VALID: Visual quality Assessment for Light field Images Dataset

Irene Viola and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

***Abstract*—In the last years, light field imaging has experienced a surge of popularity among the scientific community for its capability of rendering the 3D world in a more immersive way. In particular, several compression algorithms have been proposed to efficiently reduce the amount of data generated in the acquisition process, and different methodologies have been designed to reliably evaluate the visual quality of compressed contents. In this paper we propose a dataset for visual quality assessment of light field images (VALID). The dataset contains five contents compressed at various bitrates, using both off-the-shelf solutions and state-of-the-art algorithms. Results of objective quality evaluation using popular image metrics are included, as well as annotated subjective scores using three different methodologies and two types of visualization setups. The proposed dataset will help develop new objective metrics to predict visual quality, design new subjective assessment methodologies and compare them to existing ones, as well as produce novel analysis approaches to interpret the results.**

I. Introduction

Light Field (LF) imaging offers new ways of interaction with real-life scenarios in an immersive environment. However, the large volume of data generated in the acquisition process represents a challenge in terms of storage and transmission. The design of new compression solutions relies on subjective and objective visual quality assessment to efficiently reduce the amount of data while preserving both perceptual and immersive features. However, subjective assessment is costly and time consuming. Thus, comprehensive datasets for visual assessment of LF contents under compression artifacts are indispensable.

Several LF image datasets have been proposed in the past, comprised of both synthetic and natural scenes [1]–[3], and for object recognition and saliency map estimation [4], [5]. However, none of the datasets includes objective and subjective quality scores for compression-like artifacts. Paudyal et al. [6] propose a so-called SMART dataset including several LF images compressed at various bitrates, along with the annotated subjective scores. However, the proposed compression solutions only consider intra-based approaches to encode LF images, which were proven to be subpar with respect to pseudo-sequence based approaches [7]. Moreover, the subjective methodology that is used to collect the scores presents LF contents as conventional 2D images, which admittedly disregards any problem that may arise in the encoding of the depth information. Additionally, no data about the participants is provided, and the results are already processed in BT scores with respective confidence intervals, so it is not possible to perform outlier detection or use a subset of the rates.

In this paper we present a new dataset for visual quality assessment of light field images (VALID). The dataset is composed of uncompressed and compressed contents on various bitrates using four compression solutions. Objective quality results based on PSNR and SSIM metrics are provided, along with subjective quality assessment scores obtained using three different methodologies. Two visualization arrangements with different color bit depth are used. A summary of the contents of the dataset can be found in Table I.

II. Dataset description

A. Content and bitrate selection

Five LF lenslet images were chosen from a publicly available LF image dataset, namely I01 = *Bikes*, I02 = *Danger_de_Mort*, I04 = *Stone_Pillars_Outside*, I09 = *Fountain_&_Vincent_2* and I10 = *Friends_1* [3]. The images were carefully selected from those commonly used in literature [7], [9], [10], to provide a variety of scenarios, containing a wide range of details that would be challenging for the compression algorithms in terms of texture and disparity encoding. From each lenslet image, 15×15 perspective views of 625×434 pixels and depth of 10 bits per color channel were obtained, using the Light Field toolbox v0.4 [11], [12]. The central perspective view from the contents is depicted in Figure 1. In order to provide compression distorsions at different levels of visual quality, four bitrates were selected: 0.75 bpp, 0.1 bpp, 0.02 bpp, 0.005 bpp. The values are obtained by dividing the size of the compressed bitstream over the size of the uncompressed raw images (5368×7728 pixels).

B. Encoding solutions and data preparation

A total of five solutions were adopted to compress the LF contents. Two popular video encoders, HEVC and VP9, were selected to encode the perspective views from the LF contents as pseudo-temporal sequences. For HEVC, the software implementation x265[1] was used, with the Main10 profile. For VP9, the official implementation was employed[2]. The Quantization Parameters (QPs) and the target bitrates were selected to match the desired compression ratios for

[1]https://www.videolan.org/developers/x265.html
[2]https://www.webmproject.org/vp9/

(a) I01 (b) I02 (c) I03 (d) I04 (e) I05

Fig. 1: Central perspective view of each content from the proposed VALID dataset.

