A Novel Hierarchical Light Field Coding Scheme Based on Hybrid Stacked Multiplicative Layers and Fourier Disparity Layers for Glasses-Free 3D Displays

Joshitha Ravishankar†, Mansi Sharma†∗

Indian Institute of Technology Madras, Chennai 600036, India.

Abstract

This paper presents a novel hierarchical coding scheme for light fields based on transmittance patterns of low-rank multiplicative layers and Fourier disparity layers. The proposed scheme identifies multiplicative layers of light field view subsets optimized using a convolutional neural network for different scanning orders. Our approach exploits the hidden low-rank structure in the multiplicative layers obtained from the subsets of different scanning patterns. The spatial redundancies in the multiplicative layers can be efficiently removed by performing low-rank approximation at different ranks on the Krylov subspace. The intra-view and inter-view redundancies between approximated layers are further removed by HEVC encoding. Next, a Fourier disparity layer representation is constructed from the first subset of the approximated light field based on the chosen hierarchical order. Subsequent view subsets are synthesized by modeling the Fourier disparity layers that iteratively refine the representation with improved accuracy. The critical advantage of the proposed hybrid layered representation and coding scheme is that it utilizes not just spatial and temporal redundancies in light fields but efficiently exploits intrinsic similarities among neighboring sub-aperture images in both horizontal and vertical directions as specified by different predication orders. In addition, the scheme is flexible to

†These authors contributed equally to this work

∗Corresponding author

Email addresses: ee19d401@gmail.iitm.ac.in (Joshitha Ravishankar†), mansisharmaiitd@gmail.com, mansisharma@ee.iitm.ac.in (Mansi Sharma†)

Preprint submitted to Elsevier August 30, 2021
realize a range of multiple bitrates at the decoder within a single integrated system. The compression performance of the proposed scheme is analyzed on real light fields. We achieved substantial bitrate savings and maintained good light field reconstruction quality.

**Keywords:** Light field, Lossy compression, Layered 3D displays, Convolutional neural network, Krylov subspace, Low-rank approximation, Block Krylov singular value decomposition, Rate distortion optimization, Fourier transform, Fourier disparity layers

1. Introduction

The light field imaging and 3D display markets are growing rapidly. In recent years, we have seen autostereoscopic displays emerge as a potential alternative to stereoscopic 3D displays as it supports stereopsis and motion parallax from different viewing directions [1] [2] [3] [4]. Parallax-based and lenticular-based 3D display technology has still not reached up to the standards to simultaneously provide direction-dependent outputs without losing out on the resolution in reconstructing dense light fields. Emerging multi-layered light field displays provide continuous motion parallax, greater depth-of-field, and wider field-of-view, which are critical for reproducing realistic 3D vision. Tensor or multi-layered displays can accurately reproduce multi-view images or light fields simultaneously with high resolution using just a few light attenuating layers [5] [6] [7].

The typical structure of a multi-layered display is shown in Fig. 1. It consists of light-attenuating pixelized layers (e.g., LCD panels) stacked in front of a backlight. On each layer, the transmittance of pixels can be controlled independently by carrying out multiplicative operations. Fig. 2 illustrates the light rays which pass through different combinations of pixels in stacked layers depending on the viewing directions. Efficient representation and coding of light rays in multi-layered 3D displays are essential for adaptation on different auto-stereoscopic platforms.

Existing light field coding approaches are based on raw lenslet image [8] [9]
Figure 1: Structure of layered light field display

[10, 11, 12], geometry information [13, 14, 15, 16], scene content information [17], disparity information [14, 13, 15], epipolar plane image-based and multiplane image-based [19, 15, 19], view prediction based learning schemes [20, 21, 22, 23] or methods considering light field data as a pseudo video sequence [24, 25, 26, 27]. These compression approaches are not specifically designed for tensor or multi-layered displays.

Our previous work on light field coding exploits the spatial and temporal correlations among light field multiplicative layers [28]. We handled the inherent redundancies in light fields by approximating multiplicative layers in the image-based spatial domain. In this work, we propose a novel hierarchical coding scheme for light field compression based on a hybrid multiplicative layers [29] and Fourier disparity layers representation [30]. The current approach efficiently deals with elimination of spatial, temporal and non-linear redundancies present in the light field view subsets while working in an integrated spatial and Fourier domain representation. This is a generalized coding scheme applicable to a variety of autostereoscopic displays, in particular, useful for tensor or multi-layer light field displays [1, 2, 3, 4]. It offers much more bitrate savings and adaptability to different coding approaches and also achieves the goal of covering a range of multiple bitrates in a single unified system. Apart from supporting multi-view and layered 3D displays, our model can complement other light field coding schemes based on learning networks to support various bitrates as well [31, 32, 33, 22, 34, 23].

In the proposed coding scheme, the input light field is divided into view sub-
sets based on pre-defined Circular and Hierarchical prediction orders (Fig. 3). Three optimized multiplicative layers for each view subset using convolutional neural networks (CNN) are constructed. The key idea in our proposed coding scheme is to reduce the dimensionality of stacked multiplicative layers using the randomized Block-Krylov singular value decomposition (BK-SVD) [35]. Factorization derived from BK-SVD effectively exploits the high spatial correlation between the multiplicative layers and approximates the light field subsets for varying low ranks. Encoding of these low-rank approximated subset multiplicative layers using HEVC codec [36] is performed to eliminate intra-view and inter-view redundancies further. Thus, our scheme approximates multiplicative layers of target light field view subsets for multiple ranks and quantization parameters (QPs) in the first stage. The view subsets are then reconstructed from the decoded layers.

In the second stage, the processing of the entire approximated light field is done in the Fourier domain, following a hierarchical coding procedure. A Fourier Disparity Layer (FDL) calibration is performed to estimate disparity values and angular coordinates of each light field view [37]. These essential parameters provide additional information for the FDL construction and view prediction. They are then transmitted to the decoder as metadata. Next, we split the approximated light field into subsets as identified by four scanning or prediction orders. The first set of views are encoded and utilized for constructing the FDL representation. This FDL representation synthesizes the succeeding view subsets.
The remaining correlations between the prediction residue of synthesized views and the approximated subset are further eliminated by encoding the residual signal using HEVC. The set of views obtained from decoding the residual are employed to refine the FDL representation and predict the next subset of views with improved accuracy. This hierarchical procedure continues iterating until all light field views are coded. The critical advantages of the proposed hybrid layered representation and coding scheme are:

- Our scheme efficiently exploits spatial, temporal, and non-linear redundancies between adjacent sub-aperture images in the light field structure within a single integrated framework. The scheme leverages the benefits of hybrid multiplicative layers and Fourier disparity layers representation in our hierarchical coding and prediction model for gaining superior compression efficiency without compromising reconstruction quality. In the first stage, we process light field in the spatial domain and remove the intra-view and inter-view redundancies among subset multiplicative layers. In the second stage, while working in the Fourier domain, our scheme ensures to eliminate the non-linear redundancies among adjacent views in both horizontal and vertical directions that exhibit high similarities. Experiments with various real light fields following different scanning orders demonstrate superior compression performance of our proposed model.

