Perceptual analysis and characterization of light field content

Jesús Gutiérrez, Pradip Paudyal, Marco Carli, Federica Battisti, Patrick Le Callet

Introduction

The Light Field (LF) may be defined as the set of light-rays at every point in space traveling in every direction. The possibility of capturing this information provides a wide range of applications in various fields, such as surveillance, industrial and medical exploration, and immersive media technologies. In this sense, LF content allows novel ways to explore the captured scenes, like changing the parallax horizontally and vertically, and refocusing the content.

The novel applications provided by light field technologies entail a reconsideration of the methods for assessing Quality of Experience, starting with a proper characterization of the light field content.

This new imaging technology causes new challenges to the information processing system. To guarantee a successful development of the technology, the signal processing chain (coding, processing, delivering, storing) should take into account the peculiarities and the effects of possible impairments on the visual quality. To cope with these challenges, as experienced with previous audiovisual technologies, like 3D video, Quality of Experience (QoE) assessment is an essential factor.

Therefore, this article addresses, on one side, the requirements for properly tackling the perceptual aspects in LF processing, and on the other side, the proper characterization of LF content according to its applications.

LF basics: Perceptual perspective

Adelson and Bergen in [1] defined the plenoptic function to represent the intensity of light seen from any viewpoint, at any time instant, for any wavelength of the visible spectrum. The complexity associated with its high dimensionality can be reduced for practical imaging applications, as shown by the four-dimensional parameterization of the plenoptic function, which represents each light ray by its intersecting points with two parallel planes [2]. Therefore, the LF can be considered as a collection of perspective images of same scene, each one taken from a different viewpoint.

Content acquisition

The previous statement leads to the most intuitive way to acquire LFs, based on camera arrays [3]. However, LFs can be also obtained by using plenoptic cameras, which are based on inserting a microlens array between the camera sensor and the main lens. The main lens creates an image that is re-mapped to the sensor by the microlens array, that provides multiple views of the scene in a single shoot [4]. The differences between both alternatives entail distinct processing of the content and perceptual effects. For instance, on one side, camera arrays provide a set of views with wider baselines and better spatial resolutions. On the other side, plenoptic cameras offer the advantage of being much easier to handle and provide a denser set of views, although they entail a complex decoding process of the raw data (including demosaicing, devigneting, rectification, etc.) that can introduce artifacts and whose perceptual effects should be further explored [5].

Representation formats

Once the raw data is processed, it is possible to use different representations of the LF depending on the application under study. Among these, we can cite viewpoint images (a.k.a. sub-aperture images in plenoptic cameras, representing the scene captured from different viewpoints), the entire plenoptic image captured by the plenoptic camera, microlens images (a.k.a. elemental images or micro-images, captured by each microlens of the plenoptic camera), or epipolar images (containing depth information of the captures scene) [6].

Processing and encoding

The different representation formats of the LF are directly related to the processing that is addressed in the following. One of the main issues regarding LF imaging is the extraction of the 3D information of the captured scene. Nowadays, depth estimation and 3D reconstruction are active research areas [7]. Moreover, increasing the spatial and angular resolution of the acquired content is an important issue to be solved in order to offer improved image quality and 3D perception to the viewers [7]. Finally, given the high redundancy of LF content, many efforts are being devoted to the design of efficient compression techniques [6][8][9].

Rendering

The rendering and display of LF content are also a major issue directly influencing end users' QoE. For example, a simple approach is based on using conventional displays simulating LF applications, like interactive refocusing or viewpoint sweeping. To fully take advantage of the immersivity and interactivity of this content, head-mounted displays (HMDs) may be used if the content has been appropriately captured (e.g., omnidirectional LF content); at the same time, LF displays, which are still under development, will be the best way for visualizing LF content without the need of any other specific equipment (e.g., HMD, glasses, etc.) [3].

Related Works

This section presents an overview of the first efforts towards the QoE evaluation for LF content that have been made lately.

Datasets

Ongoing efforts are devoted to creating LF datasets, especially after the availability of affordable plenoptic cameras (e.g., Lytro and Raytrix). For example, the EPFL LF dataset provides 118 images captured with the Lytro Illum camera and covers a wide range of high-level features [10] and the SMART LF dataset collects 15 LF images designed for image quality assessment [9]. Similarly, Daudt and Guillemot published a Lytro Illum LF dataset containing 43 images for various applications, such as depth estimation, inpainting and compression [11]. Furthermore, it is also worth noticing the existence of datasets generated by different devices, such as the Stanford light field archive [12] captured with a camera array, and synthetic LF dataset generated by computer graphics [7]. However, there is still a lack of further datasets with annotated data from subjective tests to support the research on LF technologies.

