International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10563-10569

Light Field Coding based on Flexible View Ordering for Unfocused Plenoptic Camera Images

Cristian Perra

Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy.
Orcid Id: 0000-0002-1506-423X

Abstract

Plenoptic cameras are devices designed for sampling the light field entering their main lens. The raw information captured by the sensor can then be processed for rendering views at different focal distance or from different perspective. A plenoptic camera generates images with high resolution and high redundant information if compared with a conventional camera having resolution equal to that of a light field rendered view. A light field compression architecture designed for coding views obtained from unfocused plenoptic images is proposed in this paper. The intent of this proposal is to provide a lossy compression tool enabling access to group of light field views. The experimental analysis aims at exploring the performance of the proposed method. The obtained results show that the proposed flexible ordering approach can provide advanced control to the user device in selecting the set of views that are effectively required for a given rendering application.

Keywords: Light field imaging, Lenslet coding, Plenoptic coding, Light field coding

INTRODUCTION

Plenoptic cameras are digital photo cameras designed for capturing the intensity and directional information of the light field in a scene. The capturing process makes use of a microlens array and is also known as integral imaging. Light field data is composed of several micro pictures corresponding to the information captured by a single microlens. Light field data processing is required to render the final images. The main applications are the representation of the original scene at different focus and perspective.

Light fields applications are very data intensive if compared with classical digital images at the same resolution considering that each pixel conveys not only the colour information but also the directional information of the sampled light field [1], [2], [3].

Ongoing research activities are exploring the problem of light field compression for the development of novel tools for an efficient storage and transmission of light field data [4], [5].

The reconstruction of a regular image from a raw light field image is accomplished in several processing steps. A preliminary and fundamental step is an accurate signal calibration [6]. A second step is the reconstruction of a 3D model of the light field from the calibrated raw image [7], [8]. The final step is to render a view from the 3D model of the light field [9]. High storage capacity is typically required for storing raw light field data. Hence, novel data coding architectures and algorithms are required for improving data compression performances. The study of the statistical properties of images captured by a light field camera is of fundamental importance for the research and development of light field compression algorithms.

In literature, the problem of sampling and compression of light fields is addressed in different ways. Light field compression can be implemented exploiting the knowledge of geometrical information of a scene together with the camera position for modelling the plenoptic function [10].

Other approaches are based on the compression of multiview data. A planar layer representation is adopted, where each layer corresponds to the same depth of the scene. A 3D wavelet transform is then applied to uncorrelate the information reducing the entropy of the information.

Reordering the light field information before subsequent processing can result in improved compression performance. Nevertheless, this aspect deserves further investigation [11], [12] to identify the best reordering techniques.

Another approach to light field coding is based on non standard coding architectures exploiting light field partitioning in cluster minimizing the differential pulse-code modulation (DPCM) entropy before a dictionary based entropy encoder is applied [13].

A system for capturing, rendering and compression of the plenoptic data is also presented in [14]. The results show that there is high correlation in the data of a plenoptic image. Moreover, standard coding approaches are not always able to properly detect such correlation and hence are not able to achieve optimal performance.

A scalable coding scheme for focused plenoptic images is proposed in [15] where a lossy coding is based on sparse set and disparity. Standard coding architecture have been explored for addressing the problem of lossy compression of plenoptic images [16], [17] evaluating the final quality using objective methods [18], . A set of calibrated images representing a static scene from different viewpoints can be encoded by

reconstructing the 3D geometry of the scene, generating a map of view-dependent textures, and finally encoding the textures maps are encoded by set partitioning in hierarchical trees (SPIHT) [19].

Locally linear embedding and self-similarity compensated prediction are proposed in [20], [21], [22] as preprocessing tools for light field compression exploiting HEVC coding tools.

Other techniques are based on displacement intra prediction applied to both focused and unfocused images [23], [24]. User interaction is often required before completing a 3D reconstruction.

In 2014, the JPEG standardization committee launched an new activity named JPEG PLENO aiming at devising a standard framework for the representation and exchange of new imaging modalities such as light field, point-cloud and holographic imaging.

The main contribution of this paper is the proposal of an architecture for flexible ordering light field views into groups of views to be encoded and independently accessed by a user device.

The paper is organized as follows. Section II summarizes the main concepts of light field images. The proposed architecture is described in Section III. The quality evaluation procedure is described in Section IV. Section V summarizes the experimental results. Section VI draws the conclusion of the paper.