- The scheme is versatile to realize a range of multiple bitrates within a single integrated system trained using few convolutional neural networks. This characteristic of the proposed model complements existing light field coding systems or methods which usually support only specific bitrates during compression using multiple networks. Our coding model is adaptable to support various computational, multi-view auto-stereoscopic platforms, table-top, head-mounted displays, or mobile platforms by optimizing the bandwidth for a given target bitrate.

A shorter version of this work has been accepted for publication at IEEE SMC 2021 [38]. The current journal version is an extension that elaborates on
two more light field scanning patterns for view subsets. It includes new detailed results and extensive analysis of the performance of the proposed light field coding model. The rest of this article is organized into four major sections. Section 2 describes various existing light field compression approaches and their shortcomings. The proposed layered representation and coding scheme for multi-view displays is discussed in Section 3 in detail. We have elaborated our experiments specifying the implementation, results, and analysis in Section 4. Lastly, the conclusion with comprehensive findings of our proposed scheme and implications of future work are presented in Section 5.

2. Related Works

The light field image or plenoptic image contains information about the intensity and direction of light rays in space [39, 40]. It involves a large volume
of data. Thus, the storage and transmission requirements for light fields are tremendous. Efficient compression techniques leveraging the data redundancy in both the spatial and angular domains of light fields are essential. Compression or coding techniques for light fields are often broadly classified into two categories: lenslet-based or approaches based on perspective sub-aperture images.

The direct compression using raw light field lenslet image captured by plenoptic camera exploits the spatial redundancy between the microimages [8, 9, 10, 11, 12]. This approach is usually based on the existing image/video coding standards, such as JPEG [11] or high efficiency video coding (HEVC) [42]. The unique light field structural characteristics make it harder to predict the lenslet image regions with complex textures accurately. Moreover, there are numerous microimages in the raw light field image which need careful handling. The microimages are of low resolution and require customized reshaping before being fed into an HEVC encoder. Ideally, the lenslet-based compression solutions need to transmit camera parameters for further processing, and thus increase the coding burden on the compressed data stream. These drawbacks encouraged researchers to extract the light field sub-aperture images (SAIs) from raw plenoptic images and explore compression possibilities on these views.

There are various coding methods that directly take light field SAIs as the input. They are broadly categorized as content-based compression [17], disparity-based [14, 13, 18], epipolar plane image-based and multiplane image-based [10, 15, 16], pseudo sequence based [24, 43, 25, 26, 27], view synthesis based [44, 45, 20, 21], and learning-based compression methods [31, 32, 33, 22, 34, 23].

Disparity-based compression methods approximate particular views by the weighted sum of other views [14] or apply the homography-based low-rank approximation method called HLRA [13]. This approach depends on how much the disparities across views vary and may not optimally reduce the low-rank approximation error for light fields with large baselines. Dib et al. [18] proposed a novel parametric disparity estimation method to support the low-rank approximation using super rays. They efficiently exploit redundancy across the different views compared to the homography-based alignment. A shearlet
transform-based method presented in [15] categorizes the SAIs into key and
decimated views. The scheme performs well under low bitrates. The decimated
views are predicted from the compressed key views and a residual bitstream.
Chen et al. [16] used multiplane representation and strongly reduced the calculation burden on the decoder.

The views are reordered as pseudo sequences in predictive compression methods. Liu et al. [24] in their symmetric 2D hierarchical order, compresses the central view first as the I-frame (intra frame) followed by the remaining views as P-frames. Their prediction structure ensures inter-view prediction from adjacent views. Another proposal by Li et al. [13] involves dividing all the light field views into four quadrants and adopting a hierarchical coding structure within each quadrant. Ahmad et al. [25] utilize the multi-view extension of HEVC (MV-HEVC) to compress the light field in the form of a multi-view sequence. They interpret each row of the sub-aperture views as a single view of a multi-view video sequence and propose a two-dimensional prediction and rate allocation scheme. The extension of their work [26] uses hierarchical levels, where views belonging to higher levels are assigned with better quality. The higher-level views are used as references to predict the lower-level views. Synthesized virtual reference frames are generated from Adaptive Separable Convolution Network (ASCN) in another SAIs based technique [27]. Such frames are considered as extra reference candidates in a hierarchical coding structure for MV-HEVC to further exploit intrinsic similarities in light field images.

A block-basis estimation of views from translated reference views is proposed in [21]. The residuals of estimated views are also transmitted to the decoder along with the rest of the view estimation parameters. Other view synthesis compression schemes estimate image depth maps from a subset of reference light field views. The Multi-view Video plus Depth (MVD) structure is adopted for depth image-based rendering to synthesize the intermediate SAIs [15]. A pair of steps to generate noise-refined depth maps for selected perspective views is elaborated in [20].

Bakir et al. [31] presented a deep learning-based scheme on the decoder side
Figure 4: Three components of proposed light field coding scheme. (a) Overview of the encoding scheme (b) Overview of the decoding scheme.
to improve the reconstruction quality of sub-aperture images. Similarly, Zhao et al. [32] encoded only sparsely sampled SAIs, while the remaining SAIs are synthesized using a CNN from the decoded sampled SAIs as priors. However, these methods require large-scale and diverse training samples, and high quality of reconstructed views is only obtained if at least half the SAIs are taken as reference. Schiopu et al. [22] proposed a novel network that synthesizes the entire light field image as an array of synthesized macro pixels in one step. Wang et al. [33] identify a region of interest (ROI), a complex non-ROI, and a smooth non-ROI to compress light field videos framewise. A generative adversarial network (GAN) is proposed in [34] for unsampled SAIs generation. Liu et al. [23] also adopt a GAN framework to boost the light field compression. They use an image group-based sampling method to reduce more SAI redundancy and maintain the reconstructed SAI quality. A perceptual quality-based loss function is also proposed by considering the PSNR of synthetic SAIs and adversarial loss. The above-mentioned coding techniques are not explicitly designed for the representation used in multi-layer-based light field displays. They also usually train a system (or network) to support only specific bitrates during the compression.

3. Proposed Methodology

The complete workflow of our proposed representation and coding scheme with three main components is illustrated in Fig. 4. In COMPONENT I, input light field images are divided into view subsets depending on the specific prediction orders. To efficiently exploit the intrinsic redundancies in light field data, the proposed scheme constructs multiplicative layers from each view subset by employing a CNN. BLOCK I represents a CNN that converts the light field views of the input subsets into three multiplicative layers [29]. In BLOCK II, we removed the intrinsic redundancy present in subset views by exploiting the hidden low-rank structure of multiplicative layers on a Krylov subspace [35]. The low-rank approximation of each subset multiplicative layers is performed using Block-Krylov singular value decomposition (BK-SVD) by choosing various
ranks [35]. The intra-frame and inter-frame redundancies are further eliminated by encoding the approximated layers with HEVC [36]. We reconstruct the approximated views of each subset from their respective decoded multiplicative layers in BLOCK III. At the end of COMPONENT I of our scheme, we obtain approximated light field data at various ranks and quantization parameters.