Quality assessment

The new possibilities that immersive media technologies offer to the user experience require a revision of traditional methods for QoE evaluation. For example, as the appearance of 3D content entailed the consideration of evaluating visual discomfort and 3D perception in comparison with conventional video content, factors involved in the new immersive experience should currently be addressed, such as full-parallax, adaptive refocusing, interactivity, immersivity, cyber-sickness, etc. In addition, other aspects of QoE evaluation should be further investigated, such as appropriate testing environments, methodologies, and proper test content.

In this sense, some initial works have been proposed for evaluating the quality of LF content. In particular, various studies have been presented dealing with the quality assessment for LF compression algorithms. For example, Viola *et al*. [8] carried out a subjective test to compare different encoding approaches for LF images and analyzed the performance of traditional objective metrics like PSNR and

SSIM on this content. Similarly, Paudyal *et al*. [9] carried out an exhaustive analysis of subjective and objective quality evaluation of compressed LF images, using traditional methodologies and metrics. Apart from these approaches, only few works have addressed the QoE evaluation of LF in relation with other aspects, such as the visualization of LF content in LF displays (which are still under development) [13], and the effect of interactivity when the user is able to change the focus of the image and the viewpoint [14]. Taking this into account, further research on appropriate methodologies for subjective assessment and on reliable objective metrics for LF content is required to correctly evaluate perceptual and technical factors on the QoE.

Characterization of LF content

One of the main issues to deal with when assessing the QoE is the selection of contents to use in the tests under study, which should be based on visual characteristics and on the purpose of the experiment, rather than on personal preference or convenience [15]. This fact emphasizes the need for content characterization to model those aspects. In fact, important efforts have already been made to properly characterize 2D content, usually focused on analyzing spatial, temporal, and color features [16]. Moreover, the advent of 3D content with its new features (e.g., horizontal disparity, depth range, visual discomfort) showed the need for integrating novel features for a complete data characterization [17].

In this sense, the novel characteristics and applications of emerging immersive media technologies require a reconsideration of content characterization. With this aim, we proposed a framework for characterizing and selecting LF content [18], which will be summarized in the following. This framework was especially designed for QoE assessment, considering the new applications that LF technology provides, such as adaptive refocusing and full parallax.

Proposed scheme

The proposed framework is based on the analysis of various indicators representing 2D properties, together with 3D features and refocusing characteristics, given the importance of depth information provided by LFs and the novel possibilities of changing the focused elements of the content. The considered features are described in the following, and some illustrative examples are shown in Figure 1:

- *Spatial and temporal information*: The SI recommended by the ITU is widely used for this purpose, so it was adopted in the proposed framework [16]. Similarly, altough the proposed framework was dedicated to LF images, the TI recommended by the ITU may be used for describing the temporal aspects of video sequences.
- *Colorfulness*: It is an important visual feature having a significant impact on the perceptual quality of a scene. Thus, the proposed framework recomends to use the metric proposed by Hasler *et al.* [19], given its proved performance.
- *Contrast:* This property conveys meaningful perceptual information (e.g., textures, entropy, etc.). In the proposed scheme, the use of the Gray Level Co-occurrence Matrix (GLCM) is adopted for textural and contrast description [20].
- *Depth map and depth histogram*: Different approaches should be used depending on whether the LF data has been acquired by camera arrays (e.g., multi-view methods) or by plenoptic cameras (e.g., especific methods based on multi-view correspondences or occlusions [21]), due to the different acquisition properties (e.g., baseline). In the proposed scheme, for simplicity, the Lytro Desktop software was used to obtain the depth maps and from them, the histograms were computed.
- *Disparity range*: It defines the distance, in terms of pixels, corresponding to the nearest and furthest objects of the

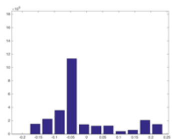
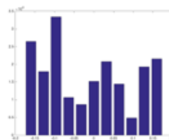
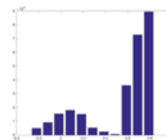
Preview			
Dataset	Own	EPFL	Own
Application	Viewpoint changing	Refocusing	Viewpoint changing & refocusing
Spatial Indicator	34.10	36.14	55.21
Colorfulness	10.12	45.85	38.97
Contrast	0.07	0.14	0.59
Refocusing Range	[-0.4,0.2]	[-0.3, 0.4]	[-1.4, 0.1]
Occluded Pixels	930	813	171
Disparity Range	[-0.16, 0.22]	[-0.16, 0.16]	[-0.22, 0.79]
Depth Distribution			

