FuRA: Fully Random Access Light Field Image Compression

Hadi Amirpour*, Christine Guillemot†, and Christian Timmerer*

*Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität, Klagenfurt, Austria
†Inria Rennes – Bretagne-Atlantique, 263 Avenue Générall Leclerc, 35042 Rennes Cedex, France

***Abstract*—Light fields are typically represented by multi-view images, and enable post-capture actions such as refocusing and perspective shift. To compress a light field image, its view images are typically converted into a pseudo video sequence (PVS) and the generated PVS is compressed using a video codec. However, when using the inter-coding tool of a video codec to exploit the redundancy among view images, the possibility to randomly access any view image is lost. On the other hand, when video codecs independently encode view images using the intra-coding tool, random access to view images is enabled, however, at the expense of a significant drop in the compression efficiency. To address this trade-off, we propose to use neural representations to represent 4D light fields. For each light field, a multi-layer perceptron (MLP) is trained to map the light field four dimensions to the color space, thus enabling random access even to pixels. To achieve higher compression efficiency, neural network compression techniques are deployed. The proposed method outperforms the compression efficiency of HEVC inter-coding, while providing random access to view images and even pixel values.**

Light field, coding, image representation, neural representation.

I. Introduction

In light field imaging, both spatial and angular information of a 3D scene is captured, resulting in a 4D function. In this way, post-capture actions, such as changing view perspective and refocusing, are enabled. Light fields are usually represented with multiple views of a 3D scene captured from different angles and tilts. Fig. 1 illustrates the capturing of multi-view images using a multi-array camera. The couple (u,v) represents the angular information (view images), while (x,y) represents the spatial information (intensity of light rays in each pixel). That is, the light field function (LF) translates positional information (x, y, u, v) into color information (L) as follows:

$$L = LF(x, y, u, v) \tag{1}$$

Light fields represent a huge amount of data, which calls for highly efficient compression methods. Conventionally, as shown in Fig. 2, light field view images are arranged in pseudo video sequences (PVS) using a predefined scan order. Inter-coding tools of a video codec are then used to exploit inter-view redundancy among different views of the PVS. For example, in JPEG Pleno HEVC anchor [1] light field views are arranged as a PVS using serpentine scan order and HEVC [2] is used to compress each PVS.

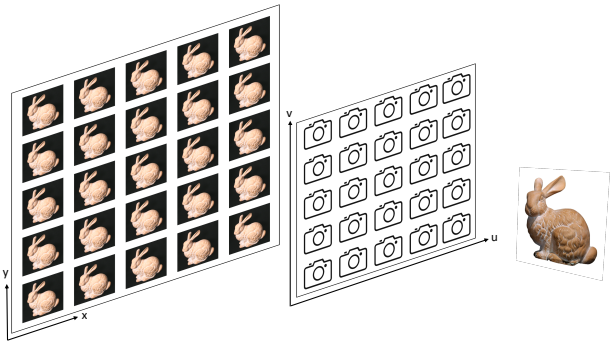


Fig. 1: Light field multi-view representation. A multi-camera array is used to capture view images from different angles and tilts.

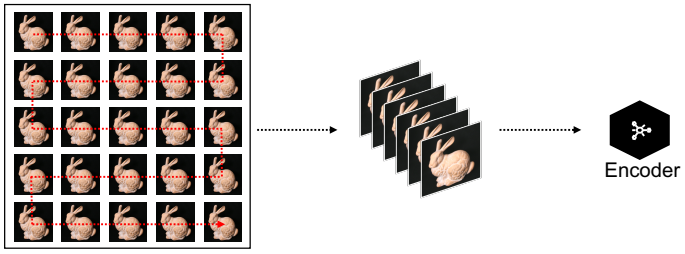


Fig. 2: Light field view images are converted to a PVS and then compressed with a video encoder.

Although using the inter-frame coding tool of a video codec is highly efficient in exploiting redundancy between light field view images, it makes view images dependent on each other, thus making it impossible the random access to the different views. On the other hand, encoding view images independently of each other (*i.e.*, , intra-coding), where views can be accessed randomly, leads to a low compression efficiency. Therefore, a trade-off between compression efficiency and random access to view images is achieved.

To address this trade-off, we propose *FuRA*, a fully random access light field image coding method. The implicit light field neural representations use a multi-layer perception (MLP) neural network to map positional information (light field pixel's information) into the color space. Although they

show compact storage, their compression efficiency is still low. We develop a neural network compression method to improve the compression efficiency of implicit light field neural representations.

