LIGHT FIELD COMPRESSION USING FOURIER DISPARITY LAYERS

Elian Dib[+]*, Mikaël Le Pendu*[*]*, Christine Guillemot*[+]

+: Inria, Campus Universitaire de Beaulieu, 35042 Rennes, France
*: V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Ireland

ABSTRACT

In this paper, we present a compression method for light fields based on the Fourier Disparity Layer representation. This light field representation consists in a set of layers that can be efficiently constructed in the Fourier domain from a sparse set of views, and then used to reconstruct intermediate viewpoints without requiring a disparity map. In the proposed compression scheme, a subset of light field views is encoded first and used to construct a Fourier Disparity Layer model from which a second subset of views is predicted. After encoding and decoding the residual of those predicted views, a larger set of decoded views is available, allowing us to refine the layer model in order to predict the next views with increased accuracy. The procedure is repeated until the complete set of light field views is encoded. Following this principle, we investigate in the paper different scanning orders of the light field views and analyse their respective efficiencies regarding the compression performance.

1. INTRODUCTION

Light field imaging is gaining in popularity for a variety of vision applications, thanks to the emergence of real light fields capturing devices, and commercially available cameras. However, light fields represent very large volumes of high dimensional data, hence the need for designing efficient compression algorithms. Many solutions proposed so far adapt standardized image and video compression solutions (in particular HEVC) to light field data (*e.g.* [1] [2] [3]). The authors in [4], [5] investigate the use of homography-based low rank models for reducing the angular dimension, while local Gaussian mixture models in the 4D ray space are considered in [6]. A depth based segmentation of the light field into 4D spatio-angular blocks is used in [7] for prediction, and the prediction residue is encoded using JPEG-2000.

In this paper, we describe a novel compression algorithm based on the Fourier Disparity Layer representation (FDL) introduced in [8]. This representation has been shown to be efficient for a variety of processing applications, namely rendering, view synthesis and denoising. The proposed compression approach iteratively reconstructs the light field using the FDL representation. More precisely, an initial subset of views is first encoded with traditional video compression methods, and is used to construct a FDL representation of the light field. The FDL representation is then used to synthesize a second subset of light field views (neighboring ones) according to a pre-defined order. Given that there may remain correlations between the prediction residue of these synthesized views, the residual signal is best encoded using a video encoder with inter coding (in the experiments we used HEVC-inter, and in particular the HM 16.10). Then, in order to predict and encode the next subset of views, a more accurate FDL representation is constructed from the previously encoded and decoded subsets. The FDL representation is thus iteratively refined after the encoding of each view subset until all the views are encoded. Two types of scanning orders are considered for synthesis and coding, one following a hierarchical scheme and the other one following a circular order. In both cases, prediction schemes with either two or four view subsets were tested. For the best performing schemes, experimental results with real and synthetic light fields show average rate savings of more than 50%, using the Bjontegaard measure, with respect to the JPEG-Pleno verification model (VM 1.1)[9].

This work was supported in part by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM) and in part by the Science Founda-tion Ireland (SFI) under the Grant Number 15/RP/2776.

2. NOTATIONS AND SCHEME OVERVIEW

Let us consider the 4D representation of light fields proposed in [10] and [11] describing the radiance along rays by a function $L(x, y, u, v)$. This representation is based on a parameterization of orientations of light rays with two parallel planes with pairs (x, y) and (u, v) respectively representing spatial and angular coordinates of light rays.

The coding algorithm is depicted in Fig.1. First, the Fourier Disparity Layer calibration described in [8] determines a set of disparity values as well as the angular coordinates of each view. These parameters are needed for the FDL construction and view prediction steps, and are transmitted as metadata to the decoder. Our coding scheme partitions the light field into subsets of views, the first subset being directly encoded as a group of pictures using a video encoder. This first set of views is used for an initial construction of the FDL representation which allows us to synthesize (or predict) the views of the second subset. The prediction residue is coded, decoded and added to the prediction to reconstruct the corresponding views. The views of the two first subsets are then used to re-compute the FDL representation that will then be used for synthesizing the views of the third subset. The algorithm continues iterating until all the light field views have been coded.

