An analysis of HEVC compression for light field image refocusing applications

Cristian Perra, Daniele Giusto
Department of Electrical and Electronic Engineering
University of Cagliari
Cagliari, Italy
cristian.perra@unica.it, ddgiusto@unica.it

***Abstract*—A light field image describes a sampling of the intensity and direction of the light rays in space. The recent technological advances in light field camera devices are fostering the development of novel imaging application. Some of the main applications are post-processing for interactive image perspective change, image refocusing, point cloud reconstruction, and image segmentation. The problem of light field data compression is a current research topic for international standardization bodies, industries, and research institutes. This paper explores the problem of compression with respect to light field image refocusing application. The main contribution of this paper is an analysis of the performance of high-efficiency video coding in terms of peak signal to noise ratio and structural-similarity index metric versus compression ratio. The experimental analysis has been performed on a reference dataset currently used for the standardization activities of ISO/IEC JPEG Pleno and shows the tradeoff between compression and objective quality at different quantization parameters.**

Keywords—light field coding, plenoptic coding, light field compression, plenoptic compression, PSNR, SSIM.

I. Introduction

A light field describes the amount of light (radiance) at every point in space from every direction. A plenoptic image is captured by sampling the radiance with camera arrays, camera with a microlens array, or camera rigs. Moreover, synthetic light field signal can be generated by software modeling, and true light fields signals can be fused with synthetic light fields for augmented reality applications.

The recent technological development of light field devices is fostering the study of new consumer and industrial applications. To understanding potentialities and limitations of novel applications based on light field technologies, several aspects must be considered. Compression, quality evaluation (both objective and subjective), and quality of experience are some of the research areas currently investigated by industries, research institutes, and standardization bodies, such as, for example, the ISO/IEC JPEG Pleno ongoing standardization activity.

Light field imaging technologies are promising to foster the development of novel industrial, entertainment, and scientific applications. Several aspects of light field imaging must be explored to determine possible limitations and performances before the development of applications. Efficient compression methods, novel objective light field image quality measures, and novel quality of experience evaluation methodologies in light field applications are among the novel research topics that are starting to be explored by the scientific community.

In this research paper, the light field data taken into consideration is obtained with camera devices with microlens array technology. Such devised are commonly known as plenoptic cameras and are used for sampling the light rays crossing the principal lens of the digital camera. Placing an array of microlens next to the digital photo-sensor is one of the possible means for sampling the light field information.

The purpose of a single microlens in the microlens array is to split the ray lights focusing on it into the different directional component that will then be sampled by one of the photosensor elements behind the microlens array.

Regular images and light field images despites are both a sampling of light rays have completely different post-processing phases. In fact, a regular camera generates a bi-dimensional signal that represents directly the visual information of the scene. Instead, light field data cannot be used as it is for a visual display of the scene and several post-processing steps are required, depending on the target application, for rendering a view to the final user. A micro-image originated by a microlens is representing a single point in the 3D space, and the color information of that point with respect to the different sampled directions.

Using computational photography algorithms, a true color image can be reconstructed from light field data as if it was shot from a regular camera. After setting information such as aperture, position, lens plane to microlens distance, it is possible to compute a synthetic view corresponding to a regular image with a given focus and aperture.

The required steps for rendering an image are as follows. At first, camera calibration is required to correct the effect of lens distortion [1], [2]. Then, a conversion from the light field data and the commonly used 4D light field representation is computed [3], [4]. Finally, the 4D light field is processed and a 2D view is rendered [5], [6].

Light field data are more demanding in terms of storage capacity requirements with respect to regular image at the same rendered image resolution. This is due to the fact that several direction components available in the light field data are used for generating a single pixel in the final rendered image.

Lossless and lossy coding algorithms have been proposed for light field data coding [7], [8], [9], [10]. Light field compression originates artifacts into the final signal affecting in this way the image quality of the final views and this problem is currently under research [11].