LIGHT FIELD IMAGING OVERVIEW

A light field describes the radiance of the light rays at every point in space from every direction. Light field signals are captured by sampling the radiance with an array of cameras or with a single camera with a microlens array, named light field camera. The two capturing techniques differs in the field of view size which is wide when capturing with an array of cameras, and is narrow when capturing with a light field camera.

A light field is a function that describes the intensity and the direction information of light rays in every point in space. A subset of the light field can be captured and sampled with a light field camera (a.k.a. plenoptic camera).

A conventional model for a light field camera is shown in Fig. 1. The camera is composed by a main lens (Fig. 1(a)), a rectangular array of lenslets (Fig. 1(b)), and a digital sensor for sampling the light rays (Fig. 1(c)). A light ray enters the photo camera in a point (u,v), then hits the microlens array in a point (s,t) and is captured and sampled by the photosensor in a point (y,x). The light field is denoted as

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

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 digital light field sampled by the photo sensor is denoted as

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

where $1 \le i \le M$, $1 \le j \le M$, $1 \le k \le P$, $1 \le l \le Q$.

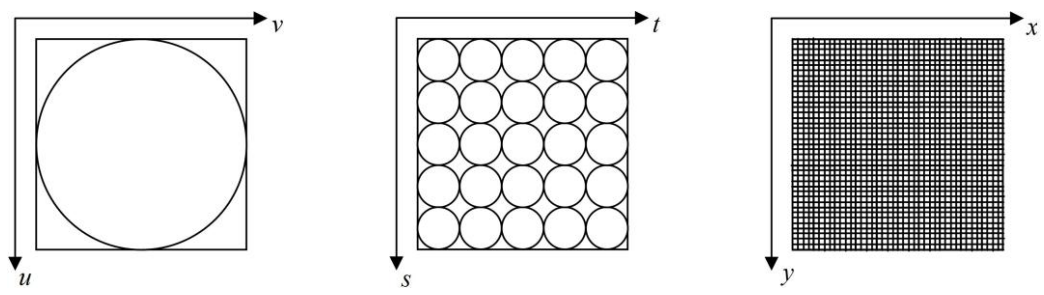
$M \times M$ are the different light ray directions that have been captured (i.e. the number of views), $P \times Q$ if the resolution of each view. A 2D representation of the stack of 2D images is obtained with the following transformation

$$I'(y',x') = LF\Big(y' \pmod M, x' \pmod M, \left\lfloor \frac{y'}{M} \right\rfloor, \left\lfloor \frac{x'}{M} \right\rfloor\Big) \tag{3}$$

The main light field computational imaging applications that have been developed so far provide the means for refocusing a picture and for changing the perspective of the image. This section presents a brief description of these applications.

Refocusing consist in processing the light field in order to render images at a given focal point. The theoretical aspects are discussed in detail in [25].

Eq. 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 proximity of the camera, when $\alpha \to 0$ the camera is focusing to infinity. A visual example of rendering light field at different focal point is shown in Fig. 2. Fig. 2(a)(e) shows the result for $\alpha = 0.2$, $\alpha = 0.6$, $\alpha = 1.0$, $\alpha = -0.6$, and $\alpha = -0.2$, respectively.



(a) (b) (c)

Figure 1: Conceptual model for a light field camera. (a) Main lens; (b) Rectangular microlens array; (c) Digital photosensor.

(a) $\alpha = 0.2$ (b) $\alpha = 0.6$ (c) $\alpha = 1.0$ (d) $\alpha = -0.6$ (e) $\alpha = -0.2$

Figure 2: Example of refocusing an image for different values of α.

(a) $(u_0,v_0) = (2,2)$ (b) $(u_0,v_0) = (7,7)$ (c) $(u_0,v_0) = (11,11)$ (d) $(u_0,v_0) = (13,13)$

Figure 3: Example of changing the perspective of an image for different values of (u_0,v_0). A diagonal parallax from the top-left to the bottom-right can be noted observing the relative movement of the edge of the table with respect to the car.

$$R(y',x') = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} LF\left(i, j, \frac{y'}{\alpha} + u'\left(1-\frac{1}{\alpha}\right), \frac{x'}{\alpha} + v'\left(1-\frac{1}{\alpha}\right)\right). \quad (4)$$

Perspective change consists in processing the light field in order to render images at a given viewpoint. Eq. 5 defines the reconstruction of perspective changed images controlled by the parameters u_0, v_0.

$$P(y',x') = L(u_0,v_0,y',x'), \quad (5)$$

where (u_0,v_0) are the coordinates of the virtual aperture point on the main lens of the plenoptic camera. The position of this point controls the viewpoint of the final image, $(u_0,v_0) = (M/2,M/2)$ correspond to the central view.