3. FOURIER DISPARITY LAYERS (FDL)

For simplicity of notation, let us consider only one 2D slice of the light field with only one spatial and one angular dimension. A view L_{u_0} of the light field at angular coordinate u_0 is then defined

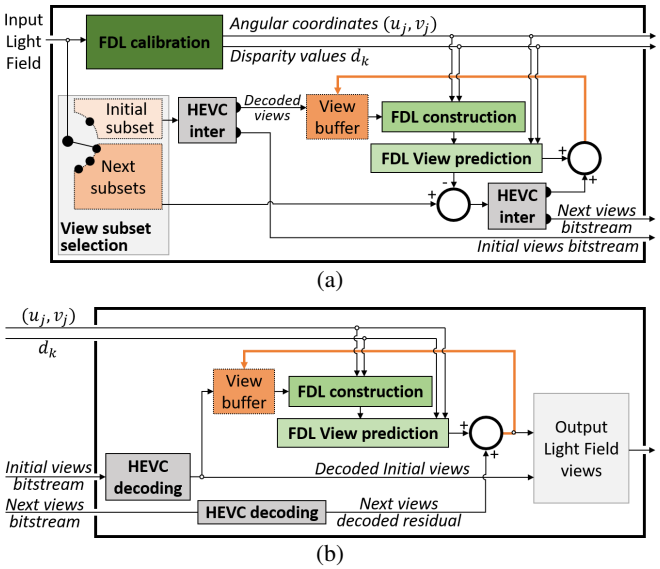


Fig. 1. Overview of the encoding (a) and decoding (b) scheme.

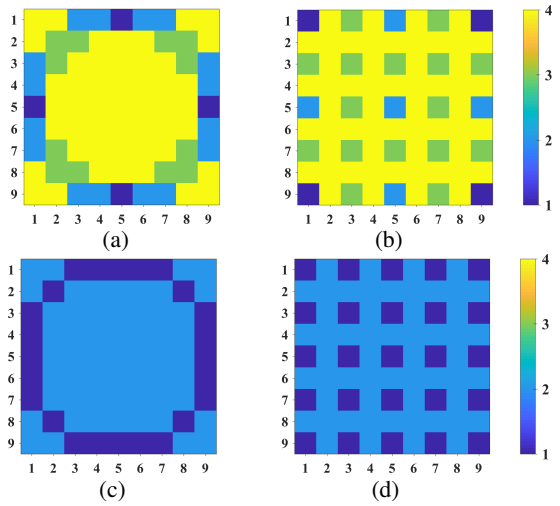


Fig. 2. View subsets for different view prediction orders: (a) Circular-4, (b) Hierarchical-4, (c) Circular-2, (d) Hierarchical-2.

by $L_{u_0}(x) = L(x, u_0)$. It has been shown in [8] that, given a set of n disparity values $\{d_k\}_{k \in [\![1,n]\!]}$, the Fourier Transform $\hat{L}_{u_0}$ of L_{u_0} can be decomposed as:

$$\hat{L}_{u_0}(\omega_x) = \sum_k e^{+2i\pi u_0 d_k \omega_x} \hat{L}^k(\omega_x). \quad (1)$$

where ω_x denotes the spatial frequency, and $\hat{L}^k$ is defined by

$$\hat{L}^k(\omega_x) = \int_{\Omega_k} e^{-2i\pi x \omega_x} L(x, 0) \mathrm{d}x. \quad (2)$$

Each function $\hat{L}^k$ can be interpreted as the Fourier transform of the central view only considering a spatial region Ω_k of disparity d_k, hence the name Fourier Disparity Layers (FDL).

Given m input views L_{u_j} ($j \in [\![1, m]\!]$) and by computing their Fourier Transforms $\hat{L}_{u_j}$, the FDL representation can be learned by solving a linear regression problem for each frequency coefficient ω_x. Constructing the FDL amounts to solving the equation $\mathbf{Ax} = \mathbf{b}$, with a Tikhonov regularization, where $\mathbf{A}$, $\mathbf{x}$, $\mathbf{b}$ are matrices and vectors of dimensions $m \times n$, $n \times 1$ and $m \times 1$ respectively. The matrix $\mathbf{A}$ is defined as $\mathbf{A}_{jk} = e^{+2i\pi u_j d_k \omega_x}$, while the vector $\mathbf{x}$ containing the Fourier coefficients of the disparity layers (for the frequency ω_x) is defined as $\mathbf{x}_k = \hat{L}^k(\omega_x)$, and $\mathbf{b}$ containing the Fourier coefficients of the input image j is defined as $\mathbf{b}_j = \hat{L}_{u_j}(\omega_x)$.

