective and refocused views, respectively).

TABLE II: One-way ANOVA on the raw scores given to test contents rendered through perspective views.

	SumSq	**DF**	**MeanSq**	**F**	**p-value**
Perspective views	3.3	2	1.63249	1.18	0.3071
Error	10148.4	7341	1.38243		

TABLE III: One-way ANOVA on the raw scores given to test contents rendered through refocused views.

	SumSq	**DF**	**MeanSq**	**F**	**p-value**
Refocused views	1.15	1	1.1489	0.91	0.341
Error	6199.85	4894	1.26683		

Once the correlation within the groups of views has been analyzed, we investigate whether there is any difference to be found between the two groups. Figure 3 shows the comparison between the MOS values assigned to all the perspective views, with respect to the MOS values associated with the refocused views. As showed in the plot, the vast majority of points fall below the $y = x$ line, signifying that the scores assigned to the perspective views were steadily higher than their refocused counterpart. The same trend can be observed not only for the test contents, but for the references as well. Despite being trained on considering only the differences between test and reference images, subjects consistently gave lower ratings to both test and reference contents when presented with refocused views, whereas for perspective views the ratings were usually higher. Results from the one-way ANOVA, summarized in Tables IV and V, confirm that the two groups reject the equal mean hypothesis ($p = 1.78763\text{e}{-25}$ and $p = 3.06596\text{e}{-121}$ for test and reference contents, respectively).

In order to assess the relevance of each coding condition on

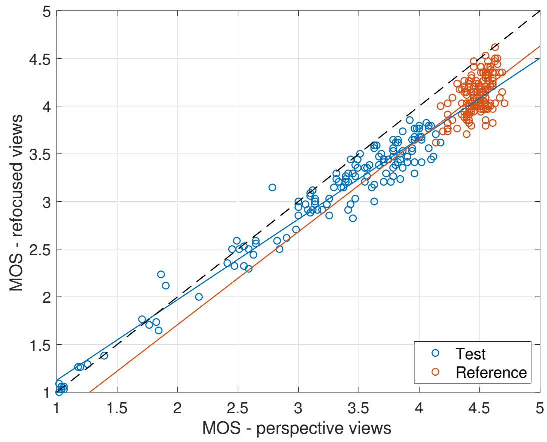


Fig. 3: MOS values for perspective views vs MOS values for refocused views, for test contents (blue) and respective references (orange), with relative linear fitting. The dashed black line represents the $y = x$ function.

TABLE IV: One-way ANOVA on the raw scores given to test contents, divided by type of view.

	SumSq	**DF**	**MeanSq**	**F**	**p-value**
Type of view	146	1	146.046	109.3	1.78763e−25
Error	16352.7	12238	1.336		

TABLE V: One-way ANOVA on the raw scores given to reference contents, divided by type of view.

	SumSq	**DF**	**MeanSq**	**F**	**p-value**
Type of view	373.88	1	373.878	560.71	3.066e−121
Error	8160.21	12238	0.667		

TABLE VI: Multi-way ANOVA (interaction model) on the raw scores given to test contents (perspective views only).

	SumSq	DF	MeanSq	F	p-value
Contents	658.3	5	131.651	179.48	0
Codecs	1935.1	5	387.015	527.63	0
Bitrates	1370.7	3	456.911	622.92	0
Perspective views	3.3	2	1.632	2.23	0.1081
Contents*Codecs	65.7	25	2.627	3.58	0
Contents*Bitrates	45.6	15	3.038	4.14	0
Contents*Persp. views	9.5	10	0.953	1.3	0.224
Codecs*Bitrates	738.4	15	49.225	67.11	0
Codecs*Persp. views	7.3	10	0.729	0.99	0.4459
Bitrates*Persp. views	2.2	6	0.372	0.51	0.8031
Error	5315.6	7247	0.733		

TABLE VII: Multi-way ANOVA (interaction model) on the raw scores given to reference contents (perspective views only).