PROPOSED ARCHITECTURE

The proposed architecture is shown in Fig. 4 and Fig. 5. The pre-processing module receives light field data (LFV, light field views) and parameters as input. An example of input parameters is shown in Fig. 6. The parameters (PS, parameter set) are represented using JavaScript Object Notation (JSON).

The light field data is a set of views obtained as described in Section . The parameters describe the partitioning scheme to be applied for clustering the views in independent coding sets (GOV, group of views).

Additional parameters are used for selecting encoding tools and coding parameters (CS, coding set). For example, the parameters in Fig. 6 are used for selecting HEVC as coding tool (”encoder”:”hevc”) with a quantization parameter 12 (”qp”:12).

The encoder module compresses each GOV according to the set of coding parameters. The encoded GOV are then multiplexed together with the PS and the CS. The final stream is then saved in the storage block. When a client makes a request for a content, the server sends to PS information to the client. The client can then make different requests for each GOV that is needed by the rendering application. The remainder of this section describes the internal architecture of each module.

The inputs of the pre-processing block are a file containing the light field data ordered in M different views and a file containing the parameters needed for describing the properties of the light field data and for describing how this views must be partitioned before encoding. The pre-processing generates different GOV that are separately sent to the encoder.

The encoder block selects the coding tool to be applied to the input views, according to the information provided in the coding parameters.

The main purposes of the multiplexer is to merge the encoded GOV, the PS, and the CS in a single container file.

The light field stream are finally stored in the storage block in order to be available for subsequent requests from the light field server.

The light field server instantiates a service listening to incoming call for a given light field file. The first answer to a request is the PS scheme. Then the client can decide to make additional requests for extracting each of the available GOV.

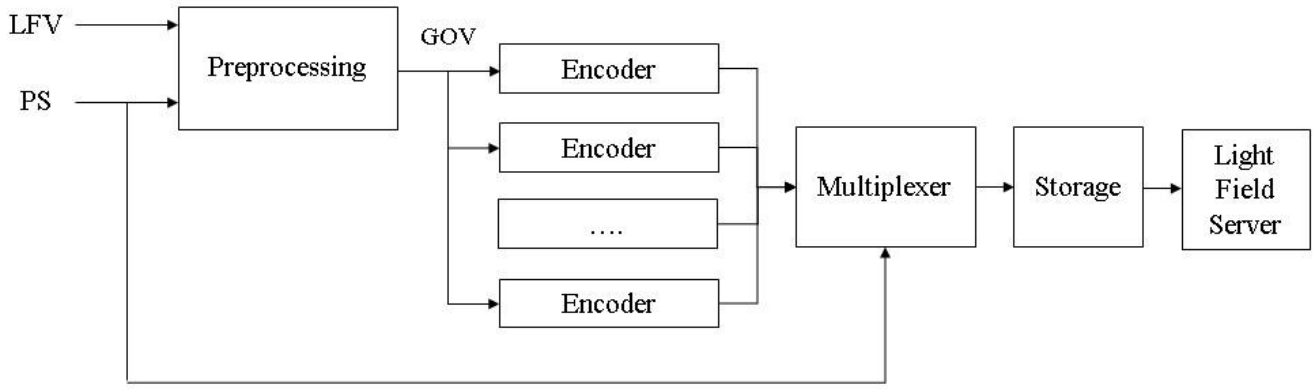


Figure 4: Encoding architecture.

