# A study on the impact of visualization techniques on Light Field perception

Federica Battisti[1], Marco Carli[1], and Patrick Le Callet[2]

*Abstract*— **Light Field imaging is a promising technology that allows to capture the whole set of light rays in a scene thus enabling the generation of perspective views from any position. This possibility can be exploited in several application scenarios, such as virtual and augmented reality or depth estimation. In this framework many issues arise due different aspects such as the large amount of generated data or to the need of dedicated and expensive hardware for Light Field capturing. Moreover, the Light Field carries information about the entire scene and the data that is delivered to the users largely differs from the traditional 2D and 3D media in terms of content and way of fruition. Dedicated rendering technology and devices for the Light Field are nowadays still not mature or quite expensive and the best option is to render the Light Field data on a conventional 2D screen. Consequently, there is the need for finding the best visualization technique that allows to exploit the information in the Light Field while being accepted by the viewers. In this paper we address this issue by considering six visualization options and by running experimental tests to study which is the technique preferred by the users.**

## I. INTRODUCTION

The Light Field (LF) expresses the radiance as a function of position and direction in regions of free space [1]. In other words, it represents the number of light rays within a specific area. The capturing of all light rays in a scene allows to generate a perspective view from any position. Therefore, LF technology can be effectively used in many applications: from accurate passive depth estimation to change of viewpoint or view synthesis, that can be useful in augmented reality content capture or movie post production.

The capturing of a LF is a quite complex procedure from the technological point of view; in fact the light field represents rays with varying positions and angles, and, in order to obtain these information, it is necessary to record the scene from multiple positions. To this aim, different techniques can be adopted: the use of camera arrays, camera gantry, or plenoptic cameras [2].

By spatially locating multiple cameras into an array, the entire LF may be collected at once. This approach is used in [3], in which a planar array of 128 cameras is exploited.

A different system is based on moving a single camera while capturing a stationary scene in order to measure the incident light rays [4]. The basic idea behind the plenoptic imaging systems is the use of a micro-lens array positioned on the focal point of the camera lens, in front of the imaging sensor as shown in Figure 1 [5], [6]. This system allows to record multiple views of a scene in a single shot, thus reducing issues related to calibration and camera synchronization. The micro-lens array records the information on the incident light direction at different positions, i.e. it records the LF. The availability of low cost acquisition devices [7], [8] allows novel applications for these imaging systems. The exploitation of the LF redundancy in the post processing and editing phases brings photographers and art directors new opportunities.

One of the main issues of this technology is related to the rendering modality. Many efforts are being devoted to the design of dedicated displays (e.g., an array of video projectors aimed at a lenticular sheet, 3D Displays [9], up to recently proposed tensor displays [10]) or devices (e.g., head mounted systems for virtual reality applications). However, up to now, these systems are very expensive and there are many challenges to be addressed (e.g., the reduced angular resolution of a LF cinema [11], [12].

The simplest and cheapest solution is the rendering of the LF data on conventional 2D screens. Since the LF gives the opportunity to render the scene from several points of view and focus points, the questions of *what* and *how* to render the scene on a 2D display arise.

In order to answer these questions, in this work an in-deep analysis of the impact of different visualization techniques of LF images on a 2D display is presented. The outcomes of this study are useful not only for better understanding the impact of the different visualization techniques on the users' satisfaction, but also for the establishment of protocols for performing subjective experiments for evaluating the LF Quality of Experience (QoE).

The rest of the paper is organized as follows. In Section II the related works and the motivations for this paper are presented, in Section III a detailed presentation of the performed tests together with an analysis of the obtained results is reported, and finally in Section IV, the conclusions are drawn.

## II. RELATED WORKS

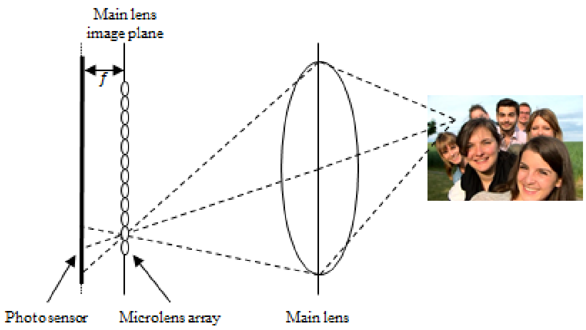
As mentioned in the Introduction, the development of techniques for LF processing is rapidly increasing and, consequently, an assessment of the quality of the produced output is needed. In this section, a review of the LF capturing and rendering systems is reported and the related works are presented.