	SumSq	DF	MeanSq	F	p-value
Contents	50.6	5	10.1204	22.23	0
Codecs	12.12	5	2.424	5.32	0.0001
Bitrates	0.98	3	0.3269	0.72	0.541
Perspective views	3.79	2	1.8956	4.16	0.0156
Contents*Codecs	14.21	25	0.5686	1.25	0.1822
Contents*Bitrates	2.02	15	0.1348	0.3	0.9959
Contents*Persp. views	5.95	10	0.5947	1.31	0.2203
Codecs*Bitrates	2.23	15	0.1486	0.33	0.9929
Codecs*Persp. views	4.65	10	0.4655	1.02	0.421
Bitrates*Persp. views	1.73	6	0.2888	0.63	0.7028
Error	3298.92	7247	0.4552		

the set of scores, we perform multi-way ANOVA on the test and reference scores, considering two-factor interactions. We first consider the scores assigned to perspective and refocused views separately. Tables VI, VII, VIII and IX show the results of the analysis. As seen before, the main effects fall above the 1% significance threshold, thus failing to reject the null hypothesis ($p = 0.1081$ and $p = 0.0156$ for scores assigned to rendered perspective views in test and reference contents, respectively, whereas for reference views the results are $p = 0.2273$ and $p = 0.7686$ for test and reference contents, respectively). Among the first order interactions, it is worth mentioning that the interaction between contents and refocused views is significant for both test and reference contents, meaning that particular combinations of the two influenced how the stimuli were scored.

Finally, we perform multi-way ANOVA on the entire set of scores, considering both test and reference contents simultaneously. Table X summarizes our findings. As can be seen, the scores assigned to different views are to be considered significantly different in a statistical sense. Moreover, the interactions between contents and views, and between codecs and views, are statistically significant. The latter is especially important, because it signals that the choice of rendering parameters can influence how different codecs are assessed, independently of the bitrate (at $p = 0.3915$, the interaction between bitrates and views is not significant). As the same interaction between codecs and views was not deemed statistically significant in the previous analyses, where the views were neatly separated

TABLE VIII: Multi-way ANOVA (interaction model) on the raw scores given to test contents (refocused views only).

	SumSq	DF	MeanSq	F	p-value
Contents	325.74	5	65.149	82.7	0
Codecs	981.81	5	196.361	249.25	0
Bitrates	602.85	3	200.95	255.07	0
Refocused views	1.15	1	1.149	1.46	0.2273
Contents*Codecs	42.09	25	1.684	2.14	0.0008
Contents*Bitrates	30.82	15	2.055	2.61	0.0006
Contents*Refoc. views	22.36	5	4.471	5.68	0
Codecs*Bitrates	393.11	15	26.207	33.27	0
Codecs*Refoc. views	7.06	5	1.411	1.79	0.111
Bitrates*Refoc. views	2.29	3	0.764	0.97	0.406
Error	3791.72	4813	0.788		

TABLE IX: Multi-way ANOVA (interaction model) on the raw scores given to reference contents (refocused views only).

	SumSq	DF	MeanSq	F	p-value
Contents	43.8	5	8.7598	9.28	0
Codecs	101.15	5	20.2309	21.43	0
Bitrates	2.11	3	0.7045	0.75	0.5244
Refocused views	0.08	1	0.0817	0.09	0.7686
Contents*Codecs	14.35	25	0.5739	0.61	0.9363
Contents*Bitrates	5.21	15	0.3472	0.37	0.9867
Contents*Refoc. views	38.01	5	7.6018	8.05	0
Codecs*Bitrates	6.38	15	0.4254	0.45	0.9639
Codecs*Refoc. views	7.1	5	1.4209	1.51	0.1846
Bitrates*Refoc. views	1.16	3	0.3878	0.41	0.7453
Error	4543.63	4813	0.944		

TABLE X: Multi-way ANOVA (interaction model) on the raw scores given to test and reference contents, for all rendered views.

	SumSq	DF	MeanSq	F	p-value
Contents	544.6	5	108.922	89.47	0
Codecs	1151.1	5	230.23	189.12	0
Bitrates	917.8	3	305.921	251.3	0
Views	497.7	4	124.435	102.22	0
Contents*Codecs	71.9	25	2.876	2.36	0.0001
Contents*Bitrates	42.4	15	2.827	2.32	0.0026
Contents*Views	78.4	20	3.922	3.22	0
Codecs*Bitrates	591	15	39.398	32.36	0
Codecs*Views	50.2	20	2.511	2.06	0.0035
Bitrates*Views	15.5	12	1.288	1.06	0.3915
Error	29648.9	24355	1.217		

based on their rendering type (perspective or refocused), the difference in the scores distribution most probably lays in how the two groups were rated. Thus, choosing either one of the two types of rendering could result in significantly different ratings, leading to biased results.