In the next section, we introduce light field compression methods, particularly those that provide random access, and also we give a quick overview of implicit light field neural representations. The proposed fully random access light field image method, *i.e.*, *FuRA*, is presented in Section III. Experimental results are given in Section IV, followed by the conclusion.

II. Related work

In this section, we first introduce light field image compression methods. The state-of-the-art light field neural representations are then presented.

A. Light Field Compression

PVS-based light field image compression methods make all view images dependent on each other; however, they exploit the redundancy among view images efficiently. Some approaches improve random access performance while providing high compression efficiency. In Linear Approximation Prior (LAP) [3], the linearity between the views is exploited. To this end, the view images are divided into two non-overlapping sets A and B. View images in set A are arranged as a PVS and it is compressed with an HEVC video encoder. The view images in set A are then used to reconstruct dropped views that exist in set B using a global optimization strategy. Therefore, view images in set B can be encoded and decoded independent of each other, but they require the encoding of all the views in set A. In HESL [4], the view images are divided into four non-overlapping PVSs, and each PVS is encoded independently, resulting in the improved random access performance. Amirpour *et al.* [5] group 3×3 view images into a macro view image (MVI) and encode them with a proper reference structure that random access performance is improved significantly.

In Multidimensional Light field Encoder (MuLE) [6] 4D light field is partitioned into smaller 4D blocks and 4D-DCT is deployed to exploit the redundancy in blocks. Independent encoding of blocks provide random access to each block; however, random access to view images is not provided since 4D blocks comprises of co-located blocks in all view images.

Various prediction structures are compared in [7] in terms of encoding efficiency and random access. The *CenterView* structure, where only the center view image is used as a reference to encode other view images, shows better performance in addressing the trade-off between compression efficiency and random access.

In [8], view images are divided into evenly distributed intra images (I-image) and predicted (P-image) images. I-images are intra-coded and P-images are inter-coded are predicted using I-images as references. However, the huge amount of intra-coded images leads to a low encoding efficiency. Similarly, in MRF [9], the certain image views are intra-coded and the remaining views are selected as predicted view images.

In [10], a hierarchical compression scheme which is suitable for interactive rendering is proposed. To this light, a tree structure is constructed and the representative view images in a higher level of the hierarchy is computed by filtering clusters of four spatially close view images. Residual view images for each hierarchy are constructed, compressed, and added to the bitstream. Since residual view images are divided into non-overlapping rectangular blocks and encoded independently random access to blocks is enabled. However, random access to view images is not provided.

B. Implicit Light Field Neural Representations

The concept of **Ne**ural **R**adiance **F**ield (NeRF) has been introduced in [11] as an implicit model mapping 5D vectors (3D coordinates plus 2D viewing directions) to opacity and color values. The model, based on multi-layer perceptrons (MLP), is trained by fitting the model to a set of input view images, and can be used to generate any view of the light field using volume rendering techniques.

This seminal work has triggered a lot of interest leading to many variants aiming at either reducing the number of input view images as in [12] or in [13], or aiming at the possibility to generalize to new scenes [13], or at learning without prior knowledge of camera parameters [14]. A solution is also proposed to learn components that can be very useful in scene rendering (e.g., albedo, illumination, normals, shading) for relighting [15], [16], [17], [18]. The authors in [19] first transform the 4D light field by leveraging Gegenbauer polynomials basis, and learn the mapping from these basis functions to color.

The concept is further generalized to X-Fields in [20] defined as sets of 2D images taken across different view, time or illumination conditions. By limiting the novel viewpoints to be on the same side of the cameras, e.g., front views only, the NeuLF method in [21] achieves 1000x speedup over NeRF during inference, while producing similar rendering quality.

III. The proposed Fully Random Access Light field image coding

In video codecs, the reference structure plays a key role in addressing the trade-off between compression efficiency and random access. To investigate its impact, we convert the $Bunny$ light field into a PVS and encode it with various reference structures (see Fig. 3), using the HEVC coding standard:

① Intra-period = 1: All view image are encoded independently (intra-coded).
② Intra-period = 2: Each intra-coded view image is used as a reference to predict its following view image (P-frame).
③ Intra-period = 4: Each intra-coded view image is used as a reference to predict its three following view images.

