HAL
open science

Learning-based lossless light field compression

Milan Stepanov, M. Umair Mukati, Giuseppe Valenzise, Søren Forchhammer, Frédéric Dufaux

To cite this version:

Milan Stepanov, M. Umair Mukati, Giuseppe Valenzise, Søren Forchhammer, Frédéric Dufaux. Learning-based lossless light field compression. IEEE International Workshop on Multimedia Signal Processing (MMSP'2021), Oct 2021, Tampere, Finland. pp.1-6, 10.1109/MMSP53017.2021.9733637 . hal-03298954

HAL Id: hal-03298954

https://hal.science/hal-03298954v1

Submitted on 23 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning-based lossless light field compression

Milan Stepanov[1], M. Umair Mukati[2], Giuseppe Valenzise[1], Søren Forchhammer[2], Frédéric Dufaux[1]
[1]*Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes*, 91190 Gif-sur-Yvette, France
[2]*DTU Fotonik, Technical University of Denmark*, 2800 Kgs. Lyngby, Denmark
[milan.stepanov, giuseppe.valenzise, frederic.dufaux]@l2s.centralesupelec.fr, [mummu, sofo]@fotonik.dtu.dk

***Abstract*—We propose a learning-based method for lossless light field compression. The approach consists of two steps: first, the view to be compressed is synthesized based on previously decoded views; then, the synthesized view is used as a context to predict probabilities of the residual signal for adaptive arithmetic coding. We leverage recent advances in deep-learning-based view synthesis and generative modeling. Specifically, we evaluate two strategies for entropy modeling: a fully parallel probability estimation, where all pixel probabilities are estimated simultaneously; and a partially auto-regressive estimation, in which groups of pixels are predicted sequentially. Our results show that the latter approach provides the best coding gains compared to the state of the art, while keeping the computational complexity competitive.**

***Index Terms*—light field, lossless coding, deep learning**

I. Introduction

Light fields are immersive visual representations that store the intensity of light rays of a scene. Techniques to capture light fields include using an array of cameras or inserting an array of microlenses (lenslets) in front of a camera sensor to acquire a collection of images from different viewpoints. Given the sheer amount of data this acquisition entails, light field coding is essential to practical applications.

In recent years, many methods for lossy compression of light fields have been proposed [1]. One of the main approaches consists of using view synthesis to radically reduce the number of views that need to be encoded by utilizing geometry information to generate remaining views at the decoder. Furthermore, by incorporating a residual signal, additional gains can be obtained.

In this work, we follow this paradigm and employ a recently proposed learning-based view synthesis technique to generate a high-quality prediction for lossless compression of light fields captured using LF cameras, e.g., Lytro. In contrast to typical lossless compression methods that use small local contexts for arithmetic coding, our approach can leverage a larger context obtained from several views. In addition, we estimate the symbol probabilities using a deep generative model [2]. While initially proposed to generate images, deep generative models, and in particular auto-regressive methods that explicitly compute the data likelihood, have been recently shown to provide accurate probability estimates for lossless coding of pictures [3], [4] and 3D point clouds [5]. To our knowledge, this is the first work applying deep conditional probability estimation in conjunction with view synthesis for coding of light fields. Our results show more than 2% improvement in bitrate compared to the best performing state-of-the-art method while being competitive in terms of execution time.

The rest of the paper is organized as follows. Section II cites the related work on lossless LF image compression with the emphasis on the approaches proposed for decoded lenslet modality. We also mention relevant generative models and recently proposed learning-based approaches for lossless image compression. Section III describes the proposed method. In Section IV, variants of the proposed method are compared, along with a comparison to state-of-the-art methods in terms of bitrate and execution time. Section V concludes the presented work and outlines possible research directions.