As already stated, a plenoptic camera is built using a lenticular array placed in front of the sensor. In particular, the

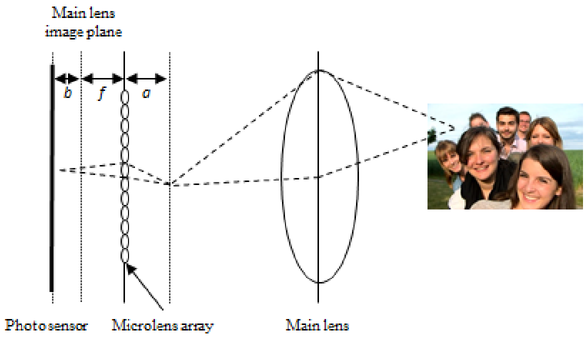
[1]F. Battisti and M. Carli are with the Department of Engineering, Roma Tre University, Rome, Italy `federica.battisti, marco.carli at uniroma3.it`

[2] P. Le Callet is with the Ecole Polytechnique de l'Université de Nantes, IRCCyN, Nantes, France `patrick.lecallet at univ-nantes.fr`

ISBN 978-90-827970-1-5 © EURASIP 2018

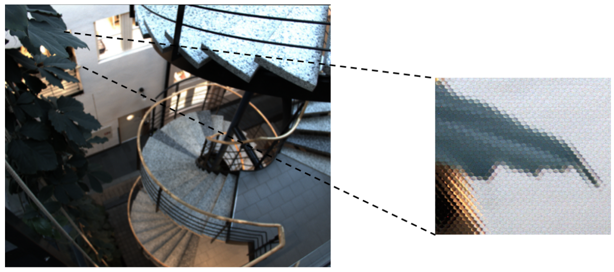



(a)

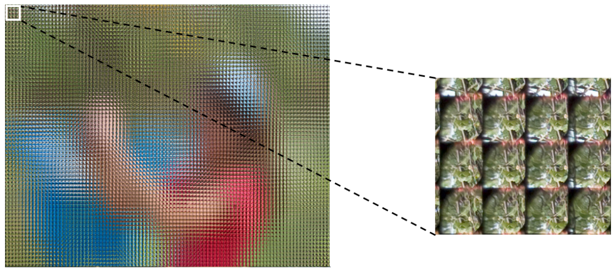



(b)

Fig. 1. Standard (a) and focused (b) plenoptic camera.



(a)



(b)

Fig. 2. Lenslet LF images: (a) standard plenoptic camera image and (b) focused plenoptic camera image.

focal plane of the micro-lens is on the camera image sensor plane, and the camera only captures angular information in each micro-lens image for a single point in the 3D space thus resulting in a low spatial resolution of the final rendered images. To overcome this drawback (a trade off the spatial resolution with angular resolution), the focused plenoptic camera is proposed in [13]. As shown in Figure 1, focused plenoptic cameras capture both angular and spatial information in each micro-lens image by setting the focal plane of micro-lenses far from the image sensor plane. The recorded information can be represented as micro-lens image and sub-apertures (multi-views) as shown in Figure 2.

The simplest way of rendering the recorded light field is to use the raw image of pixel values read by the photosensor under the micro-lens array. When displayed on a 2D screen, the raw image appears like a 2D image. When zooming it or looking closely, it is possible to notice that the raw image is composed by an array of disks. Each disk is the image recorded underneath each micro-lens and its circular shape is due to the round aperture of the lens.

In literature, the evaluation of the systems adopted for rendering LF data is performed mainly for the design or optimization of dedicated devices. For example, in [14] an analysis of typical camera setups and LF display setups is performed for providing an optimization method for virtual camera setups. In [15], the impact of a LF display in a setting supporting collaborative use in a medical scenario is presented. In [16], the evaluation of the impact of different compression systems on LF images is performed. Compressed refocused views were presented to the subjects for quality assessment. Lossy compressed LF images are also considered in [17]. In [18] a multiview autostereoscopic display has been used for testing the performances of a quality metric for dense light fields. In [19], a methodology for the subjective evaluation of light field images is presented. In [20] different light field compression algorithms are compared by means of a set of objective and subjective quality assessments. In particular the subjective assessment is performed by exploiting the Double Stimulus Impairment Scale showing the observers the compressed and un-compressed central viewpoint image. The interactivity feeling is given by the possibility of accessing and displaying the other viewpoints from the 4D LF data by dragging the mouse.