Figure 1. Characterization examples [18]

scene. To obtain this, the range of the scene in terms of distances to the objects or the camera calibration parameters, are required. Also, pixel disparities may be obtained by using estimation methods, such as the multi-view stereo algorithm [7], used in the proposed scheme.

- *Occlusions*: Although it is one of the most important problems when dealing with 3D content, only few algorithms address occlusions in LF. In the proposed scheme the amount of occluded pixels was computed using the algorithm by Wang *et al*. [21].
- *Refocusing range*: This describes the region from the nearest to the furthest elements of the scene that can be focused. For this purpose, it is possible to analyze the properties of the disparity histogram, which provides information about the depth distribution of the scene (as shown in Figure 1). Also, some objective metrics may be helpful, such as those developed for coping with the blur effect, or some specific approaches for LF content, like the Multi-focal Scene Defocus Quality (MSDQ) metric [22]. Finally, it is possible to use refocusing algorithms (e.g., "shift & sum" proposed by Ng *et al.* [4]) to determine the refocusing range going from the nearest to the furthest object in the scene. In the proposed framework, an implementation of this

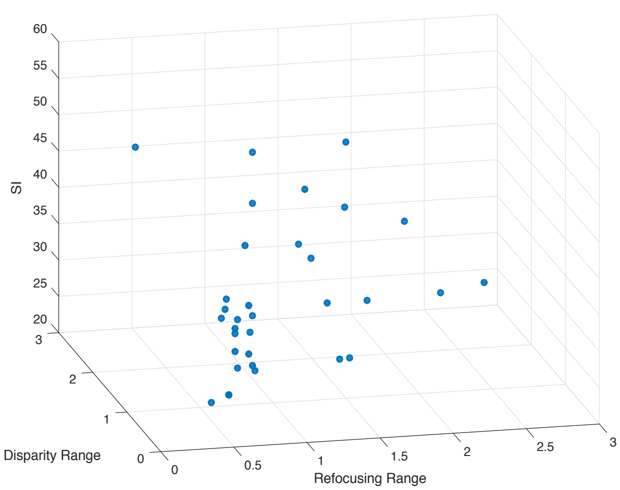


Figure 2. 3D scatterplot with SI, refocusing and disparity ranges [18].