TABLE I: Summary of contents for the VALID dataset.

Content	Bitrate (bpp)	Objective metrics	Bit depth	Display	Size	Resolution	N_P	N_R	Methodologies	Codecs
I01 I02 I04 I09 I10	0.75 0.1 0.02 0.005	$PSNR_Y$ $PSNR_{YUV}$ $SSIM_Y$ $SSIM_{YUV}$	8 bit	Samsung SyncMaster2443	24in	1920×1200	81	11	Passive Interactive Passive and interactive	HEVC VP9
			10 bit	Eizo ColorEdge CG318-4K	31.1in	4096×2160	97	-	Passive	HEVC VP9 [8] [9] [10]

HEVC and VP9, respectively. To be used for the encoding, the perspective views were padded with black pixels, converted to YCbCr format and downsampled from 444 to 422, 10-bit depth. They were then arranged in a pseudo-temporal arrangement following a serpentine order. Only the central 13×13 perspective views were encoded.

Additionally, three state-of-the-art algorithms were selected from the literature to provide up-to-date results on LF compression. In [8] authors encode a subset of the perspective views using HEVC, adopting a linear approximation prior to estimate the non-encoded views. In [9] authors arrange the perspective views into a multiview structure that can be exploited by the corresponding extension of HEVC, namely MV-HEVC. They also propose a rate allocation scheme to progressively assign the QPs in order to optimize the performance. In [10], a lenslet-based compression solution that uses depth, disparity and sparse prediction information to reconstruct the final set of views is designed. The scheme can be configured to improve the reconstruction by allocating a fraction of the bitrate to the encoding of the lenslet image using JPEG 2000, or to allow random access by encoding a subset of views.

C. Output bit depth

Two output bit depths were considered for the objective and subjective assessments. Initially, 10 bits per color channel (the original bit depth of the images) were used to test the encoding solutions. All codecs were considered for the assessments. Additionally, the output of the encoding algorithms was converted to 8 bits per color channel, to ensure compatibility with the majority of consumers' devices and rendering softwares. Multiple methodologies were assessed to give an overview of different visualization and interaction approaches. For the 8 bit depth case, only HEVC and VP9 were used.

D. Objective metrics

PSNR and SSIM were selected from the literature to provide objective assessments of the visual quality of the contents. The metrics were applied separately to each luminance and chrominance channels Y, U, V and to each perspective view (k, l), where $k = 1, ..., K$, $l = 1, ..., L$ and $K = L = 15$ represent the total number of perspective views, as generated from the toolbox. $PSNR_{YUV}$ and $SSIM_{YUV}$ were computed by means of a weighted average, assigning factor 6 to the luma channel, and factor 1 to each chrominance channel, as defined in [13].

The mean across the viewpoint images was also computed to have the average PSNR values for Y channel:

$$\widehat{PSNR}_Y = \frac{1}{(K-2)(L-2)} \sum_{k=2}^{K-1} \sum_{l=2}^{L-1} PSNR_Y(k,l), \quad (1)$$

Similarly, $\widehat{SSIM}_Y$, $\widehat{PSNR}_{YUV}$ and $\widehat{SSIM}_{YUV}$ were computed.

For the sake of completeness, the objective metrics were calculated on both the 10-bit and 8-bit outputs.

E. Subjective methodologies and test conditions

The subjective evaluations were conducted in a laboratory for subjective quality assessment, which was set up according to ITU-R Recommendation BT.500-13 [14], and equipped with adjustable neon lamps of 6500 K color temperature. The color of the background walls was mid grey, and 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-R Recommendation BT.2022 [15]. Subjects were allowed to move further or closer to the screen. Specification about the display size and resolution can be found in Table I. All monitors were calibrated 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.