Standard image and video coding algorithm have been often used for light field data coding. JPEG, JPEG 2000, and high-efficiency video coding (HEVC) standards are the most used for lossy or lossless (e.g. JPEG2000) coding [12], [13]. Light field data reorganization into video sequences before applying video coding standard compression are another possible solution to the problem of light field data coding [14].

978-1-5386-3679-4/18/$31.00 ©2018 IEEE

Light field applications development requires that novel definition of procedures for evaluating the quality of experience as perceived by the user is devised. At the same time, data transmission and data security aspects will need to be considered [15-26].

Image refocus, perspective change, and 3D point cloud reconstruction are some of the light field applications that are currently under research [27]. The comparison of different research architectures and methodologies for light field compression and quality evaluation requires the availability of reference datasets. Datasets containing light field data with associated metadata are available in [28-31].

The ISO/IEC committee has started a standardization activity called JPEG Pleno, aiming at the development, among others, of standards for light field coding.

In this paper, a method for the evaluation of the performance of HEVC compression of light field data is proposed. The experimental evaluation is performed on a dataset selected from the database proposed for the ISO/IEC JPEG Pleno standardization activity. The distortion introduced by compression is measured in terms of peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM).

The remainder of this paper is organized as follows. Section 2 discusses the main concepts of light field imaging. The proposed method for evaluating the performance of high-efficiency video coding compression in light field image refocusing application is presented in Section 3. Section 4 presents and discusses the obtained result. The conclusions of the paper are drawn in Section 5.

II. Light Field Imaging Overview

A light field describes the amount of light at every point in space from every direction. Light field data is obtained by sampling the radiance with a camera array, a single camera with a microlens array (plenoptic camera), or a camera rig.

The three capturing techniques differ in the field of view size which is wide when capturing with an array of cameras or a camera rig and is narrow when capturing with a light field camera.

A light field is described by the so-called plenoptic function that describes the intensity and the direction information of light rays at every point in space. A subset of the plenoptic function can be captured and sampled with a plenoptic camera.

A 4D light field is denoted as

$$\mathrm{L}(u, v, s, t) \tag{1}$$

Where (u, v) and (s, t) are the parameters describing the direction of the light ray and L is the intensity of the light ray.

The light field data sampled by the photo sensor is denoted as

$$\mathrm{LF}(i, j, k, l) \tag{2}$$

where $1 \leq i \leq M, 1 \leq j \leq M, 1 \leq k \leq P, 1 \leq l \leq Q$. The different light ray directions that have been captured (i.e. the number of views) are $\mathrm{M} \times \mathrm{M}$ and the resolution of each view is $\mathrm{P} \times \mathrm{Q}$.

A light field refocusing application process the light field to render images at a given focal point. Equation 4 defines the reconstruction of a focused image controlled by the parameter α. When $\alpha = 1$ the focus of the final image is at the light field camera focus plane, when $\alpha < 0$ the final image is focusing in the proximity of the camera, when $\alpha > 0$ the camera is focusing to infinity.

$$R(\mathrm{y}', \mathrm{x}') = \frac{1}{\mathrm{M}^2} \sum_{\mathrm{i}=1}^{\mathrm{M}} \sum_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{LF}\left(\mathrm{i}, \mathrm{j}, \frac{\mathrm{y}'}{\alpha} + \mathrm{u}'\left(1 - \frac{1}{\alpha}\right), \frac{\mathrm{x}'}{\alpha} + \mathrm{v}'\left(1 - \frac{1}{\alpha}\right)\right) \tag{3}$$

III. Proposed Method

The method for evaluating the performance of HEVC compression in light field image refocusing application is presented in this section. The input for the encoding process is a 4D light field as defined in Section 2.

The light field views are ordered following a spiral scan. The views at the corner of the light field are skipped because they do not contain enough visual information for being used as input for the rendering process.