II. Related work

In this paper, we focus on light fields obtained by lenslet cameras. Lossless compression of light fields can be broadly divided into two groups. The first group of methods tackles the problem of coding the raw lenslet image [6]–[10]. The other group of methods deals with rectified light field images, i.e., light fields obtained from raw sensor image through decoding procedure [11]. Helin et al. [12] first encode the center view and its quantized depth map. Then, the center view is partitioned based on the depth values, and for each partition, disparities with respect to side views are computed and transmitted to partition the side views. Residuals, predictor coefficients, and binary sparsity masks are encoded. In a follow-up work [13], the compression performance is improved by segmenting the central view using a complex color- and depth-based segmentation approach, while variable length coding is used to encode the prediction mask and the coefficients of the sparse predictor. Santos et al. [14] conduct a study on the impact of reversible color transformations and data arrangements for pseudo-sequence generation for lossless light field coding. The study shows the superiority of the forward reversible multiple component transform (RCT) [15]. Schiopu et al. [16] propose a local, context-based method for lossless compression. Each pixel is adaptively predicted from close co-located pixels in the reference frame based on the edge information. The prediction is subtracted from input values, and three matrices are generated: small residual, high residual, and error sign. The small residual matrix is encoded using context modeling defined by the regions obtained from image segmentation. In [17]–[19], minimum-rate predictors are employed that obtain competitive performance, but with a high computational cost. Another line of works achieve

This project has received funding from EU's H2020 ITN programme, under the MSCA grant agreement No 765911 (RealVision).

interesting results with a complexity similar to standard coding tools by adapting the CALIC codec [20] to operate on the Epipolar Plane Image (EPI) representation of light field [21] [22].

Recently, deep learning has been applied to lossless compression of light fields [23]–[25]. Schiopu et al. [23] design a deep convolutional neural network that uses a neighborhood of six macro-pixels to predict the current macro-pixel. Then, the residual is encoded using a modified version of the CALIC codec. Kaya et al. [24] propose CEPINET, a variant of EPINET [26], which estimates disparity maps of corner views. Textures and disparity maps of corner views and the center view are encoded. Then, a disparity map is generated for each target view by warping the closest reference disparity map to the target view location. The disparity map is divided into connected regions, and for each region, the index of the best reference view (that minimizes MSE over the region) is assigned. Finally, the residual image is also computed and encoded. In [25], view synthesis and prediction methods based on macro pixels are proposed. The authors also study the influence of the size of the reference image set and modify the CALIC codec's binary mode to utilize different causal neighborhoods.

Recently, deep learning-based auto-regressive models have been successfully applied to lossless image compression [3] [4]. Their emergence is not surprising, considering their aim is to explicitly model the distribution of underlying data utilized for entropy coding. These tools achieve similar or better performance than state-of-the-art image codecs such as BPG and FLIF. Mentzer et al. [3] propose a hierarchical approach with learned feature extractors which generate latent representations transmitted to the decoder. At the decoder, latent representations are processed by decoding blocks and are used to estimate the probability distributions necessary for decoding the latents at the higher level. In [4], the authors design a two-layer lossless compression method by leveraging the standard coding tool BPG. The input image is first encoded with BPG. Then the residual between the input and its processed variant is entropy coded using the probability distribution estimated by a neural network. The network provides parameters of the probability distribution for each pixel allowing efficient, parallelized processing.

III. Proposed method

Our approach takes inspiration from learning-based lossless image compression [4] and image generation [27] [28], which are combined with view synthesis [29] to enable the prediction of views from a set of reference views. The overall design of the proposed method is presented in Fig. 1. At the encoder side, reference views and the input view index are provided to the View Synthesis module, which predicts the input view. Then, the prediction is subtracted from the input view. The obtained residual signal is entropy coded using an arithmetic coder (AC) provided with probability distributions computed on a per-pixel basis by the Entropy Modeling block given the prediction. At the decoder, the prediction from the View Synthesis module is used to estimate probability distributions for decoding, and the decoded residual signal is added to the prediction to reconstruct the input view. View Synthesis and Entropy Model are briefly described in the following subsections.