In the next second phase, approximated subsets are used to construct Fourier disparity layer (FDL) representation of light fields [30]. The processing of approximated light field in COMPONENT II of our scheme is shown in Fig. 4a. There exist non-linear correlations between neighboring sub-aperture views in both horizontal and vertical directions in the light field structure. We particularly target these redundancies between adjacent light field views by processing in the Fourier domain as specified by different scanning or predication orders. The light field is then iteratively reconstructed by the FDL representation in a hierarchical fashion. We find the angular coordinates and disparity values of each view of the low-rank approximated light field (at different ranks and QPs) through FDL calibration and directly transmit them to the decoder (COMPONENT III) as metadata [37]. The approximated light field is divided into subsets specified by four different prediction orders. The initial subset is used in construction of the FDL representation, which is further employed to synthesize the subsequent subsets of views. The correlations in prediction residue are removed, and a more accurate FDL representation is constructed from previously encoded subsets. Thus, we iteratively refine the FDL representation in COMPONENT II until all the approximated light field views are encoded. The decoding scheme is depicted in COMPONENT III (Fig. 4b). Here, angular coordinates and disparity values of the low-rank approximated light field, along with the encoded bitstreams of approximated subsets are utilized for the final light field reconstruction.

3.1. Approximation of light field at different ranks

The input light field is divided into view subsets in COMPONENT I, depending on the specified scanning orders, and a low-rank approximated repre-
sentation of the subset multiplicative layers is obtained in the Krylov subspace. BLOCK I to III in Fig. 4 illustrate the three sub-blocks involved in this step. The details of each step are described in the following sub-sections.

3.1.1. View Subsets of Light Field

The proposed scheme divides the light field into different view subsets based on predefined scanning orders. We adopt a hierarchical pattern configuration and a circular view prediction order described in the work of Dib et al. The four chosen patterns are Circular-2 ($C_2$), Circular-4 ($C_4$), Hierarchical-2 ($H_2$) and Hierarchical-4 ($H_4$). For a $9 \times 9$ light field, the $C_2$ and $H_2$ patterns contain two view subsets and $C_4$ and $H_4$ patterns have four subsets. The exact coding orders of each subset of these four chosen scanning orders are shown in Fig. 3. In all subsets of these patterns, the light field views form a circle that spiral out from the center. Generally, the corner views of light fields are of lower quality, and thus we choose to form a circle rather than a square while scanning the views. Our proposed workflow begins with partitioning of the input light field into subsets based on $C_2$, $C_4$, $H_2$ or $H_4$ patterns (Fig. 4).

3.1.2. Light Field Views to Stacked Multiplicative Layers

The light field $L(s, t, u, v)$ characterizes radiance along light rays as a 4D function. In $L(s, t, u, v)$, the parameters $(s, t)$ specify the spatial coor-
Figure 6: The three optimal multiplicative layers produced by the CNN for the Dragon and Bunnies light field [46].

dinates and \((u, v)\) denote the angular coordinates that represent intersection of the rays with two parallel planes (Fig. 5). A light field can be interpreted as a set of directional views and can be produced using stacked layers that carry out multiplicative light ray operations [29]. Such multiplicative layers could be realized with few light-attenuating panels stacked in equally spaced intervals in front of a backlight (Fig. 1).

The normalized intensity of a light ray emitted from light field display can be described as

\[
L_m(s, t, u, v) = \prod_{x \in X} T_x(u + xs, v + xt) \tag{1}
\]

where, \(T_x(u, v)\) is the transmittance of the layer present at disparity \(x\) and \(X\) denotes the set of disparities among directional views. We determined the multiplicative layers for each subset of different view scanning order. The experiments are performed by considering the light field display composed of three layers located at \(X = \{-1, 0, 1\}\).

The three multiplicative layers obtained from each subset are optimized using a data-driven CNN based approach. The network architecture is depicted in Fig. 4a. The objective is to optimize the layer patterns. Mathematically, it is expressed as

\[
\arg\min_{T_x | x \in X} \sum_{s, t, u, v} \| L(s, t, u, v) - L_m(s, t, u, v) \|_2^2 \tag{2}
\]

The optimized multiplicative layers obtained for \(5 \times 5\) Dragon and Bunnies light field is shown in Fig. [6].
3.1.3. Low-Rank Representation of Subset Multiplicative Layers on Krylov Subspace

The randomized Block-Krylov SVD method introduced by Cameron and Christopher can optimally achieve the low-rank approximation of a matrix within $(1 + \epsilon)$ of optimal for spectral norm error [35]. The algorithm quickly converges in $\tilde{O}(\sqrt{\frac{1}{\epsilon}})$ iterations for any matrix. In our proposed scheme, the optimized multiplicative layers are compactly represented on a Krylov subspace in order to remove the intrinsic redundancy.

We denote each multiplicative layer pattern obtained from the CNN as $T_x \in \mathbb{R}^{m \times n \times 3}$, where $x \in \{-1, 0, 1\}$. The red, green, and blue colour channels of the layer $x$ are denoted as $T^r_x$, $T^g_x$, and $T^b_x$ respectively. We construct three matrices $A^{ch} \in \mathbb{R}^{3m \times n}$, $ch \in \{r, g, b\}$ as

$$A^{ch} = \begin{pmatrix} T^{ch}_{-1} \\ T^{ch}_0 \\ T^{ch}_1 \end{pmatrix}$$

The BK-SVD low-rank approximation in a Krylov subspace for each $A^{ch}$ can remove the intrinsic redundancies in multiplicative layers of the subsets. For simplicity, we will denote $A^{ch}$ as $A$ henceforth. To approximate $A$, the objective is to achieve a subspace that closely captures the variance of $A$’s top singular vectors and avoid the gap dependence in singular values. The spectral norm low-rank approximation error of $A$ is

$$\|A - DD^TA\|_2 \leq (1 + \epsilon) \|A - A_k\|_2$$

Only top $k$ singular vectors of $A$ are used in its rank-$k$ approximation $A_k$. If $D$ is a rank-$k$ matrix with orthonormal columns, the spectral norm guarantee ensures that $DD^TA$ recovers $A$ up to a threshold $\epsilon$.