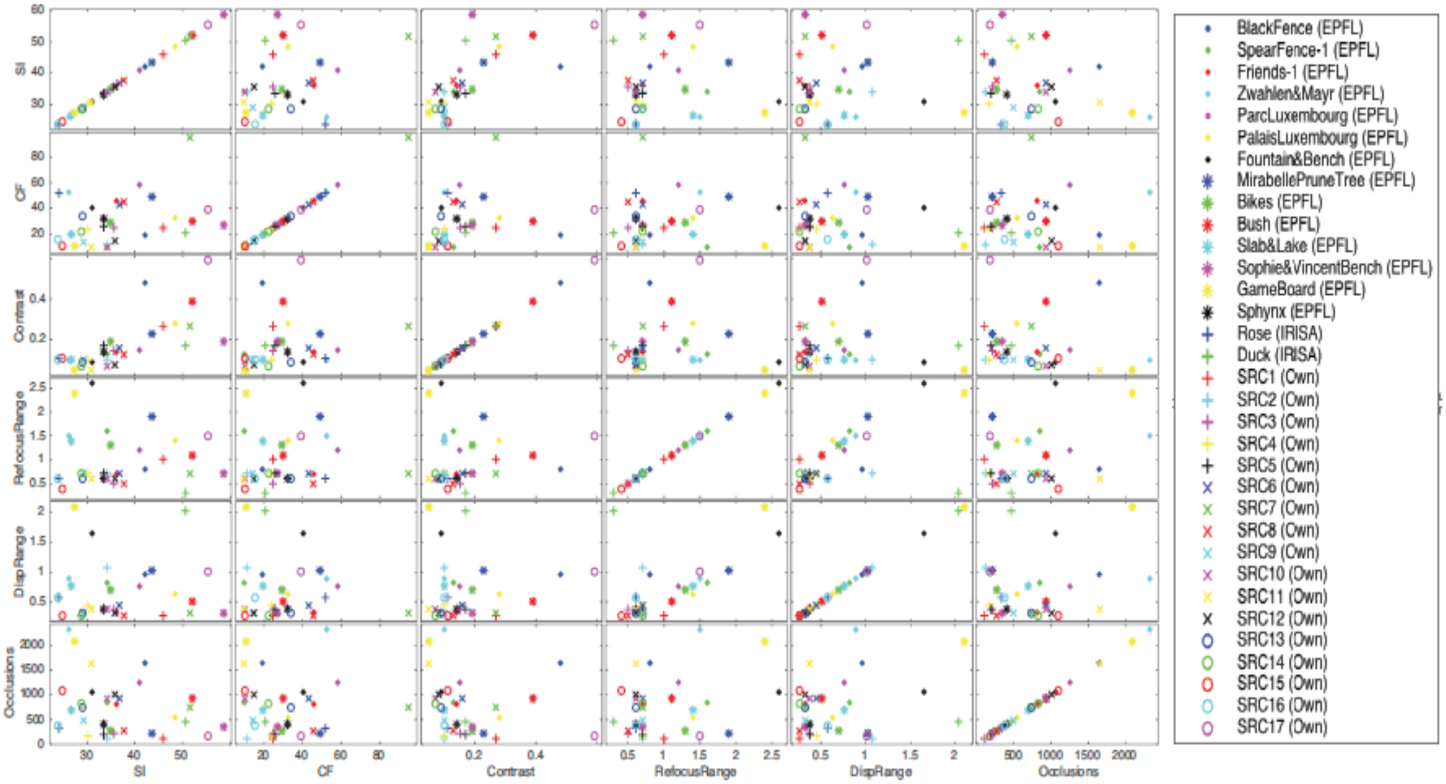


Figure 3. Scatterplot matrix of the main selected features [18].

algorithm was used to define the refocusing range [5].

The set of selected features can be graphically represented by different means based on the specific purpose, such as content selection based on a set of features. As an example, it might help to identify the lack of useful content as presented in the following. Figures 2 and 3 show two possible representations. In this case, images from different datasets where considered. A limited diversity of contents for important LF features, such as refocusing and disparity ranges might be noted. This may emphasize the need for generating and publishing more datasets for which the proposed approach for LF content characterization may be useful.

Conclusions

This article provided an overview of the perceptual aspects related to the processing and QoE assessment of LF content, and highlights the need for a revision of these aspects should be

__Jesús Gutiérrez__ is a post-doctoral researcher (Marie Curie/Prestige fellow) at the IPI group pf the LS2N of the Université de Nantes. He received the Telecommunication Engineering degree in 2008 from the Universidad Politécnica de Valencia (Spain), and the Ph.D. degree in Telecommunication in 2016, from the Universidad Politécnica de Madrid (Spain). His research interests are in the area of image and video processing, evaluation of multimedia quality of experience, and 3D and immersive media systems.

addressed, given the new possibilities and applications provided by emerging immersive technologies. In this context, we also provided some insights on proper LF content characterization as a first step towards further research on QoE assessment.

Pradip Paudyal *is Assistant Director at Nepal Telecommunications Authority (NTA), Kathmandu, Nepal. He received the M.Sc. degree in Information and Communication Engineering from Tribhuvan University, Kathmandu, Nepal, in 2010 and the Ph.D. degree from Department of Engineering at the Università degli Studi Roma TRE, Roma, Italy, in 2017. He is also serving as a Visiting Lecturer in different universities in Nepal. His research interests are in the area of multimedia signal processing, light field image processing, and perceptual quality assessment.*

Marco Carli *is Assistant Professor with the Department of Engineering at the Università degli Studi 'Roma TRE', Roma, Italy. He received the Laurea degree in Telecommunication Engineering from the Università degli Studi di Roma 'La Sapienza', Roma, Italy and the Ph.D. degree from Tampere University of Technology, Tampere, Finland. He was a Visiting Researcher with the Image Processing Laboratory, UCSB, USA (2000-2004). His research interests are in the area of digital signal and image processing with applications to multimedia communications. He is Area Editor of Elsevier Signal Processing: Image Communication.*

References

[1] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision", in "Computational models of visual processing", M. Landy and J. A. Movshon, Eds. MIT Press, pp. 3–20, 1991.