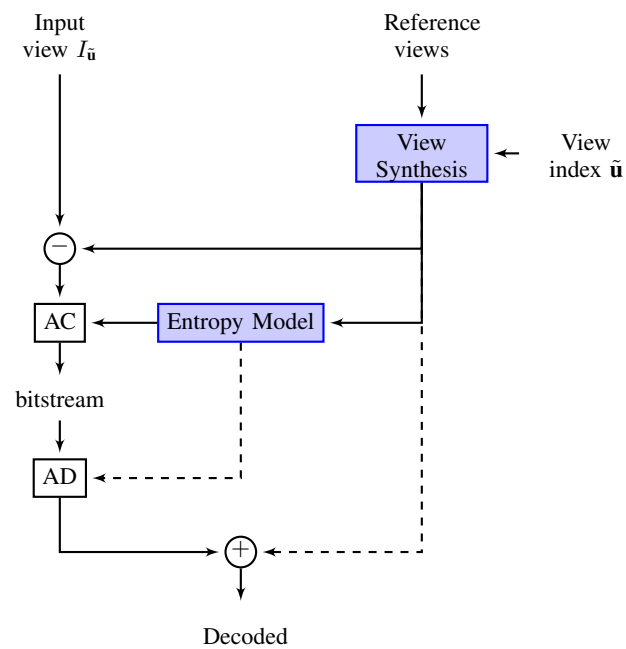

Fig. 1. The block diagram of the proposed method. Note, View Synthesis and Entropy Model are utilized at encoder and decoder side as they depend only on previously decoded information. AC and AD denote arithmetic coder and arithmetic decoder, respectively.

A. View Synthesis

To estimate a novel view, Navarro et al. [29] propose three distinctive blocks, a feature extractor, a disparity estimator and a per-pixel view selector. Given the set of corner views $I_{\mathbf{c}}$ of a light field image $L(\mathbf{x}, \mathbf{u})$, where $\mathbf{x}$ and $\mathbf{u}$ denote spatial and angular positions of light rays, and the position of a view to be synthesized $\tilde{\mathbf{u}}$, the feature extraction network $\mathcal{F}_e$ computes the feature map of each corner view $I_{\mathbf{c}_i}$ independently:

$$F_{\mathbf{c}_i} = \mathcal{F}_e(I_{\mathbf{c}_i}, \tilde{\mathbf{u}}). \tag{1}$$

The features are then concatenated in a vector $\mathbf{F}$ together with the position of the novel view $\tilde{\mathbf{u}}$ and provided to the disparity estimation network $\mathcal{F}_d$ to estimate the vector of disparity maps $\mathbf{D}$ of the novel view:

$$\mathbf{D} = \mathcal{F}_d(\mathbf{F}, \tilde{\mathbf{u}}). \tag{2}$$

Initial estimates of the novel view $W_{\mathbf{c}_i}$ are obtained by backward warping the corner views using the corresponding disparity maps,

$$W_{\mathbf{c}_i}(\mathbf{x}) = I_{\mathbf{c}_i}(\mathbf{x} + D_{\mathbf{c}_i}(\mathbf{x}) \cdot (\mathbf{c}_i - \tilde{\mathbf{u}})). \tag{3}$$

In the final stage of the scheme, the selection network $\mathcal{F}_s$ computes the contribution $M_{\mathbf{c}_i}$ for each initial estimate $W_{\mathbf{c}_i}$

$$\mathbf{M} = \mathcal{F}_s(\mathbf{W}, \mathbf{D}, \tilde{\mathbf{u}}). \tag{4}$$

where $\mathbf{M}$ and $\mathbf{W}$ are concatenated vectors of merging maps $M_{\mathbf{c}}$ and initial estimates $W_{\mathbf{c}}$. The final estimate of the novel view $\hat{I}_{\tilde{\mathbf{u}}}$ is obtained by merging initial estimates as a weighted sum

$$\hat{I}_{\tilde{\mathbf{u}}} = \hat{L}(\mathbf{x}, \tilde{\mathbf{u}}) = \sum_{\mathbf{c}_i \in \mathbf{c}} M_{\mathbf{c}_i}(\mathbf{x}) W_{\mathbf{c}_i}(\mathbf{x}). \tag{5}$$