In a previous work [21], based on the available 2D display devices, different visualization techniques were analyzed. In more details: i) *all-in-focused-view* (i.e., the central sub-aperture view of LF content is shown to the user), ii) *pseudo-video* (i.e., sub-aperture views are considered as frames for creating a pseudo-video that is displayed with circular viewing trajectory), iii) *refocused images* (i.e., displayed images are created by focusing at different planes), and iv) pseudo-video with focusing at different planes (i.e., the refocused images are created by continuously changing the focus plane). From the results collected by the performed subjective experiments, it was possible to conclude that, besides the classical 2D static image visualization, the subjects preferred the *pseudo-video* visualization technique.
Anyways, by changing the trajectory in the selection of the views used to create the pseudo-video, it is possible to create many versions of the same content. In this work, we deal with the selection of the view collection modality that allows at

ISBN 978-90-827970-1-5 © EURASIP 2018 

the same time to exploit the information of the LF and to provide the user with the feeling of naturalness.

In the following, the details of the subjective experiment performed for further understanding users' preferences are reported and the collected results discussed.

## III. Evaluation Methodology

In order to understand the best visualization technique to be adopted for light field content fruition, a subjective test has been designed and performed. To this aim, starting from the available views in the light field, we created pseudo-videos obtained by ordering the views according to different trajectories.

### A. Dataset formation

A set of 12 LF images (see Figure 3), extracted from the EPFL *Light-Field Image Dataset* [22] recorded with a Lytro Illum camera, has been used. Among all images in the dataset, we selected: *a)* one image for each available category (*Buildings, Grids, Studio, Urban, Light, Mirror and transparency, Nature*, and *People*), *b)* a second image for the categories *Grids, Mirror and transparency, Nature*, that account for LF specific capabilities due to the presence of reflections and foreground occlusions [23], and *c)* a second image for the category *Urban*, to be used for training purposes.

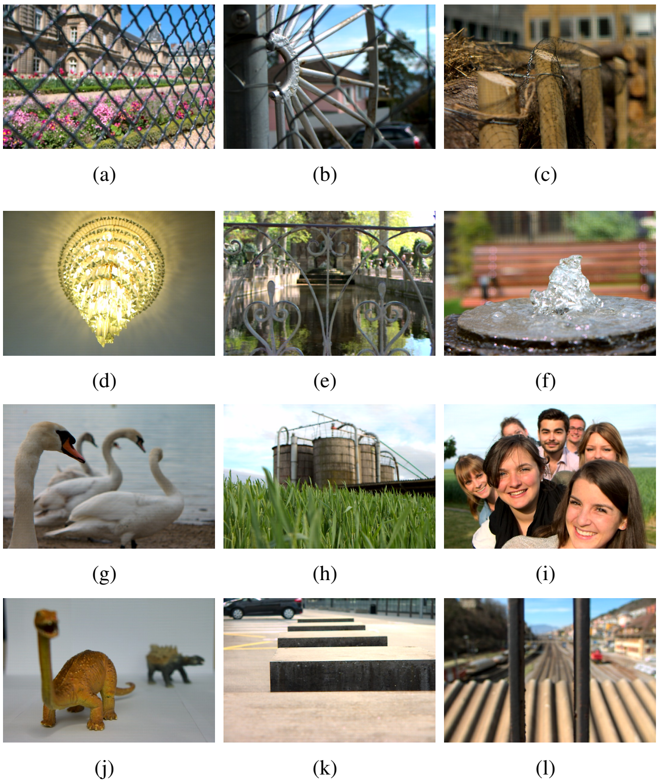



Fig. 3. Test images taken from the EPFL Light-Field Image Dataset from the different categories: (a) *Buildings*, (b)-(c) *Grid*, (d) *Light*, (e)-(f) *Mirror and transparency*, (g)-(h) *Nature*, (i) *People*, (j) *Studio*, (k)-(l) *Urban*.