[2] M. Levoy, "Light fields and computational imaging", *IEEE Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[3] K. Akeley, "Envisioning a light field ecosystem", *SID Symposium Digest of Technical Papers*, vol. 43, no. 1, pp. 459–462, Jun. 2012.

[4] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera", *Stanford Tech Report CTSR*, Apr. 2005.

[5] D. G. Dansereau, O. Pizarro, and S. B.Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras", *Proc. Of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1027–1034, Portland, USA, Jun. 2013.

[6] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, and F. Dufaux, "Full parallax 3D video content compression", in "Novel 3D media technologies", A. Kondoz and T. Dagiuklas, Eds., Springer, 2015.

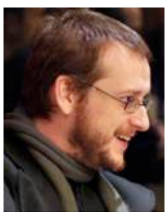
[7] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[8] I. Viola, M. Řeřábek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and Subjective Evaluation of Light Field Image Compression Algorithms", *Proc. of Picture Coding Symposium*, Nuremberg, Germany, Dec. 2016.

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

[10] M. Řeřábek and T. Ebrahimi, "New light field image dataset", *Proc. Of the International Conference on Quality of Multimedia Experience*, Lisbon, Portugal, Jun. 2016.

Federica Battisti *received the Laurea Degree (Master) in Electronic Engineering and the PhD degree from Roma Tre University in 2006 and 2010 respectively. Her research interests include signal and image processing with focus on subjective quality analysis of visual contents. She is currently assistant professor at the Department of Engineering at Roma Tre University.*



Patrick Le Callet *received both an M.Sc. and a PhD degree in image processing from Ecole polytechnique de l'Université de Nantes. Since 2003, he teaches at Ecole polytechnique de l'Université de Nantes where is now a Full Professor. His current centers of interest are Quality of Experience assessment, Visual Attention modeling and applications, Perceptual Video Coding and Immersive Media Processing. He is co-author of more than 250 publications and communications and co-inventor of 16 international patents on these topics. He serves or has served as associate editor or guest editor for several Journals such as IEEE TIP, IEEE STSP, IEEE TCSVT, SPRINGER EURASIP Journal on Image and Video Processing, and SPIE JEI. He is serving in IEEE IVMSP-TC (2015- to present) and IEEE MMSP-TC (2015-to present) and is one the founding member of EURASIP SAT (Special Areas Team) on Image and Video Processing.*

[11] R. Daudt and C. Guillemot, "Lytro Illum light field dataset", 2016. Online: https://www.irisa.fr/temics/demos/IllumDatasetLF/index.html

[12] V. Vaish and A. Adams, "The (new) Stanford light field archive," 2008. Online: http://lightfield.stanford.edu/

[13] 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", *Proc. Int. Conf. on 3D Imaging*, pp. 1–7, Liege, Belgium, Dec. 2016.

[14] I. Viola, M. Řeřábek, and T. Ebrahimi, "Impact of interactivity on the assessment of quality of experience for light field content", *Proc. of the International Conference on Quality of Multimedia Experience*, Erfurt, Germany, Jun. 2017.

[15] M. Pinson, M. Barkowsky, and P. Le Callet, "Selecting scenes for 2D and 3D subjective video quality tests", *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 50-61, Aug. 2013.

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

[17] M. Urvoy, J. Gutiérrez, M. Barkowsky, R. Cousseau, Y. Koudota, N. García, V. Ricordel, and P. Le Callet, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences", *Int. Workshop on Quality of Multimedia Experience*, pp. 109–114, Yarra Valley, Australia, Jul. 2012.

[18] P. Paudyal, J. Gutiérrez, P. Le Callet, M. Carli, and F. Battisti, "Characterization and selection of light field content for perceptual assessment", *Proc. of International Conference on Quality of Multimedia Experience*, Erfurt, Germany, Jun. 2017.

[19] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images", *Proc. of the SPIE*, vol. 5007, pp. 87–95, Jun. 2003.

[20] R. M. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification", *IEEE Trans. on systems, man, and cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[21] T.-C. Wang, A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2170–2181, Nov. 2016.

[22] W. Wu, P. Llull, I. Tosic, N. Bedard, K. Berkner, and N. Balram, "Content-adaptive focus configuration for near-eye multi-focal displays", *Proc. of the IEEE International Conference on Multimedia and Expo*, Seattle, USA, pp. 1–6, Jul. 2016.