B. Entropy model

Mentzer et al. [4] propose a Residual Compressor (RC) which takes the decoded input image, obtained through BPG coding, and estimates a set of parameters which model the probability mass function of the residual signal r_x between the input pixel value x and its decoded version $\hat{x}$, i.e., $p_m(r_x|\hat{x})$, which is used to encode/decode the residual. The probability mass function is modeled with a mixture of logistic functions

$$p(r_x) = \sum_{k=1}^{K} \pi_k p_L(r_x|\mu_k, \sigma_k), \tag{6}$$

where $p_L(x|\mu, \sigma) = \frac{e^{-\frac{r_x-\mu}{\sigma}}}{\sigma(1+e^{-\frac{r_x-\mu}{\sigma}})^2}$ and K is the number of mixtures. Furthermore, the method encodes/decodes pixels channel-by-channel and conditions the current color channel on previously decoded channel(s):

$$\tilde{\mu}_c^k = \mu_c^k + \sum_{c_i \in \mathbf{c}_{prev}} \lambda_{c_i}^k r_{x,c_i}. \tag{7}$$

We utilize the same architecture as proposed in [4] while changing the number of filters, i.e., $C_f = 64$, and the number of residual blocks to 8. More importantly, we use the *synthesized view* $\hat{I}_{\tilde{\mathbf{u}}}$ as input and compute the residuals with respect to it. The weights of the RC network are learned by minimizing the cross-entropy loss

$$J(\theta) = -\sum_{x \in \mathbf{I}_{\tilde{\mathbf{u}}}} \log_2 \left(p_m(r_x|\hat{x})\right). \tag{8}$$

The architecture of [4], which we denote as *Base* in the following, estimates the pixel distribution for every pixel in a single pass, allowing parallel and efficient processing. On the other hand, to improve the estimation locally, one could use neighboring, previously decoded pixel values as it is typically done in lossless compression schemes. In order to improve performance, we propose to divide pixels into coding groups and to use previously decoded groups to estimate the parameters for the following groups. Notice that a similar approach has been considered in a recent work for image generation [27] and point cloud coding [30].

The simplest version of this approach consists of using a group for each pixel, which corresponds to PixelCNN [2]. By increasing the size of the groups and by assuming that pixels within each group are conditionally independent, it is possible to estimate in parallel all the pixels in a group from previously decoded groups. In our experiments, we use four groups (each containing $N/4$ pixels), where each group contains every other pixel in the horizontal and vertical directions, similar to a checkerboard pattern.

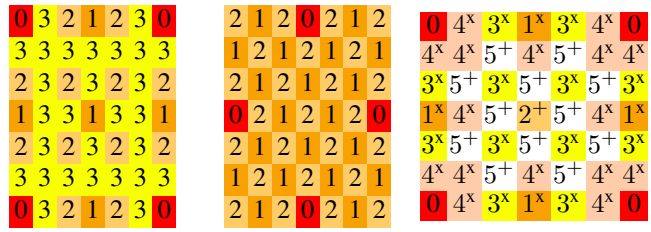

Fig. 2. Hierarchical levels for prediction of views with Corner arrangement (left) and Cross arrangement (middle). At each level i, views can be predicted from decoded views from previous levels, i.e. $i_{prev} < i$. Prediction scenarios applied in Hybrid variant (right). The number denotes the prediction level i while the superscripts denote the arrangement of the reference views used for the prediction, i.e. x and + denote Corner and Cross arrangements.

C. Training details

We train our models on 3323 LF images of size $7 \times 7 \times 376 \times 541$ from the Flowers dataset [31]. First, the view synthesis method for both arrangements, *Corner* and *Cross*, are pretrained. Here we used the same settings as done in [32]. Then, the view synthesis network parameters are fixed, and we train the residual estimation network. The input LFs are cropped along angular dimensions by selecting center 8 views. At each iteration, a 4D patch of size $7 \times 7 \times 128 \times 128$, the positions of the reference views and the target view are randomly selected. Then, we randomly perturb color channels to increase the training dataset's color diversity and finally apply gamma correction selected from an interval $[0.4, 1.0]$. We use a batch size of 16 and Adam optimizer [33] to update weights.