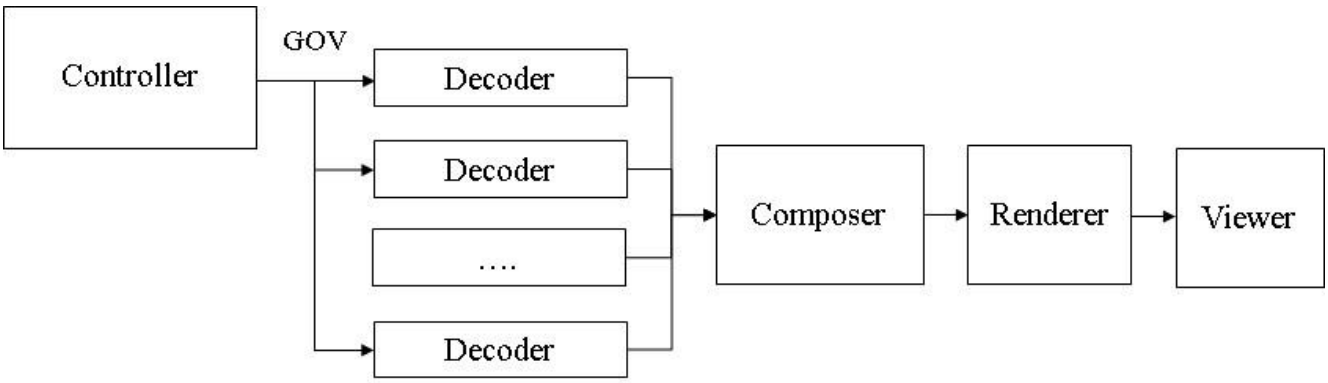


Figure 5: Decoding architecture.

The decoder receives the decoding parameters and a GOV form the controller block. Decodes the views in the GOV and sends the output to the composer block. The composer block reconstructs the light field from the available GOV.

The renderer processes the light field views for the final presentation on the client display. The type of processing performed by the renderer depends by the final application. Possible rendering processes, such as refocusing of perspective change, have been discussed in Section III.

QUALITY EVALUATION

This section discusses the procedure for the quality evaluation of the lossy compressed light fields. The analysis focuses on the quality of the reconstructed images with respect to the corresponding reference images.

The objective quality evaluation of the decompressed light field views is performed using the peak signal-to-noise ratio (*PSNR*) metric and the (*SSIM*) structural similarity index.

The RGB light field is converted into the YUV light field using the following transformation matrix (ITU-T Rec. 709):

$$\begin{aligned} Y &= 0.212600 * R + 0.715200 * G + 0.072200 * B \\ U &= -0.114572 * R - 0.385428 * G + 0.500000 * B \\ V &= 0.500000 * R - 0.454153 * G - 0.045847 * B \end{aligned} \quad (6)$$

The *PSNR* measures are defined as

$$PSNR(s,d,c) = 10log_{10}\frac{(2^b-1)^2}{MSE(s,d)} \quad (7)$$


```
{
"id":100001,
"name":"Bikes",
"viesw":15,
"viesh":15,
"width":625,
"height":434,
"colorInput":"rgb24",
"colorTransform":"yuv420",
```

```
"encoder":"hevc",
"qp":12,
"gops":[{
"gid":1,
"views":[113,81,83,85,
115,145,143,141,111
]
},{
"gid":2,
"views":[51,53,55,87,117,147,175,173,17
1,139,109,79]
}]
}
```

Figure 6: JSON coding parameters.

$$MSE(s,d) = \frac{1}{3(M-2)^2PQ} \cdot \sum_{i=2}^{M-1}\sum_{j=2}^{M-1}\sum_{k=1}^{P}\sum_{l=1}^{Q}\left(s\left(i,j,k,l,c\right)-d\left(i,j,k,l,c\right)\right)^2 \quad (8)$$

where *s* is the reference signal (i.e. the uncompressed light field *LF*), *d* is the signal to be evaluated (i.e. the decompressed light field), *b* is the bit depth of the samples, and *c* is the chroma channel (*c* = 1 denotes the *Y* channel, *c* = 2 denotes the *U* channel, and *c* = 3 denotes the *V* channel). The PSNR computed on the luminance channel is defined as

$$PSNR_Y(s,d) = PSNR(s,d,1) \quad (9)$$

The *SSIM* measures the similarity between two images. The *SSIM* considers the image degradation as perceived change in structural information and results in an index value between 0 and 1.

When reference and target images have no structural similarities then *SSIM* = 0. When reference and target are the same image then *SSIM* = 1. The mathematical description, additional information, and the algorithm of the SSIM are presented in [26].

EXPERIMENTAL ANALYSIS

The light field images used for the experimental analysis have been selected from a publicly available dataset [27].