The rate-distortion performance is shown in Fig.4. It can be seen that, with the growing number of predicted frames (P-frames), the compression efficiency is increased at the cost of

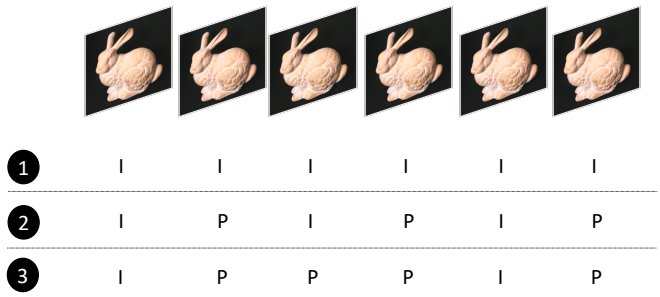


Fig. 3: Three reference structures with different intra-periods.

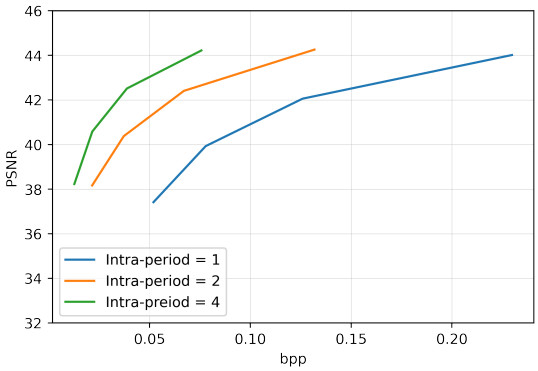


Fig. 4: Rate-distortion performance of the same PVS encoded with different intra-periods, using HEVC. With the increasing *Intra-period*, the number of Intra frames is decreased, resulting in higher compression efficiency but lower random access performance.

reduced random access performance. To solve this trade-off, we propose to use neural representations which provide easy local reconstruction.

A. Neural Network Representation

As shown in Fig.5, neural representations map positional information into color values. To this end, a multi-layer perceptron (MLP) is trained for each light field to map positional information to color information. This direct mapping allows neural representations to have access to color value of each pixel independently of all other pixels enabling full random access to view images and their pixel views. Weight values of different layers (including input layer, hidden layers, and output layer) are stored to reconstruct the corresponding light field. Hence, light fields are represented by neural networks instead of traditional 4D representation with multi-view images.

In this paper, we use SInusoidal Gegenbauer NETwork [22], or *SIGNET*, to represent static light field images since it can represent light field images more compactly than state-of-the-art compression methods. In SIGNET, a fully-connected neural network is used to learn the mapping function between each light field's positional information and its corresponding color values. To improve the learning capacity, an input transformation strategy based on the Gegenbauer polynomials with sinusoidal activation functions is deployed.

The compression efficiency of a neural network depends on the size of the network, *i.e.*, network depth (L) and network width (M). L denotes the number of layers and M denotes the number of neurons in each layer. With the growing number of M and L, the reconstruction quality of the neural network may increase at the cost of increased network size. To evaluate the compression efficiency of the neural network, we train the same network with different values of L and M for the 5×5 view images of $Bunny$ light field image. Fig.6 shows the reconstruction quality for the various L and M values. It is seen that a network with greater network depth and network width shows higher reconstruction quality. For example, at fixed L values, networks with higher M values show higher compression efficiency. However, considering the size of the network, at the same L the network with lower M shows a higher reconstruction quality. Therefore, the compression efficiency should be optimized considering network depth and network width.

Image and video compression techniques are used to exploit the redundancy between views when light fields are represented with multi-view images. In the next section, we deploy neural network compression techniques to improve the compression efficiency of the neural representation used in this paper.

B. Neural Network Compression

In image and video compression techniques usually use a transformation (*e.g.*, DCT) into transform signal information to frequency information and remove frequencies that Human Vision System (HVS) is less sensitive to them.

To compress neural networks, various techniques have been proposed [23]. Weight quantization and Huffman coding are neural network compression methods with low complexity [24]. In this paper, to quantize the weights of a trained neural network, we use k-means to cluster the weights in each layer of the trained neural network, independently of the other layers. Thus, the weights within the same cluster are assigned the same value using the linear initialization. Linear initialization linearly spaces the centroids between the [min, max] of the original weights. Therefore, n weights of each layer are clustered into k clusters, where $n >> k$. Given k clusters, only $log_2(k)$ bits are therefore needed to encode the index of each cluster, the shared weight of each cluster being quantized using a uniform scalar quantizer. These shared weights are then fine-tuned via back-propagation, to improve the reconstruction quality after quantization of the shared weights. So the total number of bits for a shared weight is the $log_2(k)$ bits to encode the index of each cluster plus the number of quantization bits for the shared weight. Fig. 7 shows the histograms of weights for the second hidden layer of an MLP that has been trained to represent the $Bunny$ light field image before and after quantization. Fig. 8 shows the reconstruction quality of the $Bunny$ light field neural representation after clustering and quantization,