Block Krylov iteration is performed working with the Krylov subspace

$$K = [I \ A \ A^2 \ I \ A^3 \ I \ \cdots \ A^q \ I]$$

As analyzed in [35], we choose to work on low degree polynomials that allow faster computation in fewer powers of $A$, and thus enabling convergence of the
BK-SVD algorithm in fewer iterations. From subspace $K$, we construct $p_q(A)\Pi$ for any polynomial $p_q(\cdot)$ of degree $q$, where $\Pi \sim N(0, 1)^{d \times k}$. The approximation of $A$ done by projecting it onto the span of $p_q(A)\Pi$ at least matches the best $k$ rank approximation of $A$ lying in the span of Krylov space $K$ \cite{35}. Further, we orthonormalize columns of $K$ to obtain $Q \in \mathbb{R}^{c \times qk}$ using QR decomposition method \cite{47}. After computing SVD of matrix $S = Q^T B B^T Q$, we found matrix $\bar{U}_k$ containing the top $k$ singular vectors of $S$. Thus, the rank-$k$ approximation of $A$ is matrix $D$, which is computed as

$$D = Q \bar{U}_k$$

Consequently, the rank-$k$ block Krylov approximation of matrices $A^r$, $A^g$, and $A^b$ are $D^r$, $D^g$, and $D^b$ respectively. Matrix $D^{ch} \in \mathbb{R}^{y \times z}$ for every colour channel $ch$. To obtain the approximated layers $\bar{T}_z$, we extract colour channels from the approximated $D^{ch}$ matrices by sectioning out the rows uniformly as

$$\bar{T}_{z}^{ch} = D^{ch}[1 : y, 1 : z]$$

$$\bar{T}_{z}^{ch} = D^{ch}[y + 1 : 2y, 1 : z]$$

$$\bar{T}_{z}^{ch} = D^{ch}[2y + 1 : 3y, 1 : z]$$

The red, green, and blue colour channels are combined to form each approximated layer $\bar{T}_{z}^{ch}$, $\bar{T}_0^{ch}$, and $\bar{T}_1^{ch}$. Thus, factorization derived from BK-SVD exploits the spatial correlation in multiplicative layers of the subset views for varying low ranks. The three block Krylov approximated layers for each subset are subsequently encoded into a bitstream using HEVC for various QPs to further eliminate intra and inter layer redundancies in the low-rank approximated representation. The low-rank representation and coding of stacked multiplicative layers on Krylov subspace is depicted in BLOCK II of Fig. 4a.

### 3.1.4. Decoding and Reconstruction of the Light Field Subsets

The decoding procedure of compressed layers and the reconstruction of light field subsets is shown in BLOCK III of Fig. 4a. We decoded three multiplicative layers from the bitstream as $\bar{T}_x$, $x \in \{-1, 0, 1\}$. We consider the integers $s^*, t^*$,
which vary depending on the number of views in each subset. The view $I_{(s^*, t^*)}$ is obtained by translating the decoded layers to $\hat{T}_x$. This step is followed by an element-wise product of colour channels of the translated layers. For a particular view $(s^*, t^*)$, the translation of every $x^{th}$ layer $\hat{T}_x$, to $\hat{T}_x$ is carried out as

$$\hat{T}_{x(s^*, t^*)}(u, v) = \hat{T}_x(u + xs^*, v + xt^*)$$

Thus, the three translated layers for every viewpoint $(s^*, t^*)$ is computed as

$$\hat{T}_{x-1(s^*, t^*)}(u, v) = \hat{T}_{x-1}(u - s^*, v - t^*)$$

$$\hat{T}_{0(s^*, t^*)}(u, v) = \hat{T}_0(u, v)$$

$$\hat{T}_{1(s^*, t^*)}(u, v) = \hat{T}_1(u + s^*, v + t^*)$$

An element-wise product of each colour channel $ch \in \{r, g, b\}$ of the translated layers give the corresponding colour channel of the subset view.

$$I_{(s^*, t^*)}^{ch} = \hat{M}_{-1(s^*, t^*)}^{ch} \odot \hat{M}_{0(s^*, t^*)}^{ch} \odot \hat{M}_{1(s^*, t^*)}^{ch}$$

The combined red, green and blue colour channels output the reconstructed light field subset at the viewpoint $(s^*, t^*)$ as $I_{(s^*, t^*)}$. At last, we merged the view subsets according to $C_2$, $C_4$, $H_2$ or $H_4$ patterns.

Thus, the spatial correlation in multiplicative layers of the subset views is exploited for different low ranks in COMPONENT I of the proposed scheme. Intra and inter-layer redundancies in the low-rank approximated representation are also well removed. We further compressed the light field by eliminating intrinsic similarities caused by non-linear correlations among neighboring views.
in horizontal and vertical directions as specified by $C_2$, $C_4$, $H_2$ and $H_4$ patterns using the Fourier Disparity Layers (FDL) representation. The following section describes the processing of approximated light fields in the Fourier domain.

3.2. Fourier Disparity Layers Representation & Light Field Processing

The Fourier Disparity Layers representation [30] samples the input BK-SVD approximated light field in the disparity dimension by decomposing it as a discrete sum of layers. The layers are constructed from the approximated light field sub-aperture views through a regularized least square regression performed independently at each spatial frequency in the Fourier domain. The FDL representation has been shown to be effective for numerous light field processing applications [30, 37, 48, 49]. We have summarised the use of FDL and encoding of low-rank approximated light field as COMPONENT II in Fig. 4a. The corresponding decoding and reconstruction of the light field subsets are illustrated as COMPONENT III in Fig. 4b. The Fourier Disparity Layer calibration, subset view synthesis, and prediction are described in the following subsections.

3.2.1. FDL Calibration

Without loss of generality, we have considered one spatial coordinate $s$ and one angular coordinate $u$ of 4D light field to present the notations in a simple manner. The approximated light field view $L_{uo}$ at angular position $u_0$ can be defined as $L_{uo}(s) = L(s, u_0)$. We can obtain Fourier coefficients of the $j^{th}$ input light field view using $n$ disparity values $\{d_k\}_{k \in [1,n]}$ [37]. The Fourier transform of $L_{uo}$ at spatial frequency $f_s$ is

$$\hat{L}_{uo}(f_s) = \sum_k e^{+2i\pi u_0 d_k f_s} \hat{L}^k(f_s)$$  \hspace{1cm} (9)

Here, the Fourier coefficients of the disparity layers for a particular frequency $f_s$ is defined by

$$\hat{L}^k(f_s) = \int_{\Omega_k} e^{-2i\pi f_s} L(s, 0) dx$$  \hspace{1cm} (10)

Each of such Fourier coefficients can be understood as the Fourier transform of the central light field view (as $u_o = 0$) by just considering the spatial region $\Omega_k$ of disparity $d_k$. 

17
The angular coordinates $u_j$ of the input views and the disparity values of the layers $d_k$ are estimated by minimizing over all frequency components $f_s^q$, $q \in \llbracket 1, Q \rrbracket$, where $Q$ is the number of pixels in each input image \textsuperscript{[30]}. By computing Fourier transforms of all $m$ approximated light field views as $\hat{L}_{u_j}(j \in \llbracket 1, m \rrbracket)$, the FDL representation is learnt by solving a linear regression problem for each frequency $f_s$.

