EFFECT OF VISUALIZATION TECHNIQUES ON SUBJECTIVE QUALITY OF LIGHT FIELD IMAGES

Pradip Paudyal, Federica Battisti, and Marco Carli

Department of Engineering, Università degli Studi Roma TRE, Rome, Italy.

ABSTRACT

Light Field imaging derives from the fundamentals of light field sampling, where the spatial information about a scene can be captured with angular information. That is, the Light Field imaging is based on a camera recording information about the intensity of light from the scene and about the direction of the light rays. The acquired data can be shown to the user in different ways, such as image with digitally extended depth of field, 3D, parallax, and 360 degree display. In this contribution, the impact of different rendering techniques on the Quality of Experience is addressed. The achieved results show that the visualization techniques may have different impact on the perceived quality even when the same content is considered.

***Index Terms*—** Light Field Imaging, Quality of Experience, Visualization, Quality Metrics, Subjective Experiment

1. INTRODUCTION

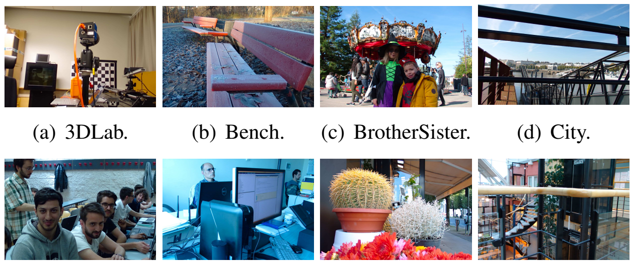
In the last decades, several efforts have been devoted for developing systems and processing algorithms to capture, distribute, display, and store images and videos. As results, many technologies like 3D television (3DTV), Free-Viewpoint Video, High Dynamic Range imaging, and Light Field (LF) imaging, are becoming popular. The latest development in the imaging field is led by LF camera. Basically, multiple views of a scene are recorded by using a camera with a microlens array inserted between the camera sensor and the main lens [1, 2]. The camera can be considered as a relay system, where the main lens creates a main image in the air, then this main image is re-mapped to the sensor by a microlens array, and thus it is able to provide multiple views of the image in a single shot [3].

This new technology can be applied in a wide range of appealing applications: interactive pictures (where focus and depth of field can be adjusted after the picture is taken), free view point display, augmented and virtual reality, parallax, 3D, and light field display [4]. LF images are prone to a wide variety of distortions during acquisition, pre-processing, compression, transmission, and post-processing; each of these steps may result in visual quality degradation. The LF imaging demands high computational power and presents image resolution and quality issues, among them: i) pre-processing steps, such as raw sensor data decoding that distorts the visual information of the scene. For this reason, demosaicing, devignetting, and color and gamma correction procedures are needed during the raw sensor data decoding [5]; ii) high compression rate systems to reduce the high redundancy typical of LF images; iii) rendering systems. LF images can be rendered in different ways. Based on the target display method and device, the LF image needs to go through specific rendering operations. The artifacts introduced by rendering depend on the display device and the adopted rendering methods, and ultimately, they may influence the perceived quality of the LF image. Therefore, the knowledge of the quality level of the processed LF image is important to design and optimize the processing algorithms and systems.

Another motivation for this work is that this technology is not commonly spread and it is unknown even to targeted consumers. Therefore, the subjective quality assessment of such a content is challenging due to the rapid development of this technology. In fact, the quality estimation of test images is already an ill-posed problem, and it becomes more challenging for LF images, due the wide rage of applications and visualization possibilities.

In literature, the subjective quality of watermarked LF images is addressed in [6]. In [7], the performances of objective quality metrics, PSNR and SSIM, are evaluated when applied to compressed LF images by using a subjective experiment [8]. A signal distortion estimation method for the LF image is presented in [9]. However, many issues are still to be solved and, up to now, no standard or validated objective and subjective quality assessment methods for LF image are available. It is useful to cite that an ongoing effort to this aim is being devoted by the JPEG PLENO community, which is targeting a standard framework for the representation and exchange of new imaging modalities such as, among others, the LF data [10].