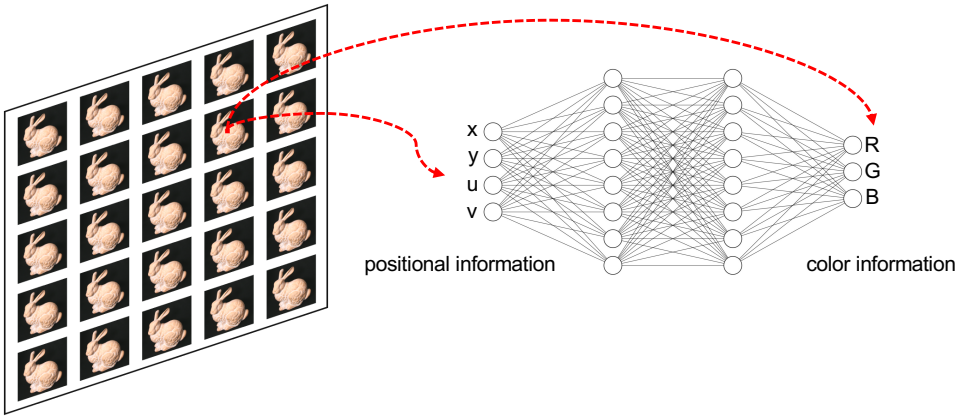


Fig. 5: Light field neural representations translate positional information into color space. In this way, random access to pixel values; consequently view images is provided.

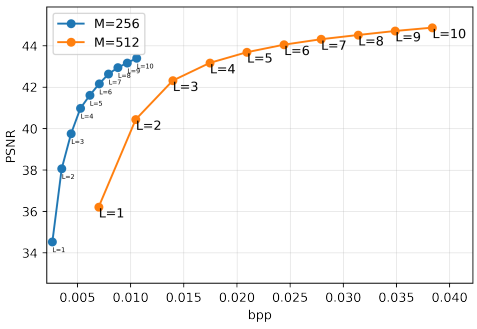
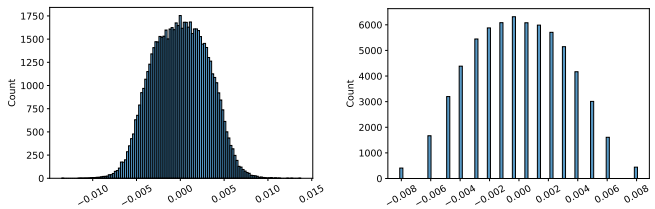


Fig. 6: The reconstruction quality of a light field image with different M and L values trained at a fixed epoch. At fixed L, the increasing M results in higher reconstruction quality; however, lower compression efficiency.



(a) before quantization (b) after quantization

Fig. 7: Histogram of weights of the second hidden layer of the $Bunny$ light field neural representation (a) before and (b) after quantization to 4 bits.

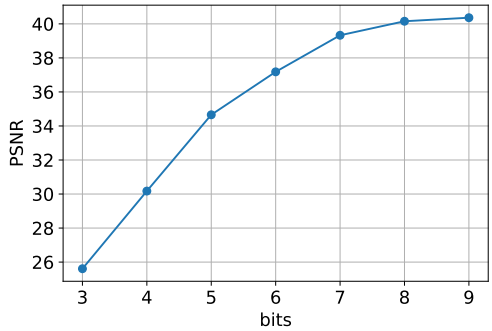


Fig. 8: Reconstruction quality of the neural representation trained for the $Bunny$ light field and quantized with different numbers of bits.