The optimization problem is formulated as $Ax = b$ with Tikhonov regularization, where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^{n \times 1}$ and $b \in \mathbb{R}^{m \times 1}$. Elements of matrix $A$ are $A_{jk} = e^{+2i\pi u_j d_k f_s}$ and $x$ contains the Fourier coefficients of disparity layers $x_k = \hat{L}_k(f_s)$. The vector $b$ contains the Fourier coefficients of $j^{th}$ input view, $b_j = \hat{L}_{u_j}(f_s)$. The solution of the optimization problem results in the disparity values $d_k$ and view positions $u_j$ that are passed as metadata information to the decoder in COMPONENT III.

### 3.2.2. FDL View Synthesis and Prediction

In section \textsuperscript{3.1.1} two basic scanning orders, circular and hierarchical, are discussed for the synthesis and coding of light field views. Each of these chosen orders has two or four view subsets. This results in four patterns $C_2$, $C_4$, $H_2$ and $H_4$. In all four cases of the view prediction orders, the images are arranged in a spiral order starting from the center of the light field for each subset. The initial view subset of every pattern is always the first subset as specified by the
scanning order. This subset is directly encoded first in COMPONENT II. For example, in Fig. 8 the blue coloured subset in $C_2$ is the first subset.

The angular coordinates $u_j$ and disparity values $d_k$ are determined by the Fourier Disparity Layer calibration [30]. These are required in the further FDL construction and view predictions. This additional information is transmitted to the decoder in COMPONENT III as metadata [37]. The initial view subset is used in the basic construction of FDL representation. This aids in the synthesis of succeeding view subsets. The residual signal is also encoded to account for the remaining correlations in the prediction residue of synthesized views. The FDL representation is then refined before prediction and encoding of the next subset of views. Thus, the FDL representation is iteratively fine-tuned at every stage, after encoding every view subset, until all the approximated input light field views are encoded.

4. Results and Analysis

The performance of the proposed compression scheme is evaluated on real light fields captured by plenoptic cameras. The experiments are performed with Bikes, Fountain & Vincent 2, and Stone pillars outside light field datasets from the EPFL Lightfield JPEG Pleno database [50]. The central views of the chosen light field images are shown in Fig. 7. The raw plenoptic images are extracted
Figure 10: Bitrate vs YUV-PSNR curves for the proposed compression scheme and Dib et al. for the three datasets in the $C_2$, $C_4$, $H_2$ and $H_4$ patterns. *(Kindly expand for better clarity)*
into 15×15 sub-aperture views using the Matlab Light field toolbox [51].

The patterns $C_2$, $C_4$, $H_2$ and $H_4$ are constructed from inner 9×9 light field views for our experiments. Subsets 1 and 2 of $C_2$ contain 24 and 57 light field views respectively. The first and second subsets $H_2$ contain 25 and 56 views respectively. In $C_4$, subsets 1, 2, 3 and 4 have 4, 16, 12 and 49 views respectively. Lastly, subsets 1, 2, 3 and 4 of $H_4$ contain 4, 5, 16 and 56 light field views respectively. The exact scanning orders of the patterns and their subsets are specified in Fig. 3.

4.1. Experimental Settings and Implementation Details

The proposed scheme is implemented on a single high-end HP OMEN X 15-DG0018TX system with 9th Gen i7-9750H, 16 GB RAM, RTX 2080 8 GB Graphics, and Windows 10 operating system. The multiplicative layers for each view subset of the four scanning orders are optimized using their corresponding convolutional neural network (CNN). The CNN contains twenty 2D convolutional layers stacked in a sequence. They have constant spatial size of tensors throughout and only the number of channels is varied. The first layer in each CNN corresponds to the tensor $L$, which contains all the pixels of subset $L(u,v,s,t)$ being handled. $L$ has number of channels equal to the number of viewpoints in respective subset and tensor $T$ (that contained all the pixels of $T_x(u,v)$) has 3 channels corresponding to the 3 light field subset multiplicative layers. Intermediate convolutional layers comprise of 64 channels each. In Fig. 4a, ‘ch’ refers to the channels in the convolutional layers. We train each CNN model for 25 epochs at a learning rate of 0.001 and a batch size of 15. The entire networks are implemented using the Python-based framework, Chainer (version 7.7.0).

The resultant output multiplicative layers for each subset of $C_2$, $C_4$, $H_2$ and $H_4$ patterns are obtained from the trained CNN models. Fig. 8 illustrates the three multiplicative layers produced for each view subset of Circular-2 and Circular-4 patterns in the Fountain & Vincent 2 data. The three multiplicative layers of each subset of Hierarchical-2 and Hierarchical-4 are shown in Fig. 9 for
Table 1: Total number of bytes for each subset of Circular-2 pattern using our proposed coding scheme and Dib et al. [37].

| QP | Scheme | Bikes | Stone pillars outside | Furman & Vincent 2 |
|----|--------|-------|----------------------|---------------------|
|    |        | Subset 1 | Subset 2 | Subset 1 | Subset 2 |
| 2  |        |        |        |        |        |
| 4  |        |        |        |        |        |
| 6  |        |        |        |        |        |
| 8  |        |        |        |        |        |
| 10 |        |        |        |        |        |
| 12 |        |        |        |        |        |
| 14 |        |        |        |        |        |
| 16 |        |        |        |        |        |
| 18 |        |        |        |        |        |
| 20 |        |        |        |        |        |
| 22 |        |        |        |        |        |

**Notes:**
- **Bikes** refers to the number of bytes for each subset of the Circular-2 pattern.
- **Stone pillars outside** and **Furman & Vincent 2** columns show the number of bytes for each scheme.

- **Proposed (Rank 4):** The proposed coding scheme ranked 4th.
- **Proposed (Rank 8):** The proposed coding scheme ranked 8th.
- **Proposed (Rank 16):** The proposed coding scheme ranked 16th.
- **Proposed (Rank 32):** The proposed coding scheme ranked 32nd.
- **Proposed (Rank 64):** The proposed coding scheme ranked 64th.