IV. Results

A. Dataset

The proposed method is evaluated on 12 LF images proposed in [34] from the EPFL dataset [35]. We use Lytro Power Tool[1] (LPT) to decode raw LFs, we apply a gamma correction of 0.45 and quantize to 8 bits. Moreover, LFs are cropped in angular and spatial dimensions to 7×7 views (we use top-left 7×7 crop of center 8×8 crop) and 320×512 pixels, respectively. Note that the outer views of decoded LFs are corrupted with vignetting artifacts and thus cropped to a smaller size.

B. Prediction variants

In the view synthesis task, typically, the whole LF is synthesized from a fixed set of reference views. On the other side, in lossless compression, as views become available after decoding, it becomes possible to dynamically select reference views and leverage reduced baseline between views for improved prediction. However, a potential drawback could be reduced random access capabilities as later views depend

[1] https://github.com/kmader/lytro-power-tools, accessed on Nov. 10th, 2020.

TABLE I
THE PERFORMANCE EVALUATION OF VARIANTS OF THE PROPOSED METHOD PRESENTED IN TERMS OF BITS PER PIXEL (BPP).

	Single		Hierarchical			
Sequence	*Corner*	*Cross*	*Corner*	*Cross*	*Hybrid*	*Hybrid (Base)*
Bikes	6.51	6.32	5.90	5.97	5.81	6.22
Danger de Mort	7.23	7.02	6.51	6.59	6.39	6.89
Flowers	7.21	6.96	6.42	6.53	6.31	6.79
Stone Pillars Outside	6.81	6.58	6.08	6.15	5.98	6.33
Vespa	6.49	6.21	5.88	5.91	5.79	6.10
Ankylosaurus	5.14	5.06	4.86	4.90	4.82	4.84
Desktop	6.84	6.63	6.24	6.30	6.15	6.62
Magnets	5.16	5.08	4.90	4.94	4.86	4.90
Fountain & Vincent	6.56	6.37	6.01	6.01	5.90	6.13
Friends	6.07	5.90	5.54	5.60	5.45	6.04
Color Chart	5.93	5.65	5.43	5.45	5.35	6.00
ISO_Chart	6.33	6.11	5.80	5.75	5.69	5.92
Average	6.36	6.16	5.80	5.84	5.71	6.06

on previously encoded/decoded views. Here we compare both cases, namely the *Single* method, which operates on a single level and predicts views from the corner references, and the *Hierarchical* where three levels are incorporated as depicted on the left part of Figure 2. In the first level, the method processes the corner views and the middle views on the boundary of the LF, i.e., between corner views. Then, in the second level, a subset of views in each quadrant is estimated. Finally, the rest of the views are predicted from their neighboring four views in the third level. Besides utilizing corner views as reference views, we also examine cross arrangement, which proved to be superior to the corner arrangement in recent work [32]. We note that the corner approach is more flexible in selecting reference view as the cross approach demands odd angular size in order to select reference views. For instance, in an LF of 4×4 angular size, in order to utilize cross arrangement, it would be necessary to operate on LFs of 3×3 angular size. Conversely, the cross arrangement provides superior prediction. Thus, we finally consider a *Hybrid* approach which benefits from the qualities of both arrangements. The approach is illustrated on the right part of Figure 2 where the numbers denote hierarchical levels while the superscripts denote arrangements. The reference views are encoded independently using the L3C method, but any method can be utilized in general.

TABLE II
THE PERFORMANCE EVALUATION IN TERMS OF BITS PER PIXEL (BPP) OF IMAGE COMPRESSION TOOLS APPLIED ON EACH VIEW SEPARATELY.

	L3C	RC	FLIF	JPEG XL
Average	8.07	8.18	7.90	7.61

Table I shows a comparison of the proposed variants. The columns marked as *Single* present the performance of the proposed method where four reference views are independently encoded and used to estimate all the other views. This setup is the most efficient from a random access perspective for the given view synthesis module. Better performance of *Cross* arrangement can be noted compared to *Corner* arrangement. These results are also aligned with previous research [32].