Different subjective methodologies were considered based on the output bit depth. For the 10-bit output depth, the

encoding solutions were tested using a "passive" methodology, using $N_P = 97$ perspective views at a rate of 10 frames per second, as recommended in [16]. However, no refocusing was applied on the views ($N_R = 0$), to exclusively compare the outcome of the encoding algorithms. The total length of the animation was 9.7 seconds. A comparison-based adjectival categorical judgement methodology with a 7-point grading scale was selected, according to ITU-R Recommendation BT.500-13 [14]. Each stimulus was displayed alongside the uncompressed reference in a side-by-side arrangement. Participants were asked to compare the quality of the test stimuli with respect to the uncompressed reference and rate it on a scale from -3 (much worse) to +3 (much better), 0 indicating no preference.

For the 8-bit output depth, three methodologies were adopted, to test the impact of different visualization and interaction approaches on the collected subjective scores. Namely, "interactive" and "passive" approaches were implemented to collect the scores, and they were subsequently combined ("passive and interactive" approach) to offer interaction while improving the consistency of the results, as suggested in [16]. In particular, for the "passive and interactive" approach, the participants were shown an animation of the images under test, and could not interact or score before the animation was concluded. To ensure a smooth interaction experience without unwanted distorsions, only the central 9×9 views were used for the tests ($N_P = 81$). Additionally, $N_R = 11$ refocused views were created following [16]. A Double Stimulus Impairment Scale (DSIS) with side-by-side visualization and 5-point grading scale, from 5 (imperceptible) to 1 (very annoying), was selected for all three methodologies. For the "passive" and "passive and interactive" methodologies, the perspective views were shown as an animation, at a rate of 10 frames per second, followed by the refocused views, going from foreground to background and from background to foreground at a rate of 4 frames per second, as suggested in [16]. The total length of the animation was 13.6 seconds. The "interactive" and "passive and interactive" methodologies were implemented using the framework proposed in [17], to allow subjects to engage with the perspective and refocused views.

In all the experiments, the position of the reference was fixed for the duration of the test, and participants were informed of its position on the screen. A training session with four training samples was established before the experiment, composed of one additional content compressed at various bitrates. The order of the stimuli was randomized for each participant, and the same content was never shown twice in a row. All subjects were examined for visual acuity and color vision using Snellen and Isihara charts, respectively. Information about the age and gender of the participants is provided separately for each test. For all the evaluations, subjective scores are provided for each stimulus and for each participant. Additionally, for the "interactive" and "passive and interactive" methodologies, the tracking values from the animation are additionally provided for each subject and for each stimulus, to help analyse user behavior.

III. Conclusion

This paper presents a new dataset for visual assessment of light field images. It provides uncompressed and compressed contents, along with objective and subjective scores. More information can be found in: https://mmspg.epfl.ch/VALID.

Acknowledgment

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).

References

[1] Synthetic light field archive. [Online]. Available: http://web.media.mit.edu/~gordonw/SyntheticLightFields/

[2] The (new) stanford light field archive. [Online]. Available: http://lightfield.stanford.edu/

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

[4] A. Ghasemi, N. Afonso, and M. Vetterli, "Lcav-31: a dataset for light field object recognition," in *Computational Imaging XII*, vol. 9020. International Society for Optics and Photonics, 2014, p. 902014.

[5] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2806–2813.

[6] 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.

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

[8] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

[9] W. Ahmad, R. Olsson, and M. Sjöström, "Interpreting plenoptic images as multi-view sequences for improved compression," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

[10] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and JPEG 2000," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

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

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

[13] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

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

[15] 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.

[16] I. Viola, M. Rerabek, and T. Ebrahimi, "Impact of interactivity on the assessment of quality of experience for light field content," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.

[17] I. Viola and T. Ebrahimi, "A new framework for interactive quality assessment with application to light field coding," in *Applications of Digital Image Processing XL*, vol. 10396. International Society for Optics and Photonics, 2017, p. 103961F.