Data extracted from the image.
Table 2: Total number of bytes for each subset of Hierarchical-2 pattern using our proposed coding scheme and Dib et al. [37].

| Subtuples | Bike | Stone pillars outside | Function & Vincent 2 |
|-----------|------|----------------------|----------------------|
|           |      | Subset 1 | Subset 2 | Subset 1 | Subset 2 | Subset 1 | Subset 2 | Subset 1 | Subset 2 | Subset 1 | Subset 2 |
|           |      | Proposed (Rank 4) | Proposed (Rank 8) | Proposed (Rank 16) | Proposed (Rank 4) | Proposed (Rank 8) | Proposed (Rank 16) | Proposed (Rank 4) | Proposed (Rank 8) | Proposed (Rank 16) | Proposed (Rank 4) | Proposed (Rank 8) | Proposed (Rank 16) |
|           |      | 287572 | 241398 | 57194 | 51268 | 287572 | 241398 | 57194 | 51268 | 287572 | 241398 | 57194 | 51268 |
|           |      | 240605 | 209976 | 51268 | 51268 | 240605 | 209976 | 51268 | 51268 | 240605 | 209976 | 51268 | 51268 |
|           |      | 253802 | 213780 | 51268 | 51268 | 253802 | 213780 | 51268 | 51268 | 253802 | 213780 | 51268 | 51268 |
|           |      | 257220 | 213780 | 51268 | 51268 | 257220 | 213780 | 51268 | 51268 | 257220 | 213780 | 51268 | 51268 |

...
Table 3: Total number of bytes for each subset of Circular-4 pattern using our proposed coding scheme and Dib et al. \[37\].

| Subset | Bikes | Stone pillars outside | Fountain A | Fountain B | Fountain C | Fountain D |
|--------|-------|-----------------------|------------|------------|------------|------------|
| 1      | Proposed (Rank 60) | 18900 | 49014 | 76951 | 50779 | 128501 |
|        | Proposed (Rank 44) | 18900 | 49014 | 76951 | 50779 | 128501 |
|        | Proposed (Rank 28) | 18900 | 49014 | 76951 | 50779 | 128501 |
| 2      | Proposed (Rank 60) | 41902 | 13999 | 5300 | 3500 | 1249 |
|        | Proposed (Rank 44) | 41902 | 13999 | 5300 | 3500 | 1249 |
|        | Proposed (Rank 28) | 41902 | 13999 | 5300 | 3500 | 1249 |
| 3      | Proposed (Rank 60) | 202694 | 37340 | 202694 | 37340 | 202694 |
|        | Proposed (Rank 44) | 202694 | 37340 | 202694 | 37340 | 202694 |
|        | Proposed (Rank 28) | 202694 | 37340 | 202694 | 37340 | 202694 |
| 4      | Proposed (Rank 60) | 3500 | 1249 | 41902 | 13999 | 5300 |
|        | Proposed (Rank 44) | 3500 | 1249 | 41902 | 13999 | 5300 |
|        | Proposed (Rank 28) | 3500 | 1249 | 41902 | 13999 | 5300 |

\[QP\] Scheme

Bikes

Stone pillars outside

Fountain A

Fountain B

Fountain C

Fountain D

Subset 1

Subset 2

Subset 3

Subset 4

Subset 1

Subset 2

Subset 3

Subset 4
Table 4: Total number of bytes for each subset of Hierarchical-4 pattern using our proposed coding scheme and Dib et al. [37].

| Subset | 1st | 2nd | 3rd | 4th |
|--------|-----|-----|-----|-----|
| Proposed (Rank 4) | 65931 | 62419 | 75760 | 62419 |
| Proposed (Rank 5) | 65931 | 62419 | 75760 | 62419 |
| Dib et al. | 65931 | 62419 | 75760 | 62419 |

Dib et al. [37]
Table 5: Bjontegaard percentage rate savings for the proposed compression scheme with respect to Dib et al. [37]. Negative values represent gains.

| Scene                    | Circular-2 | Hierarchical-2 | Circular-4 | Hierarchical-4 |
|--------------------------|------------|----------------|------------|---------------|
|                          | Subset 1   | Subset 2       | Subset 1   | Subset 2      |
| Bikes                     | 4          | -91.07697181   | -97.56541214 | -85.20018338 |
|                          | 8          | -95.066064     | -99.5369851  | -98.8452873  |
|                          | 12         | -91.9260063    | -97.9596853  | -98.2198153  |
|                          | 20         | -95.6475692    | -97.1956555  | -98.1077824  |
|                          | 40         | -94.3141049    | -94.3310249  | -97.0249826  |
|                          | 60         | -86.4629531    | -92.9321527  | -95.6279831  |
|                          |            |                |            |               |
| Stone pillars outside     | 1          | -98.74446465   | -99.7373088  | -98.9617927  |
|                          | 8          | -98.90331735   | -98.83289017 | -99.57099646 |
|                          | 16         | -98.9540372    | -98.83289017 | -99.57099646 |
|                          | 28         | -99.09494577   | -97.88383496 | -97.93405439 |
|                          | 44         | -99.69494577   | -97.88383496 | -97.93405439 |
|                          |            |                |            |               |
| Fountain & Vincent        | 1          | -70.2713249    | -93.3380536  | -87.7742571  |
|                          | 8          | -82.1462435    | -88.3314487  | -83.1595801  |
|                          | 16         | -87.949514     | -88.3314487  | -83.1595801  |
|                          | 28         | -84.445729     | -88.3314487  | -83.1595801  |
|                          | 44         | -81.3749531    | -88.3314487  | -83.1595801  |

the Bikes data. We rearranged the colour channels of these multiplicative layers as described in section 3.1.3 and then applied BK-SVD for ranks 4, 8, 16, 28, 44, 52 and 60. The approximated matrices are then rearranged back into layers and compressed using HEVC (BLOCK II of COMPONENT I). We use quantization parameters 2, 6, 10, 14, 20, 26, and 38 to test both high and low bitrate cases of HEVC. The decoding and reconstruction of BK-SVD approximated subsets for all four patterns are then performed.

Low-rank approximated subsets are then utilized to form the FDL representation of light fields. The number of layers in the FDL method are fixed to \( n = 30 \). Views in approximated Subset 1 construct the initial FDL representation. The subsequent view subsets are predicted from this FDL representation and the residues iteratively refine the FDL representation. We used HEVC to perform the encoding in COMPONENT II, choosing quantization parameters 2, 6, 10, 14, 20, 26, and 38. The final reconstructed light field central views at the end of COMPONENT III for the Stone pillars outside data is illustrated in Fig. 11 for the BK-SVD ranks 28, 44 and 60.
4.2. Results and Comparative Analysis

We compare the proposed scheme with Dib et al. [37] light field coding algorithm. The proposed coding scheme outperforms by large margins. The total number of bytes taken by our approach is comparatively far lesser than the work by Dib et al. [37] for all ranks and QPs. The corresponding results are depicted in Table 1 for $C_2$, Table 2 for $H_2$, Table 3 for $C_4$ and Table 4 for $H_4$ patterns. The bitrate vs YUV-PSNR graphs of three datasets in $C_2$, $C_4$, $H_2$ and $H_4$ configurations are illustrated in Fig. 10. For all four scanning patterns, the proposed scheme has significantly better rate-distortion results considering both subset-wise light field and entire light field.