The obtained set of views is considered as a video sequence and in passed as input to the HEVC encoder. The first frame is Intra coded while all the other frames are Inter coded. The encoded file is then decoded. The output of the decoder is passed to a light field refocusing tool that renders seven different views at different focus point.

During the coding, decoding and rendering process, some metrics are computed. The bitrate is computed from the size of the encoded bitstream.

The source focus planes are computed for the input 4D light field before compression, then the rendered focus planes are computed for the decoded 4D light field.

The PSNR is computed between corresponding input and rendered focus plane and finally the average PSNR is computed.

The reference architecture for light field refocusing application is shown in Fig. 1.

IV. Experimental analysis

The raw light fields chosen for the experimental analysis have been selected from the dataset reported in [28] and are summarized in Table I. Fig. 2 shows the thumbnails of the views extracted from the dataset. The light fields have been acquired with a Lytro Illum Camera. The microlens array in the Lytro Illum Camera is arranged as a honeycomb structure and each microlens project the light on nearly 150 pixels on a sensor having 40 Mpixel resolution.

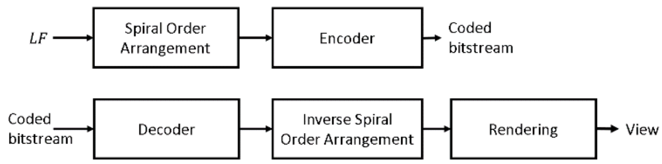


Fig. 1. Reference architecture for light field refocusing application.

TABLE I. DATASET OF RAW LIGHT FIELD IMAGES USED FOR THE EXPERIMENTAL ANALYSIS.

ID	Filename	Content
LF1	Bikes	Part of a bike
LF2	Danger_de_Mort	Metallic grid fence
LF3	Flowers	Yellow flowers on the ground
LF4	Stone_Pillars_Outside	Lined stone pillars
LF5	Fountain_&_Vincent_2	A fountain and a man
LF6	Friends_1	Faces of a group of people

White Images are stored in the Lytro Illum Camera memory to serve as calibration data and are hence associated to each light field camera in order to make possible the reconstruction of the 4D light field from the raw light field.

Raw light fields (RLF) from the dataset have resolution 5368×7728. The resolution of $\mathrm{LF}(i, j, k, l)$ is $15 \times 15 \times 435 \times 625$.

The raw light fields from the dataset are transformed into the 4D light fields. Each 4D light field is then encoded with the procedure presented in Section 3 using six different quantization parameters.

The HEVC quantization parameters that have been set for coding are $\mathrm{QP} = \{0,10,20,30,40,50\}$.

The encoded bitstreams are then decoded and rendered at seven different focus point. As described in Section 3, there is one parameter (α) controlling the focus point rendering. Seven focus points have been set for the experimental analysis and are $\alpha = \{-0.8, -0.4, -0.2, 0, 0.2, 0.4, 0.8\}$.

Fig. 3 shows an example of light field refocusing application where seven views at different focus point. It can be observed that Fig. 3(a) is focusing towards the camera, in fact the bike tire is in focus while the rest of the image is blurred.

On the contrary, Fig. 3(g) is focused far from the camera, in fact the grass at the background is in focus while the rest of the images is blurred.

Fig. 2. Central view from the light field dataset. (a) bike; (b) danger de mort; (c) flowers; (d) stone pillars; (e) fountain; (f) friends [28].