4. VIEW SYNTHESIS USING FDL

Knowing the layers and their disparities d_k, any view L_{u_0}, at angular position u_0, can be synthesized in the Fourier domain using Eq.(1), and by computing the inverse Fourier transform. One important issue that can have a strong impact on the quality of the synthesized views used for prediction, is the view synthesis (or prediction) order. Typically, views that are close to the input views of the FDL construction step are better synthesized than more distant views. Furthermore, it was observed in [8] that constructing a FDL from all the outer views at the periphery of the light field is an ideal configuration for synthesizing the inner views.

The circular prediction order illustrated in Fig. 2(a) was designed following those insights. To give an example with a light field of dimension 9×9, the circular scheme starts by encoding the views in the middle of each segment forming the periphery (in dark blue in Fig. 2(a)). The four adjacent views (light blue) are then synthesized using the FDL constructed from the base views. The prediction residue for these 16 views is then coded with HEVC inter, decoded and added to the synthesised views. The reconstructed views at the corresponding positions will then be used together with the base views to refine the FDL representation. The algorithm iterates considering next the 12 views "closing" the circle (green in Fig. 2(a)). Note that we preferred forming a circle rather than a square in order to avoid using the corner views for predicting the inner views. This choice is motivated by the fact that, in real light field datasets captured by plenoptic cameras, the corner views are generally of lower quality. A simplified circular prediction scheme was also designed as shown in Fig. 2(c), where only two subsets are used, and all the views forming the circle are contained in the initial subset.

For the comparison, we have also designed and tested more traditional hierarchical configurations as shown in Fig. 2(b) and (d) respectively with four and two subsets. The prediction schemes in Fig. 2(a),(b),(c), and (d) are referred to as Circular-4, Hierarchical-4, Circular-2, and Hierarchical-2 respectively.

5. CODING SCHEME

5.1. FDL Calibration

To construct the FDL, one first needs to estimate precise angular coordinates u_j of input views as well as the layers' disparity values d_k. The joint estimation of these two sets of parameters allows the method to be robust to common issues with real light field data, e.g. to the fact that the angular coordinates of all the views may not form precisely a square grid as it is generally assumed. These parameters u_j and d_k are found by minimizing over the Q frequency components ω_x^q (Q being the number of pixels in each input image) [8]

$$\min_{\mathbf{x}^q,\mathbf{u},\mathbf{d}} \sum_{q=1}^{Q} \left(\left\| \mathbf{A}(\omega_x^q, \mathbf{u}, \mathbf{d})\mathbf{x}^q - \mathbf{b}^q \right\|_2^2 + \lambda \left\| \mathbf{\Gamma}\mathbf{x}^q \right\|_2^2 \right), \quad (3)$$

where the input view positions u_j and the disparity values d_k are arranged in the vectors $\mathbf{u}$ and $\mathbf{d}$ respectively, and where $[\mathbf{A}(\omega_x^q, \mathbf{u}, \mathbf{d})]_{j,k} = e^{+2i\pi u_j d_k \omega_x}$. The vectors $\mathbf{x}^q$ and $\mathbf{b}^q$ contain the Fourier coefficients of, respectively, the disparity layers and the input images at the frequency ω_x^q (i.e. $\mathbf{x}_k^q = \hat{L}^k(\omega_x^q)$ and $\mathbf{b}_j^q = \hat{B}_j(\omega_x^q)$). The regularization matrix $\mathbf{\Gamma}$ is defined as a discrete approximation of the second order differential operator:

$$\mathbf{\Gamma} = \begin{bmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -2 & \end{bmatrix}. \quad (4)$$

Note that the calibration results depend on the regularization parameter λ. We optimize this parameter by performing several calibrations with different values of λ. Given a prediction order in our coding scheme (see Fig. 2), we keep the calibration results yielding the best predictions in the least square sense. For this step, the prediction of a subset of views does not take into account the compression loss of the previous subsets, in order to avoid performing the complete compression algorithm only to determine λ. Therefore, the disparity values d_k and view positions u_j are learnt for optimal predictions within our coding scheme.

