EPIC: Context Adaptive Lossless Light Field Compression using Epipolar Plane Images

Mukati, Muhammad Umair; Forchhammer, Søren

Published in:
Proceedings of 2020 Data Compression Conference

Link to article, DOI:
10.1109/DCC47342.2020.00012

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

2020 Data Compression Conference (DCC)

EPIC: Context Adaptive Lossless Light Field Compression using Epipolar Plane Images

M. Umair Mukati, Søren Forchhammer

DTU Fotonik, Technical University of Denmark
Ørsteds Plads, Kgs. Lyngby - 2800, Denmark
Emails: {mummu, sofo} @fotonik.dtu.dk

Abstract

This paper proposes extensions of CALIC for lossless compression of light field (LF) images. The overall prediction process is improved by exploiting the linear structure of Epipolar Plane Images (EPI) in a slope based prediction scheme. The prediction is improved further by averaging predictions made using horizontal and verticals EPIs. Besides this, the difference in these predictions is included in the error energy function, and the texture context is redefined to improve the overall compression ratio. The results using the proposed method shows significant bitrate-savings in comparison to standard lossless coding schemes and offers significant reduction in computational complexity in comparison to the state-of-the-art compression schemes.

1 Introduction

Light field cameras offer off-the-shelf post capture adjustment of optical parameters, such as, refocusing, view-point shifting and aperture adjustment upto some extent, in contrast to traditional cameras [1]. This ability comes at the cost of capturing and storing huge amounts of information in raw format. For instance, a typical lenslet based LF image requires 174.6 MB[1] of hard-disk space after being decoded. Therefore, compression and efficient transmission are some of the important research trends in the LF domain.

A lenslet based LF camera captures raw lenslet 2-D images. After rectification and decoding, the captured data are converted to a 4-D LF. Two of these dimensions represent angular coordinates while the other two defines the spatial coordinates of imaging system. Pairing them in different combinations, multiple LF formats can be achieved, such as, sub-aperture image (SAI), micro-image (MI) and epipolar plane image (EPI). Each of them have been used in the literature to improve the compression efficiency. Although most of the research work is based on lossy compression schemes, for lossless compression some featured techniques are listed here. Using an MI array (raw lenslet) LF format, C. Perra [2] was among one of the first to propose lossless compression of LF. Helin *et. al.* [3] segmented each SAI based on their disparities and then used sparse predictors to decorrelate neighboring views. In [4], Schiopu *et. al.* encoded the residual from an efficient adaptive predictor using a context based

[1]For an LF image captured by the Lytro Illum camera, considering, spatial resolution: 434×625, angular resolution: 15×15, three color channels and 8 bit depth.

2375-0359/20/$31.00 ©2020 IEEE
DOI 10.1109/DCC47342.2020.00012

entropy encoder. The SAIs are segmented into regions of similar pixels and defined as a unique context for coding. In [5] and [6], Santos *et. al.* used minimum rate predictors to optimize the weights of the predictors using neighboring pixels from pseudo video sequence and multiple LF formats (SAI and MI), respectively.

These techniques can achieve better compression performance compared to standard lossless image codecs, such as JPEG-LS, CALIC, AVC and HEVC. However, the compression schemes are highly complex and require much processing power and time to encode an LF image. In this paper, we propose a novel approach by modifying CALIC [7] to compress LF images. It is specifically designed to reduce the inherent redundancy in the LF structure by exploiting the EPI format. By establishing our method on CALIC's foundation, a significant reduction in computational complexity can be achieved.

2 Overview: Context Adaptive Lossless Image Compression (CALIC)

CALIC is a sequential lossless coding scheme for images [7], which encodes and decodes images in raster scan order. It operates in two modes, i.e., binary and continuous. Binary mode remains activated until the defined causal neighborhood has more than two unique values. In binary mode, three symbols are encoded using a ternary arithmetic encoder, that also includes an escape symbol, which signals the decoder to exit from the binary mode. In continuous mode of operation, apart from entropy coding, there are three important processing blocks, i.e., gradient-adjusted prediction (GAP), error energy estimation, and context modeling of prediction errors.

GAP predicts the intensity at a current pixel in two stages. In the first stage, it computes the boundaries (d_h and d_v) using the sum of horizontal and vertical differences of the intensities in the neighborhood. Conditioning on this information, it non-linearly predicts the intensity using the neighboring pixels. In the next stage, the scheme conditions the prediction errors on eight quantized error energy bands. By doing so, the entropy of the errors in the lower bins is significantly reduced, and with the help of adaptive arithmetic encoding, higher compression can be achieved. The error energy estimator Δ is defined as:

$$\Delta = d_h + d_v + 2|e_w|, \tag{1}$$

where e_w is the previous error. For computational efficiency, this estimate is quantized into eight bins whose edges are empirically chosen to minimize the total entropy of errors.