Results show that, within the same rendering group, different parameters do not lead to significantly different scores, both in test and reference contents. Thus, it is unnecessary to test different rendering parameters within the same group, as it increases the complexity of the test without bringing any added value. On the other hand, different types of rendering, such as perspective and refocused rendering, lead to statistically different results in both test and reference contents. In particular, refocused contents were consistently rated lower than their perspective counterpart. This could suggest that selecting only one of the two types of rendering could lead to biases in the way scores are distributed.

One straightforward conclusion from the analysis reported in this section would be to select the rendering parameters as to have only one view per type of rendering. However, using only one rendering parameter per group could lead to unwanted effects. For example, using only one perspective view to assess the quality of the entire light field content could be a feasible solution if the compression artifacts are homogeneously distributed among the views – that is, if the compression algorithm affects different views in equal measure. If that is not the case, selecting which view should be used in the test may become a delicate task. Indeed, a wrong selection of rendering parameters can favor or penalize certain algorithms or solutions. Moreover, compression solutions might be engineered to offer the best quality for the rendering parameters selected for the test, disregarding the quality of others. As such, results obtained by assessing only few rendering parameters might be hard to generalize. Particular care should then be paid in selecting the best rendering parameters as not to favor one algorithm over another, and to ensure a fair comparison among all the evaluated solutions.

VII. Conclusions

In this paper we performed an in-depth analysis of single-image assessment for light field contents using widely-used statistical tools. In particular, we tested whether different types of rendering (in our case, change of perspective and change of focal point) lead to statistically different scores, and if testing a variety of rendering parameters is advisable. We show that, within each type of rendering, no statistical difference can be discerned. Thus, it is sufficient to evaluate only one rendered view from each group, as the scores are statistically equivalent. However, between different types of rendering statistically significant differences can be found. We underline that such differences are present in both test and reference contents, thus they cannot be attributed to the effect of compression artifacts.

We conclude that, although it is theoretically possible to use single-image methodologies to assess light field contents, it is discouraged due to a number of issues associated with it, which might lead to biased results. Increasing the number of rendering parameters is not guaranteed to produce corresponding diversity in the scores; thus, its advantages are definitely outweighed by the increased length and cost it requires. However, caution is suggested in using only a few rendering parameters, since it could be proven ill-advised for certain compression algorithms, and could be susceptible to ad-hoc engineering to achieve the best results at the expenses of the general quality of the content. Further analysis can be devoted to how best select the rendering parameters according to the evaluation scenario, in order to remove any bias towards a particular solution.

References

[1] T. Balogh, "The HoloVizio system," in *Electronic Imaging 2006*, vol. 6055. International Society for Optics and Photonics, 2006.

[2] R. Matsubara, Z. Y. Alpaslan, and H. S. El-Ghoroury, "Light field display simulation for light field quality assessment," in *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 2015, pp. 93 910G–93 910G.

[3] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995, pp. 39–46.

[4] I. Viola, M. Řeřábek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *2016 Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[5] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[6] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[7] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[8] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[9] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[10] G. Alves, F. Pereira, and E. A. B. da Silva, "Light field imaging coding: Performance assessment methodology and standards benchmarking," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–6.

[11] J. N. Filipe, L. M. Tavora, P. A. Assuncaot, R. Fonseca-Pinto, and S. M. de Faria, "Evaluation of focus metrics in extended depth-of-field reconstruction," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[12] P. Paudyal, F. Battisti, M. Sjöström, R. Olsson, and M. Carli, "Towards the perceptual quality evaluation of compressed light field images," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 507–522, 2017.

[13] P. Paudyal, F. Battisti, and M. Carli, "Effect of visualization techniques on subjective quality of light field images," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 196–200.

[14] F. Battisti, M. Carli, and P. L. Callet, "A study on the impact of visualization techniques on light field perception," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 2155–2159.

[15] C. Perra, "Assessing the quality of experience in viewing rendered decompressed light fields," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 21 771–21 790, Aug 2018.

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

[18] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras." IEEE, Jun 2013.

[19] ——, "Linear volumetric focus for light field cameras," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 2, Feb. 2015.

[20] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd a framework for crowd-based quality evaluation," in *2012 Picture Coding Symposium (PCS)*, May 2012, pp. 245–248.

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

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