The terms *Hierarchical* denote variants of the proposed method, which allow dynamic selection of reference views in order to achieve better prediction and lower bitrate at the cost of reduced random access capabilities. The hierarchical levels of *Corner*, *Cross* and *Hybrid* sub-column are illustrated in Figure 2. First, the *Corner* variant shows better performance compared to the *Cross* variant. Interestingly, the result is opposite to the result observed for *Single* setup, which could be explained by the lower number of hierarchical levels and reduced prediction capabilities and implying the lack of flexibility of the *Cross* arrangement. Second, additional gains compared to the two arrangements can be achieved by combining them as presented in the column *Hybrid*. Finally, we also present the result of the *Base* method (without spatial grouping), which has a lower memory footprint and better execution time, but lower probability estimation accuracy. As the Entropy model of the *Base* method has four times less parameters compared to the proposed method, in order to verify that the gains in the proposed method are achieved via grouping and auto-regressive modeling, we also train the *Base* network with approximately the same number of parameters. The two *Base* models showed similar performance. Notice that all the proposed variants require a similar processing time, while they just differ in the views selection mechanism.

C. Comparison with state-of-the-art methods

In Table III, the proposed method is compared to state-of-the-art methods. We select a few general lossless schemes for image and video compression and an approach designed for lossless compression of LFs. *HEVC* denotes HM v16.22 implementation of High Efficiency Video Coding (HEVC) standard [36]. In this case, an LF image is first reshaped into a pseudo-video sequence following a serpentine scan order and encoded in the lossless mode in addition to the Main-RExt profile. *L3C* [3] and *RC* [4] represent learning-based methods for lossless image compression. *RC* method is the most relevant work as the approach is incorporated in the proposed method. Furthermore, we also evaluate the test set on FLIF codec [37] and JPEG XL [38]. In order to allow

TABLE III
THE COMPARISON OF THE PROPOSED METHOD AND AVAILABLE METHODS FROM LITERATURE IN TERMS OF BITRATE (BPP).

	General lossless schemes					LF lossless schemes	
Sequence	*HEVC*	*L3C*	*RC*	*FLIF*	*JPEG XL*	*EPIC-RCT*	*Proposed*
Bikes	6.69	7.28	7.54	6.19	5.95	6.14	5.81
Danger de Mort	7.19	7.90	8.03	6.70	6.49	6.71	6.39
Flowers	7.08	8.00	8.27	6.84	6.49	6.67	6.31
Stone Pillars Outside	6.61	6.93	7.52	6.01	5.84	6.33	5.98
Vespa	6.64	7.12	7.30	6.10	6.03	5.98	5.79
Ankylosaurus	5.43	5.84	5.79	4.77	5.48	4.83	4.82
Desktop	6.76	7.18	7.56	6.22	6.33	6.26	6.15
Magnets	5.49	6.03	5.92	4.87	5.60	4.80	4.86
Fountain & Vincent	6.71	7.48	7.35	6.14	6.14	6.03	5.90
Friends	6.10	6.50	6.89	5.47	5.38	5.69	5.45
Color Chart	6.00	6.57	6.57	5.35	5.69	5.37	5.35
ISO_Chart	6.42	6.47	6.22	5.39	5.37	5.59	5.69
Average	6.43	6.94	7.08	5.84	5.90	5.87	5.71

TABLE IV
THE COMPARISON OF TOTAL LF ENCODING AND DECODING TIMES PRESENTED IN MINUTES.

Method	*HEVC*	*FLIF*	*JPEG XL*	*EPIC-RCT*	*Proposed*	*Base*
Encoding	6.47	0.62	0.34	0.13	0.17	0.14
Decoding	0.05	0.14	0.07	0.13	0.17	0.14

image codecs to exploit angular correlation, LFs are converted to lenslet representation before processing. For reference, we also report results for independent encoding of each view in Table II.

Among the LF lossless coding approaches we select EPIC [22]. As EPIC was designed to operate independently on color channels in the YCbCr color space, we employ RCT [15] to input RGBs. We also evaluated the approach [16] with the publicly available software. However, as the results appear considerably higher compared to other methods (the code yielded an average bitrate of 9.32 bpp), we do not report them in Table III.