Further, we analyze bitrate reduction (BD-rate) of the proposed scheme with respect to Dib et al. [37] using Bjontgaard metric [52]. The average percent difference in rate change is estimated over a range of QPs for seven chosen ranks. A comparison of the percentage of bitrate savings of our proposed coding scheme with respect to the anchor method for three chosen light field datasets is shown in Table 5. For $C_2$ pattern, the proposed scheme achieves 94.53%, 96.39%, and 93.96% bitrate reduction compared to Dib et al. [37] for light fields Bikes, Stone pillars outside, and Fountain & Vincent 2 respectively. For pattern $H_2$, the proposed scheme achieves 97.23%, 97.54%, and 96.67% bitrate reduction compared to Dib et al. [37] for the light fields Bikes, Stone pillars outside, and Fountain & Vincent 2 respectively. Lastly, the $C_4$ and $H_4$ patterns have 93.09%, 95.29%, 90.71%, and 93.18%, 96.02%, 91.25% bitrate reduction respectively in the proposed scheme over Dib et al. [37] for Bikes, Stone pillars outside, and Fountain & Vincent 2.

5. Conclusion

We have proposed a novel hierarchical hybrid coding scheme for light fields based on transmittance patterns of low-rank multiplicative layers and Fourier Disparity Layers. Typical pseudo sequence based light field compression schemes [24, 25, 26, 27] do not efficiently consider the similarities between horizontal and
vertical views of a light field. Our proposed scheme not only exploits the spatial correlation in multiplicative layers of the subset views for varying low ranks, but also removes the temporal intra and inter-layer redundancies in the low-rank approximated representation of view subsets. The approximated light field is further compressed by eliminating intrinsic similarities among neighboring views of Circular-2, Circular-4, Hierarchical-2 and Hierarchical-4 patterns using Fourier Disparity Layers representation. This integrated compression achieves excellent bitrate reductions without compromising the quality of the reconstructed light field.

Our scheme offers flexibility to cover a range of multiple bitrates using just a few trained CNNs to obtain a layered representation of the light field subsets. This critical feature sets our proposed scheme apart from other existing light field coding methods, which usually train a system (or network) to support only specific bitrates during the compression. Besides, existing coding approaches are not explicitly designed to target layered displays and are usually only classified
to work for lenslet-based formats or sub-aperture images based pseudo-sequence light field representation. The proposed flexible coding scheme can support not just multi-layered light field displays, but is also adaptable to table-top or other variety of autostereoscopic displays.

In our future work, we plan to extend the proposed idea to light field displays with more than three light attenuating layers. Proof-of-concept experiments with our scheme also pave the way to form a more profound understanding in the rank-analysis of a light-field using other mathematically valid tensor-based models and coded mask cameras. We also aim to verify our scheme with physical light field display hardware. Further, we wish to adapt the proposed scheme with display device availability and optimize the bandwidth for a target bitrate. This would enable deploying the concepts of layered displays on different display platforms that deliver 3D contents with limited hardware resources and thus best meet the viewers’ preferences for depth impression and visual comfort.

Author contributions

Joshitha Ravishankar: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing- Original Draft, Visualization. Mansi Sharma: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing- Original Draft, Supervision, Project administration.

Acknowledgement

The scientific efforts leading to the results reported in this paper have been carried out under the supervision of Dr. Mansi Sharma, INSPIRE Hosted Faculty, IIT Madras. This work has been supported, in part, by the Department of Science and Technology, Government of India project “Tools and Processes for Multi-view 3D Display Technologies”, DST/INSPIRE/04/2017/001853.
We would like to thank Nikitha Varma Sunchu and P Sai Shankar Pavan Srinivas for their help in running the codes for experiments.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Surman, X. W. Sun, Towards the reality of 3d imaging and display, in: 2014 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), IEEE, 2014, pp. 1–4.

[2] T. Balogh, P. T. Kovács, A. Barsi, Holovizio 3d display system, in: 2007 3DTV Conference, IEEE, 2007, pp. 1–4.

[3] Y. Kubota, N. Umezu, Pseudo-3d display based on head tracking for viewpoint image generation, in: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), 2019, pp. 158–161. doi:10.1109/GCCE46687.2019.9015351

[4] P. Surman, X. Zhang, W. Song, X. Xia, S. Wang, Y. Zheng, Glasses-free 3-d and augmented reality display advances: From theory to implementation, IEEE MultiMedia 27 (1) (2020) 17–26. doi:10.1109/MMUL.2019.2948334

[5] T. Li, Q. Huang, S. Alfaro, A. Supikov, J. Ratcliff, G. Grover, R. Azuma, Light-field displays: a view-dependent approach, in: ACM SIGGRAPH 2020 ET, 2020, pp. 1–2.

[6] G. Wetzstein, D. R. Lanman, M. W. Hirsch, R. Raskar, Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting.
[7] M. Sharma, S. Chaudhury, B. Lall, A novel hybrid kinect-variety-based high-quality multiview rendering scheme for glass-free 3D displays, IEEE TCSVT 27 (10) (2016) 2098–2117.

[8] Y. Li, M. Sjöström, R. Olsson, U. Jennehag, Efficient intra prediction scheme for light field image compression, in: 2014 IEEE International conference on acoustics, speech and signal processing (ICASSP), IEEE, 2014, pp. 539–543.

[9] C. Perra, P. Assuncao, High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement, in: 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, 2016, pp. 1–4.

[10] Y. Li, R. Olsson, M. Sjöström, Compression of unfocused plenoptic images using a displacement intra prediction, in: 2016 IEEE ICMEW, IEEE, 2016, pp. 1–4.

[11] R. J. Monteiro, P. J. Nunes, N. M. Rodrigues, S. M. Faria, Light field image coding using high-order intrablock prediction, IEEE JSTSP 11 (7) (2017) 1120–1131.

[12] D. Liu, P. An, R. Ma, W. Zhan, X. Huang, A. A. Yahya, Content-based light field image compression method with gaussian process regression, IEEE TM 22 (4) (2019) 846–859.

[13] X. Jiang, M. Le Pendu, R. A. Farrugia, C. Guillemot, Light field compression with homography-based low-rank approximation, IEEE JSTSP 11 (7) (2017) 1132–1145.

[14] S. Zhao, Z. Chen, Light field image coding via linear approximation prior, in: IEEE ICIP, IEEE, 2017, pp. 4562–4566.

[15] W. Ahmad, S. Vagharshakyan, M. Sjöström, A. Gotchev, R. Bregovic, R. Olsson, Shearlet transform-based light field compression under low bitrates, IEEE TIP 29 (2020) 4269–4280.
[16] Y. Chen, P. An, X. Huang, C. Yang, D. Liu, Q. Wu, Light field compression using global multiplane representation and two-step prediction, IEEE SPL 27 (2020) 1135–1139.