This article is devoted to address the following three research questions: how to evaluate the subjective quality of LF images? Does the subjective quality of LF image vary based on the adopted visualization technique? What can be the possible ways of visualization for the perceptual quality evaluation of distorted LF images? In more details, from a statistical point of view, a null hypothesis, H_0: subjective quality



(a) 3DLab. (b) Bench. (c) BrotherSister. (d) City.

(e) ClassRoom. (f) COMLAB. (g) Flower. (h) Ladder.

Fig. 1. Thumbnails of the selected SRCs.

of light field image is significantly influenced based on the adopted visualization technique, has been tested.

2. SUBJECTIVE EXPERIMENT DESIGN

To address the research questions presented in the introduction, a subjective experiment has been performed by using different visualization techniques. In this section, the followed procedures are briefly reported.

2.1. Raw LF data processing

LF images are captured by using the Lytro Illum camera [11]. The camera recorded raw LF sensor data is decoded by using the Light Field Toolbox v0.4 [12]. During the decoding, demosaicing, devignetting, and colour and gamma corrections are applied [5]. As a result, the sensor data are converted into a LF data structure: a stack of 2D low-resolution RGB images in addition to the weighting image. The latter image conveys the confidence associated with each pixel, which can be used in filtering applications. The resulting dimension of the LF image is $15 \times 15 \times 434 \times 625 \times 4$, where 15×15 represents the number of views, 434×625 represents the resolution of each view and 4 corresponds to the RGB and weighting image components. In this work, the weighting component is discarded, and only the RGB color channels are considered.

2.2. Test material creation

2.2.1. Reference sequence

In this work, taking in to account the ISO 20462: psychophysical experimental methods for estimating image quality [13], eight Source Seqences (SRCs) have been selected. SRCs thumbnails are shown in Figure 1. During SRCs selection, image quality attributes, such as Colorfulness (CF), spatial perceptual information (SI), texture, and Human Visual System (HVS) properties, together with the LF imaging capabilities, depth dependence and Lambertian lighting, were considered [14]. The analysis performed on this set for highlighting image quality attributes shows that the selected sequences cover a wide range of key image content related attributes: CF and SI (as reported in Figure 2).

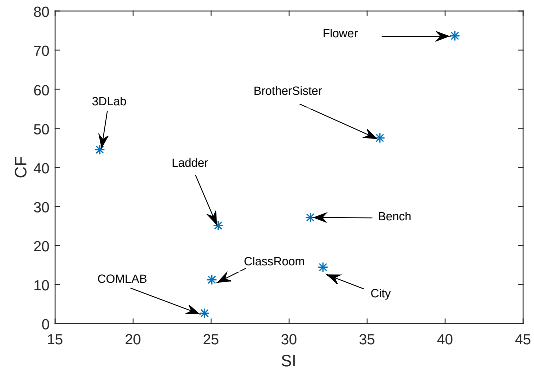


Fig. 2. SRCs in Spatial perceptual Information (SI) and Colorfulness (CF) plane.

2.2.2. Hypothetical Reference Circuits

To create the LF image test set, compression methods and Gaussian noise have been considered as Hypothetical Reference Circuits (HRCs). The encoder can be classified in two groups based on the input: lenslet-based and sub-aperture encoder. In the first one, the LF content is arranged in a lenslet structure: micro images are considered as pixels of the 2D image. This arrangement is suitable for the traditional image compression methods and for the recently proposed LF image encoders exploiting modified HEVC intra profile [7, 15]. In the latter class, the sub-aperture views of a scene are considered as video frames, and the LF image is encoded as a pseudo-video [16]. Here, the LF micro images are arranged in a lenslet structure and encoded by using JPEG, JPEG2000, and HEVC-intra profile. Four levels of distortion are considered. As in [17], the distortion level is selected based on target PSNR values: 21, 24, 27, and 30 dB; with this aim, the selected distortion levels should span a wide range of quality, from bad to excellent [18]. As a result, to create the Processed Image Sequences (PISs), compression ratios of 10 to 350 are considered for JPEG2000, quality levels from 5 to 100 are considered for JPEG, and variance of Gaussian noise from 0.001 to 0.01 are considered. Finally, for HEVC intra the selected Quantization Parameters (QPs) range from 22 to 51.