We notice that *RC* and *L3C* obtain worse performance compared to other approaches. This might be due to the domain shift between the original training data of these methods and the lenslet test data. On average, our method outperforms *FLIF*, *JPEG XL*, *EPIC* by 2.23%, 3.22% and 2.73%, respectively. Interestingly, *FLIF* and *JPEG XL* perform quite well compared to the other two methods, which are designed to exploit the correlation in LFs. It is highly likely that due to the small baseline between the views and small angular size, the lenslet format allows to exploit spatial and angular similarities efficiently. *HEVC* underperforms compared to later approaches. This result aligns with the literature results.

A per-content comparison shows that the proposed method outperforms other methods on most of the sequences. The exceptions are less natural sequences such as Ankylosaurus, Magnets, and the two charts. We note various potential reasons for this behavior. Sequences like charts are mainly flat, which is a challenging content for geometry estimation. Moreover, the noise is strongly present, and the color constancy between the reference views and target views is lacking. On the other side, object-rich sequences with diverse geometry content suit the proposed method exceptionally well, as it can be observed in the performance of Bikes, Fountain & Vincent and Friends.

In Table IV, we report the processing time of encoder and decoder with different codecs. Notice that by construction, the complexity in our approach is symmetric at the encoder and decoder side, and is especially efficient at encoding compared to conventional methods such as *FLIF*, *JPEG XL* and *HEVC*.

V. CONCLUSIONS

We have proposed a learning-based method for lossless compression of LF images. The method incorporates a view synthesis block that provides an initial estimate of the encoded view. This is later used as a context to build an entropy model, which estimates for each pixel a conditional probability to be used in adaptive arithmetic coding. By introducing partial auto-regressive relations among groups of pixels, the proposed method outperforms state-of-the-art methods in terms of bitrate while maintaining a low computational complexity.

Future work includes several directions. First, we plan to add an additional residual coding layer to improve the estimation of probabilities at the decoder side. Next, jointly training view synthesis and the entropy model should further improve performance. Last but not least, the evaluation should be extended to LFs with larger angular size and wider baseline. These data are more challenging and possibly more critical to address.

REFERENCES

[1] C. Conti, L. D. Soares, and P. Nunes, "Dense light field coding: A survey," *IEEE Access*, vol. 8, pp. 49 244–49 284, 2020.

[2] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, "PixelCNN++: A pixelcnn implementation with discretized logistic mixture likelihood and other modifications," in *ICLR*, 2017.

[3] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Practical full resolution learned lossless image compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 629–10 638.

[4] F. Mentzer, L. V. Gool, and M. Tschannen, "Learning better lossless compression using lossy compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6638–6647.

[5] D. T. Nguyen, M. Quach, G. Valenzise, and P. Duhamel, "Learning-based lossless compression of 3D point cloud geometry," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4220–4224.

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

[7] A. Miyazawa, Y. Kameda, T. Ishikawa, I. Matsuda, and S. Itoh, "Lossless coding of light field camera data captured with a micro-lens array and a color filter," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–4.

[8] F. Murgia and D. Giusto, "A comparison of raw light field lossless data compression algorithms," in *2016 24th Telecommunications Forum (TELFOR)*, 2016, pp. 1–4.

[9] I. Tabus and P. Helin, "Microlens image sparse modelling for lossless compression of plenoptic camera sensor images," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1907–1911.

[10] I. Tabus and E. Palma, "Lossless compression of plenoptic camera sensor images," *IEEE Access*, vol. 9, pp. 31 092–31 103, 2021.

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

[12] P. Helin, P. Astola, B. Rao, and I. Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," in *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2016, pp. 1–4.

[13] ——, "Minimum description length sparse modeling and region merging for lossless plenoptic image compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1146–1161, 2017.

[14] J. M. Santos, P. A. A. Assunção, L. A. da Silva Cruz, L. Távora, R. Fonseca-Pinto, and S. M. M. Faria, "Lossless light-field compression using reversible colour transformations," in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2017, pp. 1–6.

[15] "ITU-T Recommendation T.812: Information technology – JPEG 2000 image coding system: An entry level JPEG 2000 encoder,," International Telecommunication Union, Geneva, CH, Standard, 2007.