[17] X. Hu, J. Shan, Y. Liu, L. Zhang, S. Shirmohammadi, An adaptive two-layer light field compression scheme using gn-based reconstruction, ACM TOMM 16 (2s) (2020) 1–23.

[18] E. Dib, M. Le Pendu, X. Jiang, C. Guillemot, Local low rank approximation with a parametric disparity model for light field compression, IEEE TIP 29 (2020) 9641–9653.

[19] S. Vagharshakyan, R. Bregovic, A. Gotchev, Light field reconstruction using shearlet transform, IEEE transactions on pattern analysis and machine intelligence 40 (1) (2017) 133–147.

[20] X. Huang, P. An, F. Cao, D. Liu, Q. Wu, Light-field compression using a pair of steps and depth estimation, Optics express 27 (3) (2019) 3557–3573.

[21] B. Hériard-Dubreuil, I. Viola, T. Ebrahimi, Light field compression using translation-assisted view estimation, in: 2019 PCS, IEEE, 2019, pp. 1–5.

[22] I. Schiopu, A. Munteanu, Deep-learning-based macro-pixel synthesis and lossless coding of light field images, APSIPA TSIP 8.

[23] D. Liu, X. Huang, W. Zhan, L. Ai, X. Zheng, S. Cheng, View synthesis-based light field image compression using a generative adversarial network, Information Sciences 545 (2021) 118–131.

[24] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, W. Zeng, Pseudo-sequence-based light field image compression, in: IEEE ICMEW, IEEE, 2016, pp. 1–4.

[25] W. Ahmad, R. Olsson, M. Sjöström, Interpreting plenoptic images as multi-view sequences for improved compression, in: IEEE ICIP, IEEE, 2017, pp. 4557–4561.
[26] W. Ahmad, M. Ghafoor, S. A. Tariq, A. Hassan, M. Sjöström, R. Olsson, Computationally efficient light field image compression using a multiview hevc framework, IEEE access 7 (2019) 143002–143014.

[27] J. Gu, B. Guo, J. Wen, High efficiency light field compression via virtual reference and hierarchical mv-hevc, in: IEEE (ICME), IEEE, 2019, pp. 344–349.

[28] J. Ravishankar, M. Sharma, P. Gopalakrishnan, A flexible coding scheme based on block krylov subspace approximation for light field displays with stacked multiplicative layers, Sensors 21 (13) (2021) 4574.

[29] K. Maruyama, K. Takahashi, T. Fujii, Comparison of layer operations and optimization methods for light field display, IEEE Access 8 (2020) 38767–38775.

[30] M. Le Pendu, C. Guillemot, A. Smolic, A fourier disparity layer representation for light fields, IEEE TIP 28 (11) (2019) 5740–5753.

[31] N. Bakir, W. Hamidouche, O. Déforges, K. Samrouth, M. Khalil, Light field image compression based on convolutional neural networks and linear approximation, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 1128–1132.

[32] Z. Zhao, S. Wang, C. Jia, X. Zhang, S. Ma, J. Yang, Light field image compression based on deep learning, in: 2018 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2018, pp. 1–6.

[33] B. Wang, Q. Peng, E. Wang, K. Han, W. Xiang, Region-of-interest compression and view synthesis for light field video streaming, IEEE Access 7 (2019) 41183–41192.

[34] C. Jia, X. Zhang, S. Wang, S. Wang, S. Ma, Light field image compression using generative adversarial network-based view synthesis, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9 (1) (2018) 177–189.
[35] C. Musco, C. Musco, Randomized block krylov methods for stronger and faster approximate singular value decomposition, arXiv preprint arXiv:1504.05477.

[36] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (hevc) standard, IEEE TCSVT 22 (12) (2012) 1649–1668.

[37] E. Dib, M. Le Pendu, C. Guillemot, Light field compression using fourier disparity layers, in: IEEE ICIP, IEEE, 2019, pp. 3751–3755.

[38] J. Ravishankar, M. Sharma, A hierarchical coding scheme for glasses-free 3d displays based on scalable hybrid layered representation of real-world light fields, CoRR abs/2104.09378. arXiv:2104.09378 URL https://arxiv.org/abs/2104.09378

[39] M. Levoy, P. Hanrahan, Light field rendering, in: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996, pp. 31–42.

[40] S. J. Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen, The lumigraph, in: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996, pp. 43–54.

[41] W. B. Pennebaker, J. L. Mitchell, JPEG: Still image data compression standard, Springer Science & Business Media, 1992.

[42] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (hevc) standard, IEEE Transactions on circuits and systems for video technology 22 (12) (2012) 1649–1668.

[43] L. Li, Z. Li, B. Li, D. Liu, H. Li, Pseudo-sequence-based 2-d hierarchical coding structure for light-field image compression, IEEE Journal of Selected Topics in Signal Processing 11 (7) (2017) 1107–1119.
[44] T. Senoh, K. Yamamoto, N. Tetsutani, H. Yasuda, Efficient light field image coding with depth estimation and view synthesis, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1840–1844.

[45] X. Huang, P. An, L. Shan, R. Ma, L. Shen, View synthesis for light field coding using depth estimation, in: 2018 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2018, pp. 1–6.

[46] G. Wetzstein, Synthetic light field archive - mit media lab, https://web.media.mit.edu/~gordonw/SyntheticLightFields/, accessed: 01-04-2021.

[47] M. Gu, S. C. Eisenstat, Efficient algorithms for computing a strong rank-revealing qr factorization, SIAM Journal on Scientific Computing 17 (4) (1996) 848–869.

[48] M. Le Pendu, C. Ozcinar, A. Smolic, Hierarchical fourier disparity layer transmission for light field streaming, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 2606–2610.

[49] M. Le Pendu, A. Smolic, High resolution light field recovery with fourier disparity layer completion, demosaicing, and super-resolution, in: 2020 IEEE International Conference on Computational Photography (ICCP), IEEE, 2020, pp. 1–12.

[50] M. Rerabek, T. Ebrahimi, New light field image dataset, in: 8th International Conference on QoMEX, 2016.

[51] D. G. Dansereau, O. Pizarro, S. B. Williams, Decoding, calibration and rectification for lenselet-based plenoptic cameras, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 1027–1034.

[52] G. Bjontegaard, Calculation of average psnr differences between rd-curves, VCEG-M33.
[53] K. Maruyama, H. Kojima, K. Takahashi, T. Fujii, Implementation of tabletop light-field display, in: IDW, 2018.

[54] Y. Kobayashi, K. Takahashi, T. Fujii, From focal stacks to tensor display: A method for light field visualization without multi-view images, in: IEEE ICASSP, IEEE, 2017, pp. 2007–2011.

[55] K. Maruyama, Y. Inagaki, K. Takahashi, T. Fujii, H. Nagahara, A 3-d display pipeline from coded-aperture camera to tensor light-field display through cnn, in: IEEE ICIP, IEEE, 2019, pp. 1064–1068.