5.2. Prediction Scaling

Additionally, in order to cope with possible variations of illumination between views, we perform a scaling of each predicted view. Let the image I be one of the views, and I_{pr} its prediction obtained using the FDL view synthesis. The final prediction $\widetilde{I_{pr}}$ is then computed as $\widetilde{I_{pr}} = s \cdot I_{pr}$, where the scaling factor s is derived by the encoder to minimize the sum of squared error of the scaled prediction as follows:

$$\arg\min_s \|I - s \cdot I_{pr}\|_2^2 = \frac{\langle I, I_{pr} \rangle}{\|I_{pr}\|_2^2}, \quad (5)$$

where $\langle \cdot, \cdot \rangle$ is the inner product. The value of s is transmitted as metadata in order to perform the same prediction scaling on the decoder side.

5.3. Encoding of metadata

All the transmitted parameters (i.e. scaling factors, set of disparity values and angular coordinates), are encoded in the double precision floating point format with a fixed length encoding, resulting in 64 bits for each parameter. The additional cost is accounted for in the results presented in Section 6.

5.4. Encoding of initial views and view residuals

The initial set of views is directly encoded as a video sequence using HEVC 8 bits with inter coding. For the other views, the prediction residual is also encoded in HEVC. Since the prediction residual requires one additional bit of precision compared to the original signal, we avoid the risk of precision loss by using HEVC 10 bits. Furthermore, due to imperfections either in the original signal or in the FDL view synthesis, correlations may remain between the prediction residual of the different predicted views. Therefore, at each iteration, the current subset of views to encode is arranged in a sequence and encoded using HEVC-inter. For both the circular and the hierarchical prediction schemes, the views are arranged in a spiral order starting from the center of the light field, and only considering the views of the current subset.

In order to optimize the bitrate allocation, we use different QP parameters in the HEVC encoding of the different subsets of views. Let QP_1 be the base parameter used for the initial subset of views, then we use $QP_t = QP_{t-1} + 1$ as the parameter for the subset of index t in the coding order.

6. EXPERIMENTAL RESULTS

6.1. Experimental Setting

Experiments were performed using the luminance component of light fields coming from the HCI [12], INRIA [13] and ICME 2016 Grand Challenge [14] datasets. The HCI dataset contains synthetic light fields with 9x9 views of 768x768 pixels (Buddha, Butterfly, StillLife) and with 9x9 views of 512x512 pixels (Greek, Sideboard), the INRIA and ICME datasets originally contain light fields captured by a Lytro Illum camera with 15x15 views of 625x434 pixels from which we keep the 9x9 central views (to avoid strong vignetting issues) of 616x424 pixels (to remove the black borders and ensure the size is a multiple of 8 for the HEVC coding). The Lytro LFs have been extracted using the Matlab Light Field Toolbox v0.4 [15] with gamma correction.

The performances of the proposed FDL coding schemes have been compared to JPEG Pleno VM 1.1 [9] and to the HEVC coding of all the views arranged in a spiral order. The latter method can be seen as a particular case of our approach where all the views are in the initial set. It is therefore referred to as the Circular-1 scheme. For Circular-1 and our FDL coding schemes, the HEVC encoding is performed using the HM 16.10 with the Main-RExt profile and the random access configuration. We have used a range of quality parameters $QP \in \{5, 10, 15, 20, 30, 40\}$ (which corresponds to the QP parameter of the initial view subset in our approach). For our FDL coding schemes, we have fixed the number of layers in the FDL method to $n = 30$. Although the random access configuration was used for HEVC coding (intra period of 32), the random access capability of our approach is limited by the fact that the views of the last subset are dependant on those of the previous subsets. In practice, for the four variants in Fig. 2, the percentage of the bitstream required to decode one view in the worst case is about 80% at high bitrates ($QP = 40$) and about 95% at low bitrates ($QP = 5$).

For JPEG Pleno, the required input central disparity maps have been generated using [16] and the configuration files provided with JPEG Pleno VM 1.1 have been adapted for use on 9x9 light fields.

6.2. Results

The rate-distortion results are presented in Fig. 3 for two real and two synthetic light fields. For all the tested light fields, the bitrate savings of the different variants of our coding scheme, as well as the direct HEVC encoding (i.e. 'C1') are presented in Table 1. They have been computed with the Bjontegaard metric [17] using JPEG-Pleno as a reference.