This error is additionally conditioned on 144 texture-contexts to further improve the coding efficiency. Due to the high number of contexts, instead of using them with adaptive arithmetic coding, the expectation of error within each context is maintained and the expected error from the corresponding context is subtracted from the error through a feedback mechanism. For better performance, a combined context is formed by combining 144 texture-contexts with 4 quantized error energy levels.

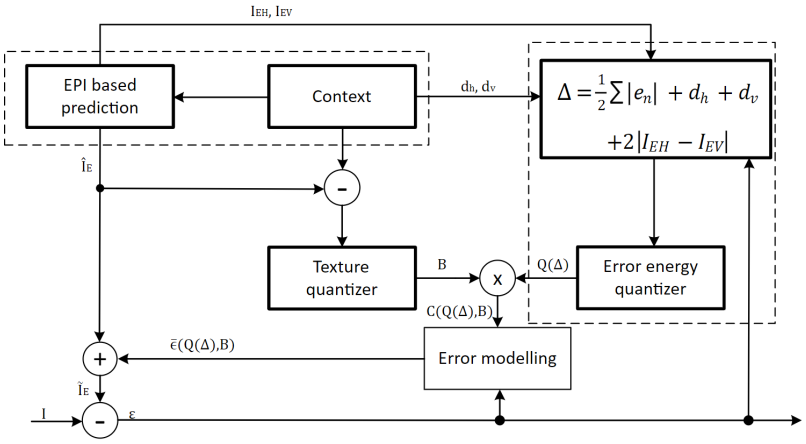

Figure 1: Block diagram with modification highlighted by blocks with bold borders.

3 Context Adaptive Compression of Epipolar Plane Images (EPIC)

A light field is a set of light rays passing through an LF camera surface. The radiances of these rays are recorded by LF camera, which can be indexed through several LF representations. With an assumption that light rays travel along a straight path and there is no loss of energy as its propagates, a 4-D parallel plane representation of LF $L(s,t,u,v)$ can be achieved, which proved to be simple in terms of processing and efficiency [8]. Out of these coordinates, u and v define the spatial location, where the ray intersects on the first plane, whereas, s and t represent the angular orientation of a ray in horizontal and vertical directions. An additional dimension is used to index the RGB color components in a 5-D representation.

EPI is an LF format, which is formed by fixing either s and u, the vertical indices of the two planes, or by fixing horizontal indices, t and v, termed as horizontal EPI (H-EPI) and vertical EPI (V-EPI), respectively. An example of an EPI image is shown in Figure 2. Considering the structure of EPI in LF images, the method described in this paper proposes three folds improvement to all the important blocks of CALIC, i.e, prediction, error energy estimation and context modeling of prediction errors. The block diagram in Figure 1 describes the working of continuous mode in CALIC and highlights the blocks which are modified in this proposal.

3.1 EPI Slope based Prediction (ESP)

A light field image contains highly correlated data, which makes it possible to achieve high compression ratios. Capturing this correlation using different LF formats can be complex and time consuming. However, due to the nature of an EPI (i.e., linear edges or smooth surfaces as shown in Figure 2), it is comparatively easier to capture

Figure 2: EPI created using "Bikes" LF from [9] by indexing 8^{th} row and 400^{th} row of angular and spatial coordinates, respectively.

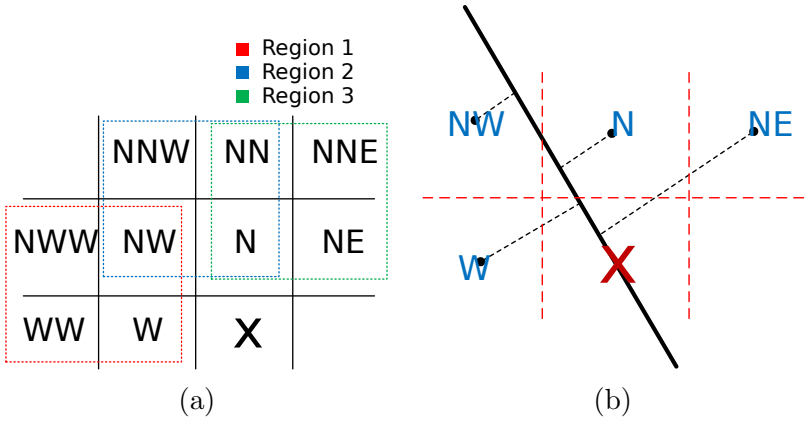

Figure 3: (a) Dotted boxes representing regions used for slope estimation; (b) Perpendicular distance of neighbors to the edge in EPI passing through the center of the current pixel.