The use case that has been considered for experimenting the proposed architecture considers a partitioning of the light field data into a group of views arranging the light field views as a sub-sampled spiral view starting from the central view (the GOV is composed of 37 views). This GOV is dubbed *base layer* in the remainder of this section.

The GOV is encoded with HEVC at different quantization parameter values (qp=0,18,28,32,36,40; qp=0 corresponds to the highest quality).

The reference architecture is the encoding of the whole light field arranged in spiral order starting from the central view [11].

The experiments on the use case aim at evaluating the coding gain that can be achieved introducing the proposed flexible view ordering method.

The rate-distortion results for two light fields are shown in Figs. 7-10. As expected, the *PSNR* of the base layer is higher than the *PSNR* of the corresponding views extracted from the reference frame. The *SSIM* shows similar results confirming the higher quality of the views.

Table 1 shows the Bjontegaard PSNR difference (BD-PSNR column) and bitrate reduction (Rate column). The average PSNR difference is 3.98*dB*, while the average bitrate reduction is −67%.

Table 1: BD-PSNR gain and bitrate reduction of the base layer with respect to the reference algorithm.

Light Field Name	BD-PSNR	Rate
Bike	3.99	-68%
Danger de Mort	4.00	-66%
Flowers	3.92	-65%
Stone Pillars Outside	3.94	-69%
Fountain & Vincent 2	4.05	-67%
Friends 2	3.97	-65%

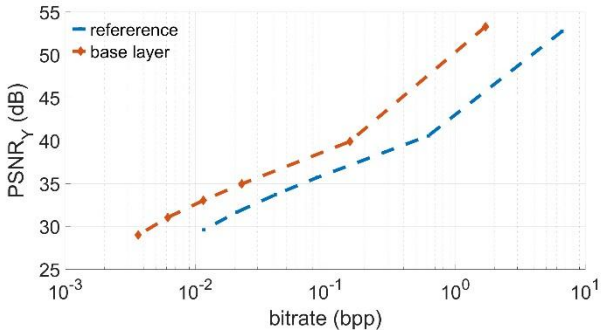


Figure 7: Rate-distortion curves showing $PSNR_Y$(dB) vs *bitrate* (bpp) for light field *Bikes*.

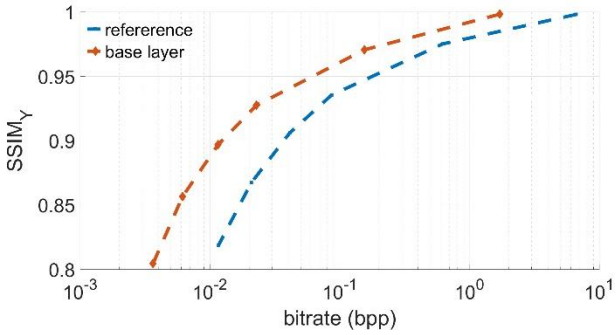


Figure 8: Rate-distortion curves showing $SSIM_Y$(dB) vs *bitrate* (bpp) for light field *Bikes*.

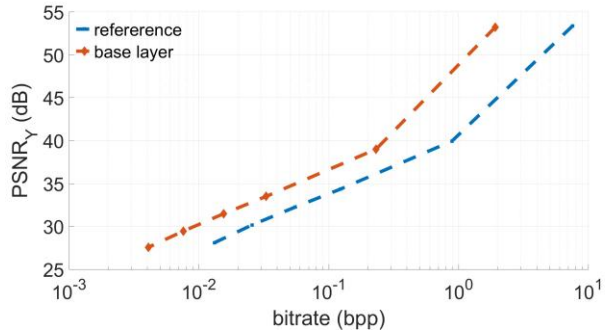


Figure 9: Rate-distortion curves showing $PSNR_Y$(dB) vs *bitrate* (bpp) for light field *Danger de Mort*.

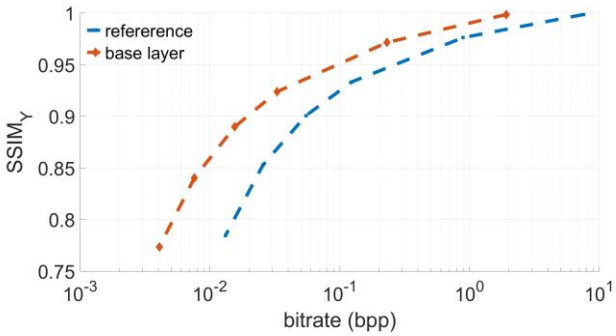


Figure 10: Rate-distortion curves showing $SSIM_Y$(dB) vs *bitrate* (bpp) for light field *Danger de Mort*.