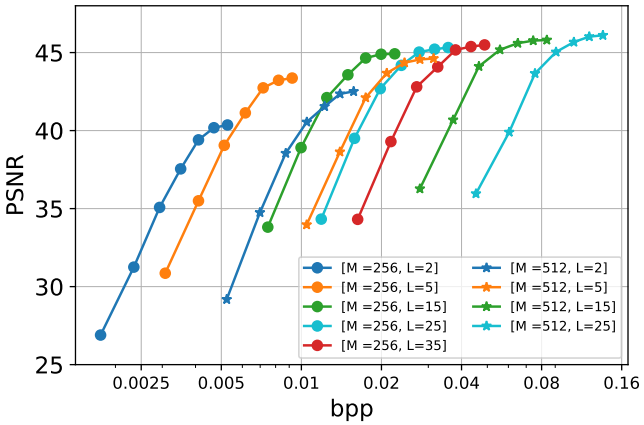
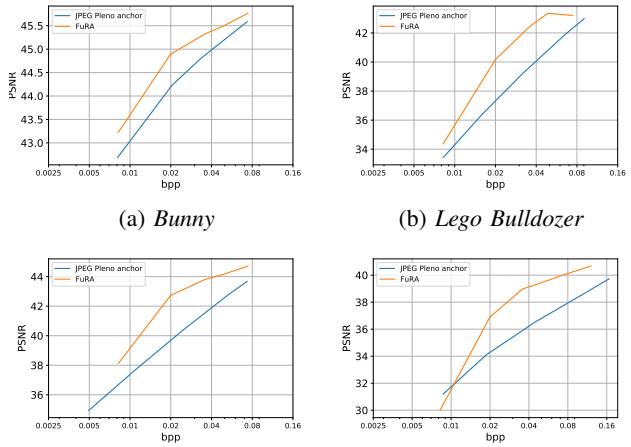


Fig. 9: Compression efficiency of neural representations for the $Bunny$ light field trained with various M and L values and quantized with a number of bits going from 3 to 9.

using a number of bits going from 3 to 9. The number of quantization bits can also be used as a rate controller to encode a neural representation at different bitrates. To further compress the trained neural network, Huffman encoding is deployed. Huffman encoding uses variable-length codewords to losslessly encode k clusters [25].

Fig. 9 shows the compression efficiency of $Bunny$ light field neural representations trained at various M and L values and compressed with different numbers of quantization levels. Since the reconstruction quality of compressed neural representations with various M and L values saturates after



(a) *Bunny* (b) *Lego Bulldozer*

(c) *Eucalyptus Flowers* (d) *Tarot Cards*

Fig. 10: Rate-distortion curves in terms of PSNR vs. bpp for the proposed *FuRA* and JPEG Pleno HEVC anchor.

certain bitrate points, a convex-hull is formed to determine the optimal rate-distortion curve for each light field.

IV. Experimental Results

In this section, we present the implementation details and evaluate the performance of the proposed method.

Four light field images, namely *(i) Bunny*, *(ii) Lego Bulldozer*, *(iii) Eucalyptus Flowers*, and *(iv) Tarot Cards* were selected for the evaluation. JPEG Pleno HEVC anchor [1] was selected as the benchmark. For each light field image, an MLP is trained with 45 epochs at $M \in \{256, 512\}$, $L \in \{2, 5, 15, 25, 35\}$, and the quantization bits in the range of 3 to 9. After quantization, the network is trained 5 more epochs and then Huffman coding is applied. The rate metric is defined as the number of bits per pixel (bpp). That is, the size of compressed bitstream divided by the number of pixels in the whole light field. The rate-distortion curves for test light fields are shown in Fig. 10. The proposed method, *FuRA*, outperforms the JPEG Pleno HEVC anchor which exploits the inter-view redundancy among view images; however, *FuRA* favours random access to images and even to pixels.

V. Conclusion

Traditional light field image coding solutions typically use the inter-coding tool of a video codec to exploit the inter-view redundancy among view images. In this way, a higher compression efficiency is achieved, however at the cost of sacrificing the random access functionality, since view images are made dependent on each other. In this paper, we proposed *FuRA*, a fully random access light field image coding scheme which employs implicit light field neural representations. Neural representations translate positional information into color information. Thus, they provide access to color information of each pixel independently of other pixels. To achieve high compression efficiency, neural network compression techniques, namely quantization and Huffman coding are deployed. The experimental results reveal that the proposed method outperforms HEVC inter-coding, in terms of compression efficiency, while providing random access to view images and even to pixels.

Acknowledgement

The financial support of the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association are gratefully acknowledged. Christian Doppler Laboratory ATHENA: https://athena.itec.aau.at/. This work has also been in part supported by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM, and in part by the french ANR research agency in the context of the project DeepCIM.

References

[1] F. Pereira, C. Pagliari, E. da Silva, I. Tabus, H. Amirpour, M. Bernardo, and A. Pinheiro, "JPEG Pleno Light Field Coding Common Test Conditions v3.2," *Doc. ISO/IEC JTC*.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[3] S. Zhao and Z. Chen, "Light Field Image Coding via Linear Approximation Prior," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 4562–4566.