the correlation through neighboring pixels and thus high quality predictions can be made. This LF format can improve the prediction performance using CALIC's GAP predictor but only upto a certain level, because it is designed for general prediction. Considering the EPI structure, we propose an EPI slope based predictor (ESP), which first estimates the slope of the edges in EPI regions and then predicts the intensity through quadratic interpolation.

Let the intensity of the current pixel of the LF being decoded is $LF(\bar{s}, \bar{t}, \bar{u}, \bar{v})$. Then, the corresponding H-EPI can be formed by:

$$I^H(t, v) = LF(s = \bar{s}, t, u = \bar{u}, v),$$

where t and v can be varied to access intensities of neighboring pixels in $I^H(t, v)$. As shown in Figure 3a, the causal neighborhood of the current pixel is first divided into three different regions. Then for each region, the gradient direction is computed using a 2×2 horizontal and vertical edge filters. To estimate the slope robustly, variance of each region is used as a weight in the weighted average of slopes.

Given that the intensity remains the same along the direction of the edge in EPI, we first estimate perpendicular distance of neighboring indices using point-to-line distance formula. These distances along with the intensity of the neighboring pixels are used to estimate intensity at the current pixel using 1-D quadratic interpolation. A small error in slope estimation, noise in the region and the proximity of distance can cause the estimation to overshoot; therefore, only a nearby set of neighbors of

$I^H(\bar{t}, \bar{v})$ are used to predict the intensity $\hat{I}_e^H$, as shown in Figure 3b. Through the least squares solution, the coefficients α, β, γ of a quadratic interpolator are estimated to minimize the following expression:

$$\arg\min_{\alpha,\beta,\gamma} \sum_{j \in J} \left(I_j^H - \left[\alpha d_j^2 + \beta d_j + \gamma \right] \right)^2,$$

where $J = \{N, W, NW, NE\}$, as shown in Figure 3b. The intensity $I^H(\bar{t}, \bar{v})$ to be predicted lies at the origin ($d = 0$) and γ represents the y-intercept of the quadratic curve, as well as the prediction $\hat{I}_e^H$. Therefore, for efficiency only γ is computed.

Since the neighborhood region can be smooth or can contain edges, the sum of horizontal (d_h) and vertical (d_v) boundaries is taken as an indicator and used to improve the prediction in the following way:

$$\hat{I}^H = \frac{c_e \hat{I}_e^H + c_s \hat{I}_s^H}{c_e + c_s}$$

$$\hat{I}_e^H = \gamma, c_e = d_h + d_v, \hat{I}_s^H = \frac{1}{4} \sum_j^J I_j, c_s = 6$$

As in CALIC, the boundaries d_h and d_v are calculated by taking the sum of horizontal and vertical differences of the neighboring pixels, respectively. The value for c_s is determined by minimizing the prediction error on a set of LF images in [9]. Similarly, the described method is applied to V-EPI, to predict $\hat{I}^V$. An improved prediction $\hat{I}$ is achieved by averaging these predictions, thus using a simple 4-D predictor. In comparison to GAP, ESP achieved a 22% reduction in the prediction error, when tested on the same set of LF images.

3.2 Error energy estimation

Although the improvement in the prediction step significantly reduces the entropy of the prediction error, yet by conditioning this error on a function, which closely correlates to the error energy, can further improve compression ratio. CALIC relies on the variance in the neighborhood and the previous prediction error to define an error energy function in Eq. 1. Being inspired by the on-line residual estimation method in [10], we use the difference in predictions (i.e. $|\hat{I}^H - \hat{I}^V|$) as an additional component in the error energy function, such that the overall error energy function becomes:

$$\Delta = d_h + d_v + \frac{1}{2} \sum_k |e_k| + 2|\hat{I}^H - \hat{I}^V|$$

Apart from the difference in prediction, the sum of prediction errors of the k neighbors of current pixel is also utilized, where $k \in \{\{N, W\} \in \text{H-EPI}, \{N, W\} \in \text{V-EPI}\}$. The estimator Δ is quantized to eight bins. The bin edges are optimized

to minimize overall conditional entropy of errors, on a set of EPIs extracted from different LFs.