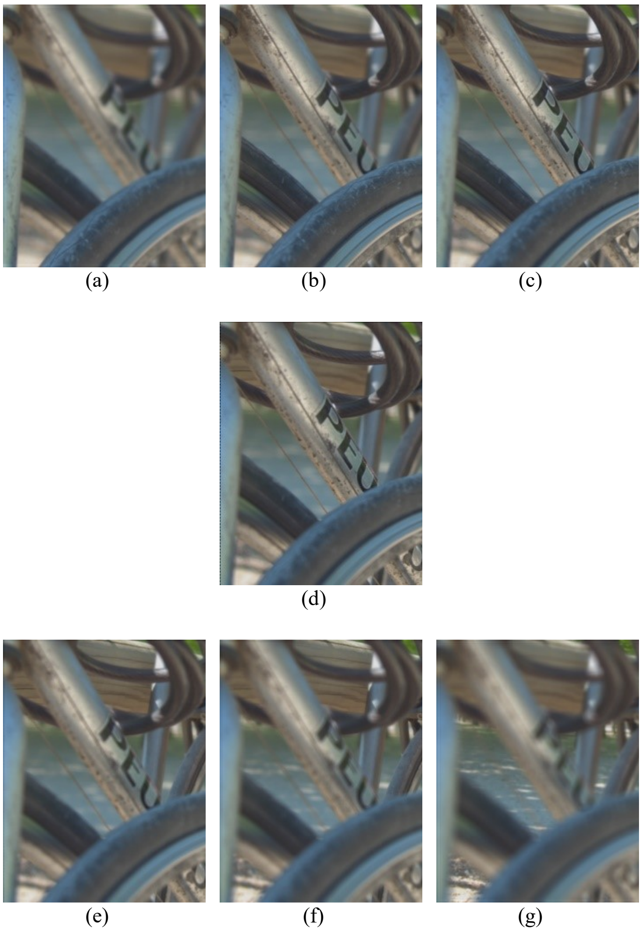


Fig. 3. Light field refocusing application. (a)Refocus at $\alpha = -0.8$; (b)Refocus at $\alpha = -0.4$; (c) Refocus at $\alpha = -0.2$; (d) Refocus at $\alpha = 0$; (e)Refocus at $\alpha = 0.2$; (f)Refocus at $\alpha = 0.4$; (g)Refocus at $\alpha = 0.8$.

The other focus points are intermediate between the two limit cases and Fig. 2(d) is focusing in the metallic part of the bike. The objective quality evaluation has been performed in terms of peak-signal-to-noise-ratio (PSNR) and structural similarity index (SSIM).

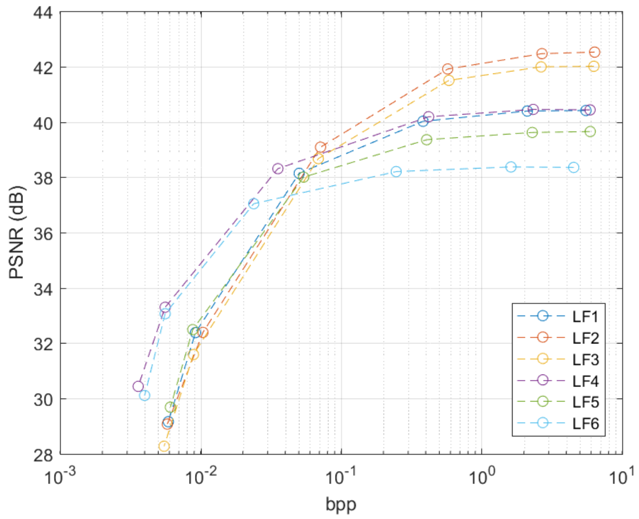
Fig. 4(a) shows the obtained PSNR results. The horizontal axis is the bitrate in logarithmic scale, the vertical axis is the PSNR in dB. Six rate-distortion point are shown corresponding the six different quantization parameters

Fig. 4(b) shows the obtained SSIM results. Again, the horizontal axis is the bitrate in logarithmic scale, the vertical axis is the PSNR in dB. Six rate-distortion point are shown corresponding the six different quantization parameters.