2.3. Selected visualization techniques

As mentioned in Section 1, the LF content can be visualized in many ways. In this work, based on the publicly available 2D display devices, only four visualization techniques are selected:

- All-in-focused-view: the central sub-aperture view of LF content is selected as an all-in-focused view. This view is considered as the Extended Depth of Field (EDoF) image. In fact, the EDoF image uses the light coming from a larger lens aperture, i.e. it captures light more efficiently and allows less grainy images with higher SNR [19]. The issue *(I1)* associated with this technique is that evaluating the LF content (comprising of 15×15 views) by using only a central view may not be sufficient to describe the quality of the LF image.

- Pseudo-video (PV) with circular viewing trajectory: to address issue *(I1)*, this technique considers sub-aperture views as frames for creating a pseudo-video. The frames are displayed with the circular viewing trajectory. The associated issue *(I2)* is that, due to the lack of natural motion information, this type of visualization technique is not usual, and may lead to visual discomfort in the subjects. Moreover, the transition within the video is not smooth.

- Refocused images: one appealing application of LF imaging is digitally refocused images. Therefore, in this work, subjective quality of the test LF image is assessed with the help of randomly selected two refocused images, which are created by focusing at different planes. In the following, the refocused images are indicated as Focused at plane *P1* ($FP1$) and Focused at plane *P2* ($FP2$). To create the refocused images, the Matlab function *LFFiltShiftSum* has been used. Specifically, the subaperture images have been shifted based on a parameter, referred as a slope, and are added together to create the refocused image. The point of focus is determined by the selected slope value. The issue *(I3)* in this case is that considering only two refocused images may not be sufficient to address the quality of a LF image. In fact, many refocused images can be obtained by focusing at different slopes.

- Pseudo-video with focusing at different slope: to address issue *(I3)*, refocused images are created by changing the slope from -2 to +2, and the Refocused images are shown to the subjects as a pseudo-Video (RV) [12]. In this work the length of the created pseudo-videos is set to 5s.

2.4. Subjective experiment setup

In literature, many protocols have been proposed for multimedia quality assessment [20]. Based on the assessment length, complexity, provided accuracy, and stability, the Absolute Category Rating (ACR) method is generally considered for video quality assessment [21]. In this work, our goal is to evaluate the overall visual Quality of Experience (QoE) of LF images. The selected test materials (images and pseudo-videos) provide a wide range of quality. Therefore, the ACR method [20] has been selected as quality assessment protocol. In the subjective experiment, 17 test images for each SRC, i.e. $17 \times 8 = 136$ LF images, are evaluated by the subjects. In more details, each SRC ic composed by the reference image and 16 distorted images (4 images for each artifacts: JPEG, JPEG2000, HEVC intra encoding, and Gaussian noise).

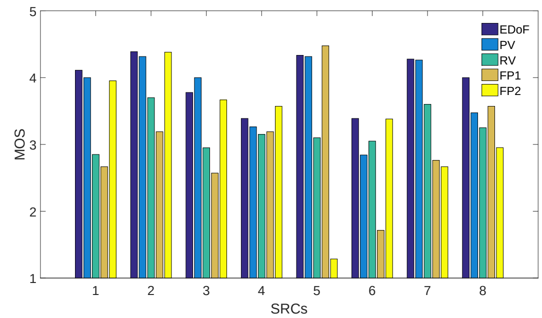


Fig. 3. MOS for the SRCs.

The experiment is divided in two sessions to maintain the total length of the experiment less than 30 minutes for retaining the attention of the subjects. Each session is further divided in two sub-session, and the same visualization technique is adopted in each sub-session. A sufficient number of subjects (19) has been used for each sub-session [22].

The subjective experiment was conducted by using the Graphical User Interfacer (GUI) presented in [23]. The images and videos were displayed in a computer monitor (DELL U2413 digital monitor, with resolution of 1920x1200 pixels), and background of the desktop set to gray. Verbal instructions followed by a written instructions (prepared by using the format suggested in ITU-T Rec. P.910 [20]) were given to the subjects. Before the subjective experiment, training sessions were scheduled for each visualization techniques to make the subjects familiar with the LF content and visualization techniques. This training session also helps the subject to map the annoyance quality range.