$$q_1 = 6, q_2 = 20, q_3 = 32, q_4 = 50, q_5 = 80, q_6 = 110, q_7 = 170$$

3.3 Texture context and quantization

Context modeling of prediction errors can capture higher order structure in the neighborhood, which can help in further minimization of conditional entropy. Due to the large number of possible contexts, CALIC computes the expectation of prediction errors within each context and through a feedback mechanism removes it from the prediction error corresponding to its context. The texture context is defined by the intensities of the local neighborhood:

$$C = \{x_0, \ldots, x_6, x_7\}$$

$$C = \{I_N, I_W, I_{NW}, I_{NE}, I_{NN}, I_{WW}, 2I_N - I_{NN}, 2I_W - I_{WW}\}$$

This context is quantized to an 8-bit binary number $B = b_7 b_6 \ldots b_0$ using the prediction.

$$b_k = \begin{cases} 0 & \text{if} x \geq \hat{I}[i,j] \\ 1 & otherwise \end{cases} \tag{2}$$

After applying the threshold, this context becomes sensitive to random noise due to high sensitivity of the pixels and small surface area of micro-lenses in lenslet based LF cameras. The slope of the edge in EPI is dependent on the disparity of an object. Therefore, given a current pixel, the slope of the edge remains the same in a close neighborhood of the current pixel in H-EPI and V-EPI, respectively. Thus after averaging the intensities of corresponding neighboring pixels, the texture will remain the same and the sensitivity to noise can be reduced, as shown in Figure 4. Therefore, we modify the context such that:

$$C = \{I'_N, I'_W, I'_{NW}, I'_{NE}, I'_{NN}, I'_{WW}, 2I'_N - I'_{NN}, 2I'_W - I'_{WW}\},$$

where, I'_x is the average of intensities of corresponding neighbor x in I^H and I^V.

4 Experimental performance analysis

To analyze the performance of the proposed scheme quantitatively, lenslet based LF images are used in this paper. Two datasets are used for evaluation.

4.1 Test dataset preparation

Dataset-1 consists of four LF images from the JPEG-Pleno test set [11], i.e., *(Bikes, Danger_de_Mort, Fountain_and_Vincent_2, Stone_Pillars_Outside)*, which are a part

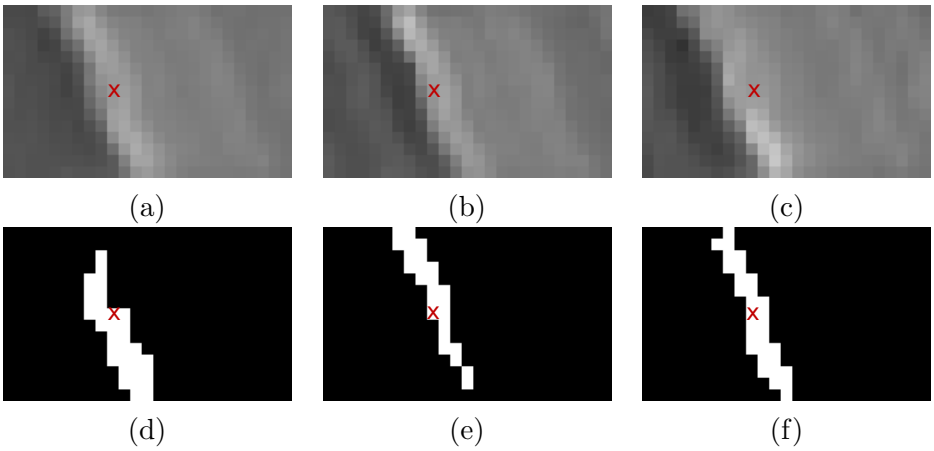

Figure 4: Representing a patch of EPI around current pixel 'x' (represented with a red cross sign in the image) in (a) horizontal EPI, (b) vertical EPI and (c) average of these patches. Binary context around 'x' generated using Eq. 2 for corresponding (d) horizontal EPI, (e) vertical EPI and (f) average of these patches.

Table 1: [**Dataset-1**] Average compression rate measured in bits-per-pixel calculated for JPEG-Pleno's (EPFL) dataset (in 24-bit $YUV444$ format) using different lossless compression schemes.