CONCLUSIONS

This paper proposed an architecture for organizing light field information in group of views that can be flexibly organized for supporting light field applications requiring a subset of the light field information for rendering a given scene. The group of views are encoded using HEVC standard and are stored and made available for extraction upon request from a user device to a light field server. The proposed architecture has been experimented considering a use case where a base layer containing a set of a spiral-ordered views is needed by the user device. Obtained results have shown that high *PSNR* (40*dB*) and *SSIM* (0.95) are achieved at low bit-rates (0.6*bpp*).

ACKNOWLEDGMENT

This research activity has been partially funded within the CagliariPort2020 project (MIUR, SCN 00281) and Cagliari2020 project (MIUR, PON04a2 00381).

REFERENCES

[1] P. A. A. Assuncao, "Evolution from 3D video to light-field coding and transmission over future media networks," in *2016 18th Mediterranean Electrotechnical Conference (MELECON)*, April 2016.

[2] C. Perra and D. Giusto, "A framework for view progressive coding of light field," *IEEE International Conference of Consumer Elecronics*, January 2018.

[3] C. Perra and D. Giusto, "Light field compression on sliced lenslet array," *International Journal of Internet Technology and Secured Transactions*, 2017.

[4] C. Perra and D. Giusto, "Raw light field image compression of sliced lenslet array," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2017.

[5] C. Perra and D. Giusto, "JPEG 2000 compression of unfocused light field images based on lenslet array slicing," in *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 2017, pp. 27–28.

[6] F. Murgia, S.-B. Fernandez, C. Perra, and D. Giusto, "Unfocused plenop-´ tic camera calibration," in *22nd Telecommunications forum*, Belgrade, Serbia, November 2014.

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

[8] 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.

[9] F. Murgia, D. Giusto, and C. Perra, "3D reconstruction from plenoptic image," in *Telecommunications Forum Telfor (TELFOR), 2015 23rd*, November 2015, pp. 448–451.

[10] N. Gehrig and P. L. Dragotti, "Geometry-driven distributed compression of the plenoptic function: Performance bounds and constructive algorithms," *Image Processing, IEEE Transactions on*, vol. 18, no. 3, pp. 457–470, 2009.

[11] 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*, November 2015.

[12] C. Perra, "On the coding of plenoptic raw images," in *22nd Telecommunications forum*, Belgrade, Serbia, November 2014.

[13] C. Perra, "Lossless plenoptic image compression using adaptive block differential prediction," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing*, April 2015.

[14] 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.

[15] Y. Li, M. Sjostr¨ om, 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, pp. 80–91, 2016.

[16] 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.

[17] C. Perra, "Light field image compression based on preprocessing and high efficiency coding," in *2016 24th Telecommunications Forum (TELFOR)*, November 2016.

[18] C. Perra, "On the quality evaluation of lossy compressed light fields," in *2016 24th Telecommunications Forum (TELFOR)*, November 2016.

[19] M. Magnor and B. Girod, "Model-based coding of multiviewpoint imagery," in *Visual Communications and Image Processing 2000*. International Society for Optics and Photonics, 2000, pp. 14–22.

[20] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field hevc-based image coding using locally linear embedding and self-similarity compensated prediction," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[21] C. Conti, P. Nunes, and L. D. Soares, "Hevc-based light field image coding with bi-predicted self-similarity compensation," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[22] C. Conti, L. D. Soares, and P. Nunes, "Improved inter-layer prediction for light field content coding with display scalability," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2016, pp. 99710P–99710P.

[23] Y. Li, R. Olsson, and M. Sjostrom, "Compression of unfocused plenoptic¨ images using a displacement intra prediction," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[24] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, "Coding of focused¨ plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1308– 1319, July 2016.

[25] R. Ng, M. Levoy, M. Bredif, 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.

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

[27] M. Rerabek, L. Yuan, L. A. Authier, and T. Ebrahimi, "EPFL Light-Field image dataset," ISO/IEC JTC 1/SC 29/WG1, Tech. Rep., 2015.