3. RESULTS AND DISCUSSION

Before processing the results, to remove the opinion scores given by the subjects which are strongly deviated from the mean behavior, the outlier detection method presented in ITU-T Rec. BT.500 [22] has been used. As a result, one subject for extended depth of field (EDoF) images and eight subjects for parallax display were detected as outliers. Whereas, scores given by all the subjects were accepted for refocused views and for refocusing video. After the outlier detection, for each test LF image, Mean Opinion Scores (MOS) and associated Confidence Interval (CI), which is derived from the standard deviation and size of each sample, are estimated, as presented in [22].

Subjective quality of SRCs, estimated by using the different visualization techniques, is shown in Figure 3. It shows that the MOS scores are comparable and uniformly high for EDoF and PV. However, the MOS scores are varying noticeably for the other methods. To confirm the results, a reliability analysis is performed. The reliability of the opinion scores is measured with the help of individual agreement with the mean behavior. The agreement between the scores given by an individual subject to the MOS is shown in Table 1. It shows that the opinion scores collected using EDoF images and par-

Coeff.	EDoF	PV	FP1	FP2	RV
PLCC	0.808	0.853	0.7268	0.7674	0.587
SRCC	0.818	0.851	0.7097	0.7681	0.539

Table 1. Average individual agreement with MOS for different visualization techniques.

	HRCs (p_{value})	SRCs (p_{value})
EDoF	$P(F > 30.84) = 0$	$P(F > 1.75) = 0.102$
PV	$P(F > 44.01) = 0$	$P(F > 1.3) = 0.253$
RV	$P(F > 11.3) = 0$	$P(F > 5.52) = 0$
FP1	$P(F > 3.03) = 0$	$P(F > 26.83) = 0$
FP2	$P(F > 2.6) = 0.0017$	$P(F > 30.08) = 0$

Table 2. Result of ANOVA test.

allax video have very high individual agreement, which is measured in terms of Pearson Linear Correlation Coefficient (PLCC) and Spearman's rank correlation coefficient (SRCC). It is important to notice that only the reliable opinion scores (highest average individual agreement) can be used to benchmark the processing algorithms [22]. Based on these results, we can not reject the $H0$.

Moreover, the differences in the opinion scores given to the test LF images, by using different visualization methods, are analysed with the help of Analysis of Variance (ANOVA). The achieved results are summarized in Table 2. A small p_{value} (< 0.05), indicates the difference between column means is significant. The F-statistic is the ratio between the mean squared errors: between groups variations and within group variation. It shows that, in EDoF image visualization, the MOS scores are significantly different for the HRCs and the subjective quality is less affected by the image content. The same result is achieved for PV visualization technique. However, for refocused images and video, the subjective quality is mainly influenced by the image content other than by the adapted artifacts (HRCs).

Finally, correlation analysis is performed on the collected opinion scores using different visualization techniques. The achieved results, shown in Figure 4, indicate that: i) the correlation coefficient between the scores collected through EDoF image and parallax video is very high, ii) there is a remarkable correlation between the scores collected using refocused video and parallax video, and iii) there is low level of correlation between the scores collected using refocused views. This is due to the fact that, in refocused images out-of-focus regions have been blurred, and this has a dominant impact on the subjective quality; the used refocusing method (shift-sum) reduces the level of introduced artifacts by adding all the sub-aperture views in a refocused view; the refocusing method introduced blur artifact that has a very strong impact more than Gaussian blur, JPEG2000 produced ringing and blur, and low level of JPEG and HEVC intra introduced block artifacts.

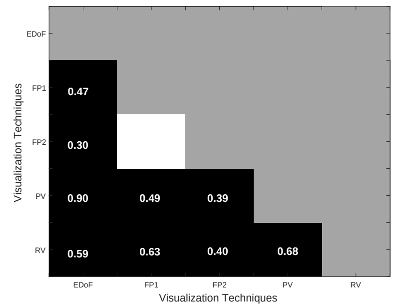


Fig. 4. Correlation analysis results between the visualization methods. The black box represents the correlation on the opinion scores is significant, whereas white box indicate that there is no significant correlation. The value given in the blue box is the correlation coefficient.