Light Fields	HEVC	AVC	JPEG-LS	CALIC	EPIC
Bikes	8.80	8.49	9.04	8.80	**7.98**
Danger de Mort	9.54	9.21	9.70	9.38	**8.71**
Fountain and Vincent 2	9.30	9.01	9.67	9.44	**8.51**
Stone Pillars Outside	8.59	8.26	8.95	8.71	**7.94**
Average (bpp)	9.05	8.74	9.34	9.08	**8.28**

of the EPFL light field dataset [9], are used in the experiments. These LF images are pre-processed as described in [12], followed by a contrast adjustment through gamma correction. **Dataset-2** composed of 12 LF images are taken directly from [9], which differs in the sense that their contrast is not adjusted. The analysis on this dataset is provided to serve as a comparison with the prior literature for LF compression. The LF images in both the datasets have an angular resolution of 15×15 and a spatial resolution of 434×625. The actual light fields are 10-bit RGB images which are converted to 8-bit $YUV444$ representation for our testing using the guideline in JPEG-Pleno's common testing condition criteria [13].

4.2 Encoder configuration

The proposed method is compared with four state-of-the-art lossless image and video compression techniques, i.e., HEVC, AVC, JPEG-LS and CALIC. To generate results

Table 2: **[Dataset-2]** Average compression rate measured in bits-per-pixel calculated for EPFL light field dataset (in 24-bit $YUV444$ format) using different lossless compression schemes.

Light Fields	HEVC	AVC	JPEG-LS	CALIC	EPIC
Bikes	7.82	7.69	8.11	7.86	**7.17**
Danger_de_Mort	7.27	7.17	7.53	7.30	**6.77**
Flowers	7.38	7.28	7.76	7.51	**6.87**
Stone_Pillars_Outside	7.54	7.34	8.02	7.73	**7.11**
Vespa	7.08	6.99	7.33	7.16	**6.57**
Ankylosaurus_&_Diplodocus_1	7.86	7.78	8.77	8.50	**7.38**
Desktop	6.03	5.97	6.68	6.56	**5.71**
Magnets_1	7.85	7.74	8.64	8.35	**7.31**
Fountain_&_Vincent_2	8.49	8.40	8.91	8.65	**7.82**
Friends_1	6.85	6.68	7.45	7.15	**6.40**
Color_Chart_1	8.11	8.06	8.95	8.74	**7.74**
ISO_Chart_12	8.23	8.23	8.57	8.30	**7.64**
Average (bpp)	7.54	7.44	8.06	7.82	**7.04**

using the first two methods, the LF images are converted into 225-frame video sequences by scanning and concatenating sub-aperture images following a serpentine scanning order. The X.265 library is used with $MAIN444-8$ profile and *lossless* mode for HEVC [14] and X.264 library is used with $QP = 0$ and $very-slow$ preset mode for AVC [15]. To compare with JPEG-LS [16] and CALIC [7], reference software with standard lossless configuration and official source code is used, respectively. The LF data is provided to JPEG-LS after converting it into lenslet array format with a resolution of $6510 \times 9375 \times 3$. For input to the CALIC, the YUV channels are concatenated vertically to form a 2-D image of resolution 19530×9375.

4.3 Results and discussion

The total size of uncompressed LF is $174.6MB$. For performance analysis, bits-per-pixel (bpp) is used as a compression rate metric. It is calculated by dividing the total number of bits after compression by the total number of pixels for all three channels (i.e. $15 \times 15 \times 625 \times 434 \times 3$). For each of the LFs, the achieved compression rate using the reference schemes and the proposed technique are listed in Tables 1 and 2. The proposed scheme achieves the highest compression performance compared to the reference techniques for all scenarios. On average, JPEG-LS achieves the lowest compression ratio while AVC achieves the second highest compression ratio. In comparison to HEVC, AVC, JPEG-LS and CALIC, our method achieves a significant bit reduction on **Dataset-1** of 8.51%, 5.26%, 11.35% and 8.81%, respectively, while on

Dataset-2, it achieves a reduction of 6.63%, 5.37%, 12.66% and 9.97% respectively.

It can be observed that the compression performance on **Dataset-2** is better than **Dataset-1**. That is due to the contrast adjustment on the LF images through gamma correction to generate **Dataset-1**. This process scales up the lower energy irradiances, which also increases entropy due to noise in the low energy regions. Overall, the proposed method is able to compress LF images in **Dataset-1** and **Dataset-2** to 34.5% and 29.3% of the uncompressed LF size.