[4] H. Amirpour, M. Pereira, and A. Pinheiro, "High Efficient Snake Order Pseudo-Sequence Based Light Field Image Compression," in *2018 Data Compression Conference*, Mar. 2018, pp. 397–397, ISSN: 2375-0359.

[5] H. Amirpour, A. Pinheiro, M. Pereira, F. J. P. Lopes, and M. Ghanbari, "Efficient Light Field Image Compression with Enhanced Random Access," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 2, pp. 44:1–44:18, Mar. 2022.

[6] M. B. de Carvalho, M. P. Pereira, G. Alves, E. A. B. da Silva, C. L. Pagliari, F. Pereira, and V. Testoni, "A 4D DCT-Based Lenslet Light Field Codec," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 435–439.

[7] V. Avramelos, J. D. Praeter, G. V. Wallendael, and P. Lambert, "Random Access Prediction Structures for Light Field Video Coding with MV-HEVC," *Multimedia Tools and Applications*, pp. 1 – 21, 2020.

[8] M. Magnor and B. Girod, "Data Compression for Light-field Rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338–343, 2000.

[9] Cha Zhang and Jin Li, "Compression of Lumigraph with Multiple Reference Frame (MRF) Prediction and Just-in-time Rendering," in *Proceedings DCC 2000. Data Compression Conference*, 2000, pp. 253–262.

[10] S. Pratapa and D. Manocha, "RLFC: Random Access Light Field Compression using Key Views," *CoRR*, vol. abs/1805.06019, 2018.

[11] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in *Eur. Conf. on Computer Vision (ECCV)*, 2020, pp. 405–421.

[12] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelNeRF: Neural Radiance Fields from One or Few Images," *IEEE. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 4576–4585, 2021.

[13] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2021.

[14] Z. Wang, S. Wu, W. Xie, M. Chen, and V. Prisacariu, "NeRF−−: Neural Radiance Fields Without Known Camera Parameters," *arXiv preprint arXiv:2102.07064*, 2021.

[15] M. Boss, R. Braun, V. Jampani, J. Barron, C. Liu, and H. Lensch, "NeRD: Neural Reflectance Decomposition from Image Collections," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2021, pp. 12684–12694.

[16] P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. Barron, "NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis," *IEEE. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 7491–7500, 2021.

[17] X. Zhang, S. Fanello, Y. Tsai, T. Sun, T. Xue, R. Pandey, S. Orts, P. Davidson, C. Rhemann, P. Debevec, J. Barron, R. Ramamoorthi, and W. Freeman, "Neural Light Transport for Relighting and View Synthesis," *ACM Trans. on Graphics (TOG)*, vol. 40, pp. 1–17, 2020.

[18] X. Zhang, P. Srinivasan, B. Deng, P. Debevec, W. Freeman, and J. Barron, "NeRFactor: Neural Factorization of Shape and Reflectance Under an Unknown Illumination," *ACM Trans. on Graphics (TOG)*, vol. 40, no. 6, dec 2021.

[19] B. Y. Feng and A. Varshney, "SIGNET: Efficient Neural Representations for Light Fields," in *Proceedings of the International Conference on Computer Vision (ICCV 2021)*, 2021.

[20] M. Bemana, K. Myszkowski, H.-P. Seidel, and T. Ritschel, "X-Fields: Implicit Neural View-, Light- and Time-Image Interpolation," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2020)*, vol. 39, no. 6, 2020.

[21] Z. Li, L. Song, C. Liu, J. Yuan, and Y. Xu, "NeuLF: Efficient Novel View Synthesis with Neural 4D Light Field," in *arXiv:2105.07112v6*, Dec. 2021.

[22] B. Y. Feng and A. Varshney, "SIGNET: Efficient Neural Representation for Light Fields," 2021, pp. 14224–14233.

[23] R. Mishra, H. P. Gupta, and T. Dutta, "A Survey on Deep Neural Network Compression: Challenges, Overview, and Solutions," Tech. Rep. arXiv:2010.03954, arXiv, Oct. 2020, arXiv:2010.03954 [cs, eess] type: article.

[24] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," Tech. Rep. arXiv:1510.00149, arXiv, Feb. 2016, arXiv:1510.00149 [cs] type: article.

[25] J. van Leeuwen, "On the Construction of Huffman Trees," in *ICALP*, 1976.