We observe in Table 1 that the Circular-2 (respectively Circular-4) scheme systematically outperform Hierarchical-2 (respectively Hierarchical-4). This result supports the idea that using a set of connected peripheral views to construct the FDL rather

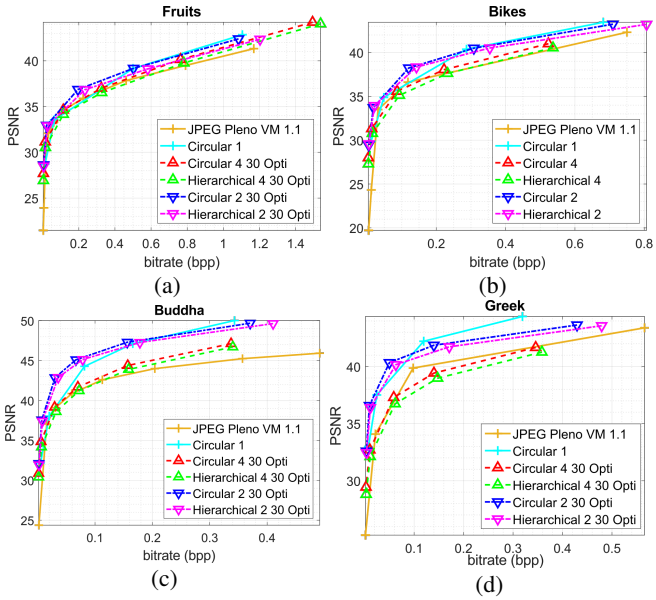


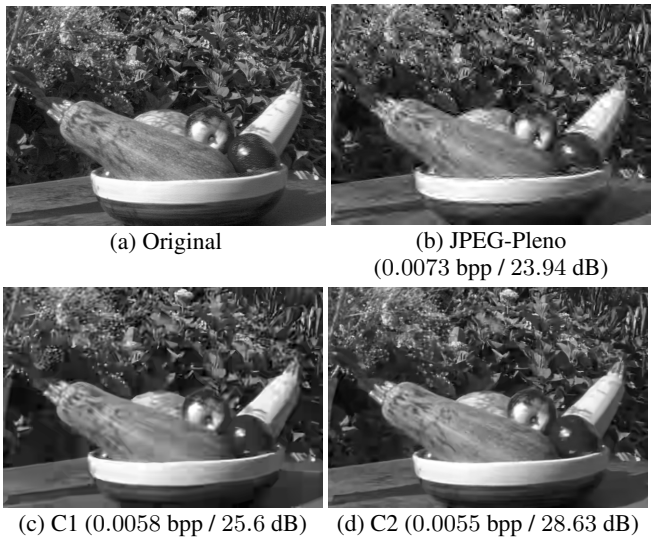
Fig. 3. PSNR-Rate performance on real (top) and synthetic (bottom) light fields from (a) INRIA, (b) ICME and (c,d) HCI datasets.

Method	H2	C2	H4	C4	C1
Bench	-25.53	-34.3	2.76	-15.05	-10.29
Fruits	-35.54	-46.02	-4.05	-22.12	-11.67
Toys	-48.17	-51.2	-4.56	-16.24	-52.67
Bikes	-41.18	-47.31	1	-15.82	-20.05
D_de_M	-29.78	-40.42	-3.62	-21.61	1.54
F_&_V_2	-45.89	-48.64	-8.18	-27.44	-28.37
Friends_1	-46.41	-48.75	-21.5	-37.39	-21.41
S_P_I	-28.25	-33.22	41.08	7.19	-41.06
Vespa	-43.87	-47.55	21.07	-7.49	-23.45
Greek	-58.44	-67.31	20.47	-2.14	-52.24
Sideboard	-56.31	-64.14	54.27	27.09	-13.63
Buddha	-62.69	-68.21	-8.19	-26.97	-31.73
Butterfly	-86.03	-86.96	-72.6	-75.16	-75.6
StillLife	-49.8	-60.99	44.85	16.51	-60.11
Average	-50.11	-56.5	-1.54	-20.42	-35.42

Table 1. Bjontegaard percentage rate savings for the Hierarchical-2 (H2), Circular-2 (C2), Hierarchical-4 (H4), Circular-4 (C4), and Circular-1 (C1) schemes with respect to JPEG Pleno VM 1.1 (negative values represent gains).

than a more scattered set substantially improves the view prediction. Additionally, the variants of our coding scheme with only two subsets (i.e. Circular-2, Hierarchical-2) have significantly better performances than the ones with 4 subsets (i.e. Circular-4, Hierarchical-4). This may be due to high frequency artifacts in the predicted views caused by the FDL construction from a too sparse set of known views.