The state-of-the-art in lossless LF compression, such as SMPC [3], pseudo video sequence MRP [5], or multi-reference MRP [6], use highly complex systems for encoding LF images. It is reported in [6], that the MRP techniques requires 13-15 hours to encode an LF image on average. Using the same LF images in **Dataset-2** but with RGB color channel, an average compression of 9.20 bpp is achieved using the proposed technique, while SMPC achieved 9.23 bpp on the same dataset, but with higher encoding complexity. In contrast, techniques such as, HEVC (RExt mode) [14], MRP (PVS) or MRP (MR), offers superior compression performance, but at the cost of hours of encoding time. The proposed method requires only 11 minutes to encode a single LF image in our non-optimized implementation. The implementation is symmetric and therefore the decoder requires similar amount of time for decoding. Due to CALIC source code implementation not being available, this technique was implemented by modifying a third party open-source C++ implementation. Therefore, a significant reduction in computation time can be achieved if the implementation is optimized for the speed. The experiments are performed on HP ZBook 15 G4 Intel Core i7-7700HQ CPU at 2.80GHz and 16GB RAM.

5 Conclusion

This paper presents a lossless coding scheme by modifying some important blocks of CALIC to improve the compression performance on LF images. Considering the structure of EPI, we have proposed the construction of a new predictor that can achieve significant improvements compared to CALIC's GAP predictor. With the addition of difference in prediction to error energy function and designing of a combined texture context, we further reduce the conditional entropy of the prediction errors. Eventually, we demonstrated that our compression scheme achieves better performance compared to state-of-the-art lossless compression techniques. This method currently deals with 8-bit lenslet based LF images. As a future goal, we plan to improve it, by optimizing the implementation for speed and by including support for 10-bit and for HDCA light fields addressing challenges due to a wider baseline.

Acknowledgement

This work was supported by the EU's H2020 ITN Programme through the MSCA Grant Agreement (RealVision) under Grant 765911.

6 References

[1] Ren Ng *et. al.*, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[2] Cristian Perra, "Lossless plenoptic image compression using adaptive block differential prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 1231–1234.

[3] Petri Helin, Pekka Astola, Bhaskar Rao, and Ioan Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," in *Proceedings of IEEE 3DTV-Conference*, 2016, pp. 1–4.

[4] Ionut Schiopu, Moncef Gabbouj, Atanas Gotchev, and Miska M. Hannuksela, "Lossless compression of subaperture images using context modeling," in *Proceedings of IEEE 3DTV Conference*, 2017, pp. 1–4.

[5] João M. Santos, Pedro A. A. Assuncao, Luis A. da Silva Cruz, Luis M. N. Tavor, Rui Fonseca-Pinto, and Sergio M. M. Faria, "Lossless coding of light field images based on minimum-rate predictors," *Journal of Visual Communication and Image Representation*, vol. 54, pp. 21–30, 2018.

[6] João M. Santos, Pedro A. A. Assuncao, Luis A. da Silva Cruz, Luis M. N. Tavor, Rui Fonseca-Pinto, and Sergio M. M. Faria, "Lossless compression of light fields using multi-reference minimum rate predictors," in *Data Compression Conference*, 2019, pp. 408–417.

[7] Xiaolin Wu and Nasir Memon, "Context-based, adaptive, lossless image coding," *IEEE Transactions on Communications*, vol. 45, no. 4, pp. 437–444, 1997.

[8] Marc Levoy and Pat Hanrahan, "Light field rendering," in *International Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 31–42.

[9] Martin Rerabek and Touradj Ebrahimi, "New light field image dataset," in *International Conference on Quality of Multimedia Experience*, 2016.

[10] Xin Huang and Soren Forchhammer, "Improved side information generation for distributed video coding," in *IEEE Workshop on Multimedia Signal Processing*, 2008, pp. 223–228.

[11] JPEG Pleno, "Light field dataset according to common testing conditions," `https://jpeg.org/plenodb/lf/pleno_lf//`, [Online; accessed 28-10-2019].

[12] Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034.

[13] "JPEG PLENO - Light Field Coding Common Test Conditions," ISO/IEC JTC 1/SC29/WG1, Vancouver, Canada, 2018.

[14] Gary J. Sullivan, Jill M. Boyce, Ying Chen, Jens-Rainer Ohm, C Andrew Segall, and Anthony Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE Journal of selected topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013.

[15] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[16] Marcelo J. Weinberger, Gadiel Seroussi, and Guillermo Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Transactions on Image processing*, vol. 9, no. 8, pp. 1309–1324, 2000.