[16] I. Schiopu, M. Gabbouj, A. Gotchev, and M. M. Hannuksela, "Lossless compression of subaperture images using context modeling," in *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2017, pp. 1–4.

[17] J. M. Santos, P. A. Assunção, L. A. da Silva Cruz, L. M. Távora, R. Fonseca-Pinto, and S. 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320318300518

[18] J. M. Santos, P. A. A. Assunção, L. A. d. S. Cruz, L. M. N. Távora, R. Fonseca-Pinto, and S. M. M. Faria, "Lossless compression of light fields using multi-reference minimum rate predictors," in *2019 Data Compression Conference (DCC)*, 2019, pp. 408–417.

[19] J. M. Santos, L. A. Thomaz, P. A. Assunção, L. A. Cruz, L. Távora, and S. M. Faria, "Lossless coding of light fields based on 4D minimum rate predictors," *arXiv preprint arXiv:2104.06252*, 2021.

[20] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE transactions on Communications*, vol. 45, no. 4, pp. 437–444, 1997.

[21] M. U. Mukati and S. Forchhammer, "EPIC: Context adaptive lossless light field compression using epipolar plane images," in *2020 Data Compression Conference (DCC)*, 2020, pp. 43–52.

[22] ——, "Epipolar plane image-based lossless and near-lossless light field compression," *IEEE Access*, vol. 9, pp. 1124–1136, 2021.

[23] I. Schiopu and A. Munteanu, "Macro-pixel prediction based on convolutional neural networks for lossless compression of light field images," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 445–449.

[24] E. C. Kaya and I. Tabus, "Corner view disparity estimation for lossless light field compression," in *Proc. 1st Eur. Light Field Imag. Workshop (ELFI)*, 2019, pp. 1–4.

[25] I. Schiopu and A. Munteanu, "Deep-learning-based macro-pixel synthesis and lossless coding of light field images," *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e20, 2019.

[26] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, "EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4748–4757.

[27] S. Reed, A. van den Oord, N. Kalchbrenner, S. G. Colmenarejo, Z. Wang, Y. Chen, D. Belov, and N. de Freitas, "Parallel multiscale autoregressive density estimation," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 2912–2921. [Online]. Available: http://proceedings.mlr.press/v70/reed17a.html

[28] A. Kolesnikov and C. H. Lampert, "PixelCNN models with auxiliary variables for natural image modeling," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 1905–1914. [Online]. Available: http://proceedings.mlr.press/v70/kolesnikov17a.html

[29] J. Navarro and N. Sabater, "Learning occlusion-aware view synthesis for light fields," *Pattern Analysis and Applications*, pp. 1–16, 2021.

[30] D. Thanh Nguyen, M. Quach, G. Valenzise, and P. Duhamel, "Multiscale deep context modeling for lossless point cloud geometry compression," *arXiv e-prints*, pp. arXiv–2104, 2021.

[31] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2243–2251.

[32] M. U. Mukati, M. Stepanov, G. Valenzise, S. Forchhammer, and F. Dufaux, "Improved deep distributed light field coding," *IEEE Open Journal of Circuits and Systems*, vol. 2, pp. 325–337.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2015.

[34] M. Rerábek, T. Bruylants, T. Ebrahimi, and F. Pereira, "ICME 2016 grand challenge: Light-field image compression - call for proposals and evaluation procedure," 2016.

[35] M. Rerábek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[36] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. CSVT*, vol. 22, pp. 1649–1668, 2012.

[37] J. Sneyers and P. Wuille, "FLIF: Free lossless image format based on maniac compression," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 66–70.

[38] J. Alakuijala, R. van Asseldonk, S. Boukortt, M. Bruse, I.-M. Comșa, M. Firsching, T. Fischbacher, E. Kliuchnikov, S. Gomez, R. Obryk *et al.*, "JPEG XL next-generation image compression architecture and coding tools," in *Applications of Digital Image Processing XLII*, vol. 11137. International Society for Optics and Photonics, 2019, p. 111370K.