Compared to JPEG-Pleno, more than 50% rate gains are obtained on average for our variants with two subsets, This is a substantial improvement over the average 35% gain of the Circular-1 method which directly encodes all the views in HEVC with the same configuration. It can also be observed in Figs. 3 and 4 that particularly large PSNR gains are obtained for very low bitrates, especially compared to JPEG-Pleno. This can be explained by



(a) Original (b) JPEG-Pleno (0.0073 bpp / 23.94 dB)

(c) C1 (0.0058 bpp / 25.6 dB) (d) C2 (0.0055 bpp / 28.63 dB)

Fig. 4. Visual comparison of the reconstructed top left view image for JPEG Pleno VM 1.1, Circular-1 (C1) and Circular-2 (C2).

the very limited amount of additional data needed to perform our predictions. On the other hand, JPEG-Pleno requires the transmission of a complete disparity map, which results in a significant overhead for low bitrate coding.

Note finally that another advantage of our approach is to allow scalable light field coding. Since the FDL model can be used to synthesize any view of the light field, a complete light field can be reconstructed by the decoder at each iteration. By decoding additional view subsets, a more accurate light field is obtained. In that sense, despite their lower performance, the schemes with 4 iterations (i.e. Circular-4, Hierarchical-4) may be preferable in some scenarios as they provide additional levels of scalability. Furthermore, Circular-4 still clearly outperforms the JPEG-Pleno anchor in most cases (20% bitrate savings on average).

7. CONCLUSION

We have presented a novel light field compression method based on the Fourier Disparity Layer representation. In the proposed scheme, the light field is partitioned into several subsets of views, where the first subset is encoded as a video sequence using HEVC. The next subsets are iteratively predicted from the previously encoded and decoded ones thanks to the Fourier Disparity Layer view synthesis, and only the prediction residual is encoded with HEVC. We have shown that defining view subsets forming a circular pattern is better suited to this prediction scheme than the more conventional hierarchical coding order. In the tested configurations with only two subsets of views, significant bitrate savings were obtained in comparison to both JPEG-Pleno and a direct encoding of all the views with HEVC. Using more subsets with our circular coding order still outperforms JPEG-Pleno, but reduces the performance compared to the case of two subsets. However, it provides additional levels of scalability since the FDL model refined with each decoded subset can be used to synthesize any view of the light field.

8. REFERENCES

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

[2] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*, Jul. 2016.

[3] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*. IEEE, 2016, pp. 1–4.

[4] X. Jiang, M. Le Pendu, R. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1132–1145, Oct. 2017.

[5] E. Dib, M. Le Pendu, X. Jiang, and C. Guillemot, "Super-ray based low rank approximation for light field compression," in *Data Compression Conference (DCC)*, Mar. 2019.

[6] R. Verhack, T. Sikora, L. Lange, R. Jongebloed, G. Van Wallendael, and P. Lambert, "Steered mixture-of-experts for light field coding, depth estimation, and processing," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*, 2017, pp. 1183–1188.

[7] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and jpeg 2000," in *IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2017, pp. 4567–4571.

[8] M. Le Pendu, C. Guillemot, and A. Smolic, "A fourier disparity layer representation for light fields," *arXiv:1901.06919 (under review in IEEE Trans. Image Proc.)*, 2019.

[9] ISO/IEC JTC 1/SC29/WG1 JPEG, "Jpeg pleno light field coding vm 1.1," Doc. N81052, 2018.

[10] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 1996, SIGGRAPH '96, pp. 31–42, ACM.

[11] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proc. SIGGRAPH*, 1996, pp. 43–54.

[12] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *VMV Workshop*, 2013, pp. 225–226.

[13] "INRIA Lytro image dataset," `https://www.irisa.fr/temics/demos/lightField/LowRank2/datasets/datasets.html`.

[14] "ICME 2016 Grand Challenge dataset," `http://mmspg.epfl.ch/EPFL-light-field-image-dataset`.

[15] D.G. Dansereau, "Light Field Toolbox for Matlab," 2015.

[16] X. Jiang, M. Le Pendu, and C. Guillemot, "Depth estimation with occlusion handling from a sparse set of light field views," in *IEEE Int. Conf. Image Process. (ICIP)*, 2018.

[17] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," *document VCEG-M33, ITU-T VCEG Meeting*, 2001.