For all considered images, the Spatial perceptual information (SI) and the Colorfulness (CF) values have been computed and the obtained scores are plotted in Figure 4. In more details, SI allows to characterize the scene content. In this work, the ITU recommended method [24] has been adopted. According to this approach, the luminance $Y$ of the image is first filtered by using a *Sobel* filter. Then, the standard deviation of the pixels in each filtered image is computed as SI:

$$SI = \sigma_{space}[Y_{Sobel}], \tag{1}$$

where $\sigma_{space}$ is the standard deviation over the pixels, and $Y_{Sobel}$ is the *Sobel* filtered luminance plane of the image.

CF allows to characterize the video perceptual quality and the naturalness of the signal by analyzing the variety and intensity of colors in the image. The CF is computed following the approach presented in [25]:

$$CF = \sigma_{rgyb} + 0.3\mu_{rgyb}, \tag{2}$$

where $\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}$, $\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2}$, $rg = R - G$, and $yb = 0.5(R+G) - B$, $\sigma$ is the standard deviation and $\mu$ is the mean value. The R, G, and B are the red, green, and blue color components of the image.

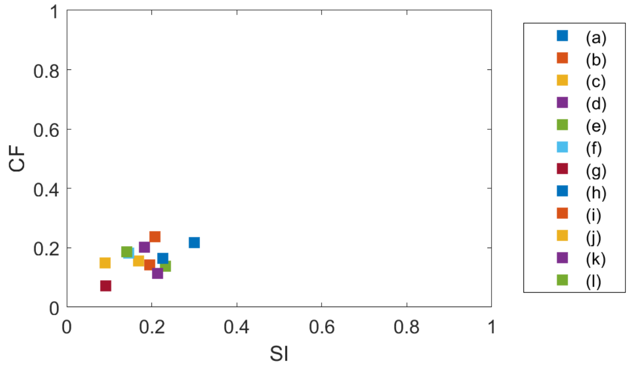


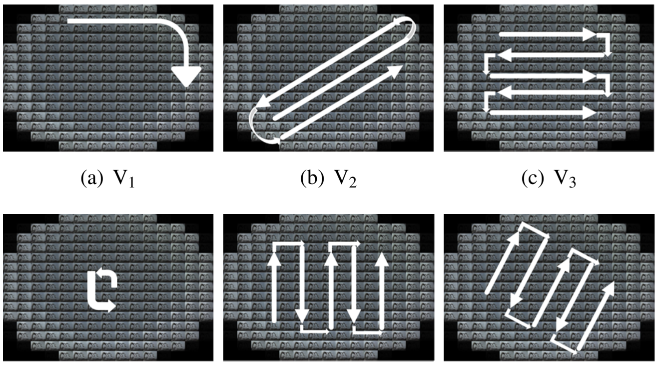

Fig. 4. Spatial Information (SI) and Colorfulness (CF) of the images in the dataset.

Among the possible options for creating the pseudo-video we selected the ones depicted in Figure 5 and described in the following:

- $V_1$: spiral scan from the external to the internal views in clockwise direction;
- $V_2$: diagonal scan in a spiral fashion starting from the view on the left inferior corner;
- $V_3$: horizontal scan from left to right starting from the view on the left superior corner;
- $V_4$: spiral scan in counter-clockwise direction starting from the center view;
- $V_5$: vertical scan from bottom to top starting from the view on the left inferior corner;
- $V_6$: diagonal scan from left to right.

Moreover, each created video was played with three values of frames per second: 10, 15, and 20 fps.

ISBN 978-90-827970-1-5 © EURASIP 2018 



(a) $V_1$ (b) $V_2$ (c) $V_3$

(d) $V_4$ (e) $V_5$ (f) $V_6$

Fig. 5. Analyzed visualization techniques.

### B. Subjective experiment

Overall, 28 subjects, drawn from a pool of students from the Department of Engineering and with age in the range [23-27], participated to the experiments. The tests were run in a controlled environment and each session lasted less than 30 minutes as recommended in [26], [27].

Each subject was asked to express his/her appreciation of the visualized content on a scale from 1 to 5 corresponding to *Bad* to *Excellent* in an Absolute Category Rating scheme [24]. Before starting the test, a training phase was carried out to allow the subjects to familiarize with the content and the rating scheme. As already mentioned, one of the two images belonging to the category *Urban* (Figure 3(l)) was used for this purpose.