Metrics	ρ	EDoF	PV	RV	FP2	FP1
PSNR	PLCC	0.792	0.9	0.572	0.429	0.384
	SRCC	0.782	0.906	0.516	0.43	0.398
SSIM	PLCC	0.792	0.9	0.573	0.429	0.384
	SRCC	0.867	0.86	0.457	0.365	0.331

Table 3. Performance of objective 2D image quality metrics, when applied for LF image, measured in terms of correlation coefficients (ρ).

Furthermore, the collected opinion scores are also used to benchmark the well known the 2D image quality metrics: PSNR and Structural Similarity (SSIM) index [24]. The achieved results shown in Table 3 indicate that the opinion scores collected by using EDoF images and PV have significantly high correlation coefficients (PLCC and SRCC) compared to refocused images. Moreover, the correlation is higher for pseudo-video in comparison to the all-in-focused image.

4. CONCLUSION

This article presents the impact of visualization techniques: all-in-focused image, pseudo-video, refocused views, and pseudo-video using refocused views, on subjective quality of LF image. The achieved results show that there is high correlation between the scores collected using all-in-focused image and pseudo-video with circular trajectory. A similar result is achieved by analyzing the individual agreement versus the average behavior. Therefore, the pseudo-video has been noticed as an appropriate solution for the subjective quality assessment of LF image. Next, the performance of validated 2D image quality metrics, PSNR and SSIM, when applied to LF image, is evaluated. The results indicate that there is a significant correlation between the estimated quality scores and the subjective scores collected by using all-in-focused image and pseudo video.

5. REFERENCES

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[2] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Int. Conf. on Computational Photography (ICCP)*. IEEE, 2009, pp. 1–8.

[3] T. Georgiev and A. Lumsdaine, "Depth of field in plenoptic cameras.," in *Eurographics (Short Papers)*, 2009, pp. 5–8.

[4] P. A. Kara, M. G. Martini, P. T. Kovacs, S. Imre, A. Barsi, K. Lackner, and T. Balogh, "Perceived quality of angular resolution for light field displays and the validy of subjective assessment," in *Int. Conf. on 3D Imaging (IC3D)*. IEEE, 2016, pp. 1–7.

[5] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE conf. on computer vision and pattern recognition*. IEEE, 2013, pp. 1027–1034.

[6] P. Paudyal, F. Battisti, A. Neri, and M. Carli, "A study of the impact of light fields watermarking on the perceived quality of the refocused data," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE, 2015, pp. 1–4.

[7] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *32nd Picture Coding Symposium*, 2016, number EPFL-CONF-221601.

[8] I. Viola, M. Řeřábek, and T. Ebrahimi, "A new approach to subjectively assess quality of plenoptic content," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2016, pp. 99710X–99710X.

[9] H. Shidanshidi, F. Safaei, and W. Li, "Estimation of signal distortion using effective sampling density for light field-based free viewpoint video," *IEEE Trans. on Multimedia*, vol. 17, no. 10, pp. 1677–1693, 2015.

[10] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "Jpeg pleno: Toward an efficient representation of visual reality," *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, Oct 2016.

[11] Lytro Illum Camera, "Online: `https://illum.lytro.com/illum`," 2017.

[12] D. Dansereau, "Light field toolbox v0.4," 2016.

[13] B. W. Keelan and H. Urabe, "Iso 20462: a psychophysical image quality measurement standard," in *Electronic Imaging*. International Society for Optics and Photonics, 2003, pp. 181–189.

[14] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, and M. Carli, "Smart: a light field image quality dataset," in *7th Int. Conf. on Multimedia Systems*. ACM, 2016, p. 49.

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

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

[17] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[18] H. R. Sheikh, M. F. Sabir, and A. C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[19] R. Ng, *Digital light field photography*, Ph.D. thesis, Stanford University, 2006.

[20] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," 2008.

[21] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," in *Second Int. workshop on Quality of multimedia experience (QoMEX)*. IEEE, 2010, pp. 82–87.

[22] ITU-R Rec. BT.500, "Methodology for the subjective assessment of the quality of television pictures," 2012.

[23] M. Nuutinen, T. Virtanen, O. Rummukainen, and J. Häkkinen, "Vqone matlab toolbox: A graphical experiment builder for image and video quality evaluations," *Behavior research methods*, vol. 48, no. 1, pp. 138–150, 2016.

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