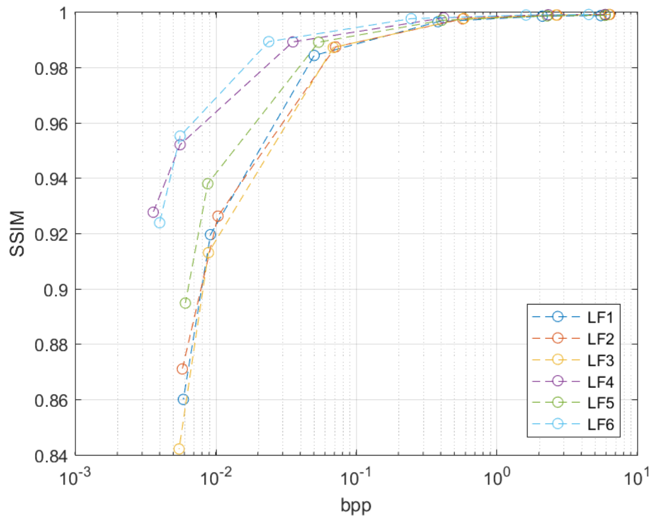
It can be observed that for all rates above 10^{-1}bpp the PSNR quality is higher than 38dB, which can be acceptable for some consumer applications.

This observation is also confirmed by the SSIM results reported in Fig. 3(b) where for all rates above 10^{-1}bpp the SSIM index is higher than 0.98.

Recalling that the experiments have been performed at fixed quantization parameters it can be noted that there is a high variability in terms of PSNR results at medium bitrates (nearly $5 \cdot 10^{-1}$bpp) while this variability is less evident at low bitrates (nearly $5 \cdot 10^{-2}$bpp).



(a)



(b)

Fig. 4. Rate distortion results. (a) PSNR (db) vs bitrate (logarithmic scale); (b) SSIM vs bitrate (logarithmic scale).

V. CONCLUSIONS

Novel coding tools are needed to efficiently store and transmit light field data. Current research activities are exploring different coding architectures mainly based on the exploitation of state of the art coding algorithms. Several aspects of light field imaging need to be explored to fully understand possible limitations and performances before the development of novel coding tools. Efficient compression methods, objective quality measures and techniques for evaluating the quality of experience in viewing rendered light fields, are among the novel research topics that are starting to be explored by the scientific community. In this paper, coding performance of high efficiency video coding applied to the 4D light field organized as a video sequence is evaluated considering light field refocusing applications as a use case. The experimental analysis has been performed measuring the resulting distortion at different bit rates in terms of PSNR and SSIM quality metrics. The proposed evaluation procedure, the rate distortion results, and the knowledge of the tradeoff between coding compression parameters and objective quality of rendered views can help in the design of novel light field refocusing application.

ACKNOWLEDGMENT

This work has been partially supported by the Italian Ministry of University and Research (MIUR), within the Smart City framework (project: PON04a2_00381 “CAGLIARI2020”, and by the Italian region of Sardegna, within the top-down cluster action program (project POR FESR Sardegna 2014/2020, “DigitArch”).

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. S. Landy and J. A. Movshon, Eds. Cambridge, MA: MIT Press, 1991.

[2] F. Murgia, S.-B. Fernández, C. Perra, and D. Giusto, “Unfocused plenoptic camera calibration,” in 22nd Telecommunications forum, Belgrade, Serbia, November 2014.

[3] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, “Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction,” in Computer Vision (ICCV), 2013 IEEE International Conference on, Dec 2013.

[4] D. Dansereau, O. Pizarro, and S. Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, June 2013.

[5] S.-C. Chan, K.-T. Ng, Z.-F. Gan, K.-L. Chan, and H.-Y. Shum, “The plenoptic videos: capturing, rendering and compression,” in Circuits and Systems, 2004. ISCAS’04. Proceedings of the 2004 International Symposium on, 2004.

[6] F. Murgia, D. Giusto, and C. Perra, “3D reconstruction from plenoptic image,” in 23rd Telecommunications Forum (TELFOR), Belgrade, Serbia, 2015.