### C. Data analysis

The collected scores are first screened for outliers detection according to the procedure described in [27]. From this analysis no outliers are found. Then, the mean value and the %95 Confidence Interval are computed for each visualization technique as follows:

$$\bar{u}_v = \frac{1}{n}\sum_{i=1}^{n} u_{iv}$$

where $u_{iv}$ is the score of observer $i$ for a given visualization $v$, and $n$ is the number of subjects.
The 95% Confidence Interval is calculated as:

$$[\bar{u}_{iv} - \delta_{iv}, \bar{u}_{iv} + \delta_{iv}]$$

where:

$$\delta_q = t_{0.05}\frac{S_q}{\sqrt{n}}$$

with $t_{0.05}$ being the $t$ value for a significance level of 95% that is equal to 1.96. The standard deviation for each visualization, $S_v$, is given by:

$$S_v = \sqrt{\sum_{i=1}^{n}\frac{\left(\bar{u}_{iv} - u_{iv}\right)^2}{(n-1)}}.$$

The obtained values are reported in Table I and plotted in Figure 6.

### D. Discussion

From the analysis of the achieved results it can be noticed that:

- the frame rate does not have a noticeable impact on the subjective scores; in fact, the largest difference in the mean values of the collected MOS is 0.4 for $V_6$ and it can be negligible as shown by the Confidence Intervals;
- the mean values are always smaller than 3 thus meaning that, on average, the subjects evaluated the quality of the video sequences as less than *Fair*. A possible explanation for this behavior may be due on the selected video content. In fact, as shown in Figure4, the images in the dataset present low scores of SI and CF that account for scene content and video naturalness;
- the preferred visualization is $V_3$ in which the views are selected by starting from the one positioned in the top-left corner and proceeding in a horizontal scan; in this way the video simulates an horizontal shift of the recording camera and the transitions between consecutive frames results to be smooth;
- the less favorite visualization techniques is $V_2$. This can be easily explained by the fact that consecutive views of the displayed video contain both an horizontal and a vertical shift thus resulting in a lack of smoothness and naturalness.

| Frame rate | Visualization | Mean | 95% Confidence Interval |
|---|---|---|---|
| **10 fps** | $V_1$ | 2.25 | 0.19 |
| | $V_2$ | 1.70 | 0.14 |
| | $V_3$ | 2.80 | 0.20 |
| | $V_4$ | 2.25 | 0.18 |
| | $V_5$ | 2.91 | 0.19 |
| | $V_6$ | 2.77 | 0.20 |
| **15 fps** | $V_1$ | 2.37 | 0.19 |
| | $V_2$ | 1.70 | 0.13 |
| | $V_3$ | 2.82 | 0.19 |
| | $V_4$ | 2.29 | 0.19 |
| | $V_5$ | 2.64 | 0.21 |
| | $V_6$ | 2.30 | 0.19 |
| **20 fps** | $V_1$ | 2.40 | 0.19 |
| | $V_2$ | 1.70 | 0.14 |
| | $V_3$ | 2.58 | 0.20 |
| | $V_4$ | 2.29 | 0.19 |
| | $V_5$ | 2.64 | 0.20 |
| | $V_6$ | 2.19 | 0.20 |

TABLE I
MEAN AND 95% CONFIDENCE INTERVAL OF THE COLLECTED MOS.

## IV. CONCLUSIONS

In this paper a study on the impact of 2D visualization techniques for light fields has been presented. More specifically, six different visualizations have been tested through subjective tests to provide an insight on the preferred visualization. The obtained results show that, among the considered options, the best way to exploit the content of the light field while reproducing it on a 2D screen is to horizontally scan the views in order to reduce the shift among consecutive frames. This study highlights the issue of the

ISBN 978-90-827970-1-5 © EURASIP 2018 

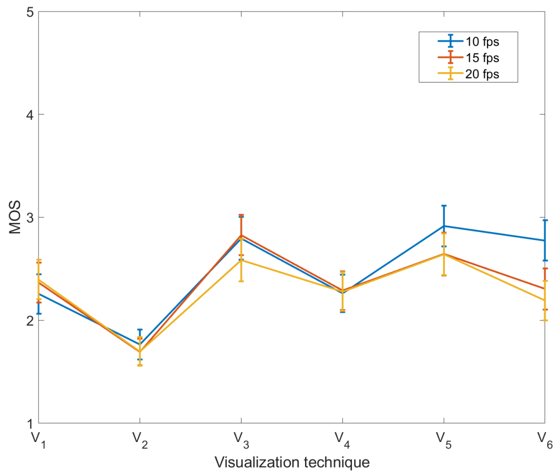



Fig. 6. MOS and 95% Confidence Interval for the performed experiments.

impact of content selection. In fact, it could be interesting to extend this study to LF images that span a wider range of values of CF and SI [23], and to verify how and if different values have an impact on the average and overall subjective scores.

## ACKNOWLEDGMENT

The authors would like to thank Pia Francesca Talia for her support in performing the experimental tests.

## REFERENCES

[1] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1-4, pp. 51–151, 1939. [Online]. Available: http://dx.doi.org/10.1002/sapm193918151

[2] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[3] B. Wilburn, M. Smulski, H.-H. K. Lee, and M. Horowitz, "The light field video camera," in *Proceedings of Media Processors, Electronic Imaging*. SPIE, 2002.

[4] M. Ziegler, R. op het Veld, J. Keinert, and F. Zilly, "Acquisition system for dense lightfield of large scenes," in *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, June 2017, pp. 1–4.

[5] L. Gabriel, "Epreuves reversibles. photographies integrals," *Comptes-Rendus Academie des Sciences*, vol. 146, 1908.

[6] G. Lippmann, *Sur la théorie de la photographie des couleurs simples et composées par la méthode interférentielle*. na, 1894.

[7] Lytro, "Lytro illum," https://www.lytro.com/, online; accessed 21 February 2018.

[8] Raytrix, "Light field technology," https://raytrix.de/, online; accessed 21 February 2018.

[9] S. Pastoor and M. Woepking, "3-D displays: A review of current technologies," *Displays*, vol. 17, no. 2, pp. 100 – 110, 1997.

[10] Y. Kobayashi, K. Takahashi, and T. Fujii, "From focal stacks to tensor display: A method for light field visualization without multi-view images," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 2007–2011.

[11] P. A. Kara, A. Cserkaszky, S. Darukumalli, A. Barsi, and M. G. Martini, "On the Edge of the Seat: Reduced Angular Resolution of a Light Field Cinema with Fixed Observer Positions," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, 2017.

[12] P. A. Kara, A. Cserkaszky, A. Barsi, and M. G. Martini, "The Couch, the Sofa, and Everything in between: Discussion on the Use Case Scenarios for Light Field Video Streaming Services," in *International Young Researcher Summit on Quality of Experience in Emerging Multimedia Services (QEEMS)*, Erfurt, 2017.

[13] T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, 2010.

[14] P. T. Kovcs, K. Lackner, A. Barsi, V. K. Adhikarla, R. Bregovi, and A. Gotchev, "Analysis and optimization of pixel usage of light-field conversion from multi-camera setups to 3d light-field displays," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 86–90.

[15] M. Agus, F. Bettio, A. Giachetti, E. Gobbetti, J. A. Iglesias Guitián, F. Marton, J. Nilsson, and G. Pintore, "An interactive 3d medical visualization system based on a light field display," *The Visual Computer*, vol. 25, no. 9, pp. 883–893, Sep 2009. [Online]. Available: https://doi.org/10.1007/s00371-009-0311-y

[16] 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.

[17] C. Perra, "On the quality evaluation of lossy compressed light fields," in *2016 24th Telecommunications Forum (TELFOR)*, Nov 2016, pp. 1–4.

[18] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," *CoRR*, vol. abs/1704.07576, 2017. [Online]. Available: http://arxiv.org/abs/1704.07576

[19] L. Shan, P. An, D. Liu, and R. Ma, "Subjective evaluation of light field images for quality assessment database," in *Digital TV and Wireless Multimedia Communication*, G. Zhai, J. Zhou, and X. Yang, Eds. Singapore: Springer Singapore, 2018, pp. 267–276.

[20] I. Viola, M. Rerabek, 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.

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

[22] EPFL, "Light-field image dataset," https://mmspg.epfl.ch/EPFL-light-field-image-dataset , online; accessed 01 June 2018.

[23] P. Paudyal, J. Gutiérrez, P. Le Callet, M. Carli, and F. Battisti, "Characterization and selection of light field content for perceptual assessment," in *Int. Conf. on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

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

[25] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Electronic Imaging 2003*. International Society for Optics and Photonics, 2003, pp. 87–95.

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

[27] ITU-T Rec. P.913, "Audiovisual quality in multimedia services: methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," 2016.

ISBN 978-90-827970-1-5 © EURASIP 2018