[7] C. Perra, “On the coding of plenoptic raw images,” in 22nd Telecommunications forum, Belgrade, Serbia, November 2014.

[8] C. Perra and P. Assuncao, “High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement,” in IEEE International Conference on Multimedia and Expo (ICME), Seattle, USA, July 2016.

[9] C. Perra and D. Giusto, “JPEG 2000 compression of unfocused light field images based on lenslet array slicing,” in IEEE International Conference on Consumer Electronics, Las Vegas, USA, January 2017.

[10] C. Perra, F. Murgia, and D. Giusto, “An analysis of 3D point cloud reconstruction from light field images,” in International Conference on Image Processing Theory, Tools and Applications, Oulu, Finland, December 2016.

[11] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, “Scalable coding of plenoptic images by using a sparse set and disparities,” Image Processing, IEEE Transactions on, vol. 25, no. 1, 2016.

[12] R. S. Higa, R. F. L. Chavez, R. B. Leite, R. Arthur, and Y. Iano, “Plenoptic image compression comparison between JPEG, JPEG2000 and SPITH,” Cyber Journals: JSAT, vol. 3, no. 6, 2013.

[13] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod, “Light field compression using disparity-compensated lifting and shape adaptation,” Image Processing, IEEE Transactions on, vol. 15, no. 4, 2006.

[14] M. Magnor and B. Girod, “Data compression for light-field rendering,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 10, no. 3, 2000.

[15] C. Perra, “A framework for user control over media data based on a trusted point,” in 2015 IEEE International Conference on Consumer Electronics (ICCE), Jan 2015.

[16] C. Perra, P. A. Pes, and D. D. Giusto, “High-frequency error recovery in JPEG XR coded images,” in 18th IEEE International Conference on Image Processing, Sept 2011.

[17] C. Perra, “Re-encoding JPEG images for smart phone applications,” in 21st Telecommunications Forum (TELFOR), Belgrade, Serbia, Nov 2013.

[18] B. Saba, C. Perra, and D. Giusto, “A scalable image browsing application based on JPEG-XR,” in 12th International Workshop on Image Analysis for Multimedia Interactive Services, April 2011.

[19] B. Saba, C. Perra, and D. D. Giusto, “Image transmission over error prone channels,” in 19th Telecommunications Forum (TELFOR), Belgrade, Serbia, Nov 2011.

[20] C. Perra, “A low computational complexity blockiness estimation based on spatial analysis,” in 22nd Telecommunications Forum (TELFOR), Belgrade, Serbia, Nov 2014.

[21] C. Perra and D. Giusto, “An image browsing application based on JPEG XR,” in 2008 International Workshop on Content-Based Multimedia Indexing, June 2008.

[22] C. Perra, “Wireless cooperative access to images,” in 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), July 2013.

[23] B. Saba, C. Perra, and D. Giusto, “Interactive image viewing in mobile devices based on JPEG XR,” Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, vol. 79 LNICST, 2012.

[24] C. Perra, “Wireless cooperative access for next generation mobility applications,” in 21st Telecommunications Forum (TELFOR), Belgrade, Serbia, Nov 2013.

[25] G. J. Sullivan, J. R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, Dec 2012.

[26] C. Perra, “Assessing the quality of experience in viewing rendered decompressed light field,” Multimedia Tools and Applications, 2018.

[27] F. Murgia, C. Perra, and D. Giusto, “3D point cloud reconstruction from single plenoptic image,” Telfor Journal, vol. 8, no. 1, 2016.

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

[29] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, “Data formats for high efficiency coding of lytro-illum light fields,” in International Conference on Image Processing Theory, Tools and Applications, Orléans, France, November 2015.

[30] T. Georgiev, “Todor Georgiev Gallery of Light Field Data,” 2015. [Online]. Available: http://www.tgeorgiev.net/Gallery/

[31] C. Perra, “Lossless plenoptic image compression using adaptive block differential prediction,” in IEEE 40th International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, April 2015