Shearlet Transform-Based Light Field Compression Under Low Bitrates

Waqas Ahmad, Suren Vagharshakyan, Mårten Sjöström, Atanas Gotchev, *Member, IEEE*, Robert Bregovic, *Member, IEEE*, and Roger Olsson

***Abstract*— Light field (LF) acquisition devices capture spatial and angular information of a scene. In contrast with traditional cameras, the additional angular information enables novel post-processing applications, such as 3D scene reconstruction, the ability to refocus at different depth planes, and synthetic aperture. In this paper, we present a novel compression scheme for LF data captured using multiple traditional cameras. The input LF views were divided into two groups: key views and decimated views. The key views were compressed using the multi-view extension of high-efficiency video coding (MV-HEVC) scheme, and decimated views were predicted using the shearlet-transform–based prediction (STBP) scheme. Additionally, the residual information of predicted views was also encoded and sent along with the coded stream of key views. The proposed scheme was evaluated over a benchmark multi-camera based LF datasets, demonstrating that incorporating the residual information into the compression scheme increased the overall peak signal to noise ratio (PSNR) by 2 dB. The proposed compression scheme performed significantly better at low bit rates compared to anchor schemes, which have a better level of compression efficiency in high bit-rate scenarios. The sensitivity of the human vision system towards compression artifacts, specifically at low bit rates, favors the proposed compression scheme over anchor schemes.**

***Index Terms*— Light field (LF) coding, multiple camera system (MCS) coding, multi-view extension of high-efficiency video coding (MV-HEVC), shearlet.**

I. Introduction

THE spatial and angular information of scenes has attracted significant attention in various 3D capturing [1], [2], processing [3] and rendering applications [4]–[7]. The idea of capturing angular information along with spatial information was initially proposed by G. Lippmann in 1908 [8]. With the advancement in computing technology, the light field (LF) was captured using multiple traditional cameras [9]. Each camera captures a single perspective of the scene; thus, the capturing system records a sparsely sampled LF. Moreover, advancements in optical technology and the pursuit of dense sampling led to capturing LF using a single plenoptic camera. Initially, the plenoptic camera was introduced for the consumer market [10], and later, commercial applications were also targeted [11]. In a plenoptic camera, a lenslet array is placed between the main lens and image sensor to multiplex the spatial and angular information of the scene. Recording the spatial information of the scene from different perspectives provides an opportunity to perform various post-processing applications; however, it also increases the amount of captured data. Standard image and video encoders can be used to compress LF data. However, such encoders do not take into account the correlation present in LF data; hence, they provide low compression efficiency. A recent call for proposal from JPEG Pleno [12], reflects the importance of novel compression solutions for LF data. Recently, various compression schemes have been proposed with the aim of efficiently compressing LF data. These proposals for LF compression can be divided into two major groups based on acquisition technology: the plenoptic camera and the multi-camera system. However, few compression schemes are applicable to both types of captured data [13], [14].

A. Related Work

In 2016, a handful of compression schemes [15]–[19] were presented as a response to grand challenges [20], [21] of plenoptic image compression. The majority of the presented schemes introduced novel tools in standard high efficiency video coding (HEVC) image encoder to compress plenoptic images. Li *et al.* [16], [22] proposed a bi-prediction mode capability within an HEVC image compression framework for the compression of plenoptic images. In addition to 33 intra prediction modes in HEVC, each block was allowed to take predictions from previously encoded blocks. A similar approach was proposed by Monteiro et al., who added two novel tools to the HEVC image compression scheme. Each block can take predictions from other blocks by using the self-similarity (SS) and local linear embedding (LLE) operators [18]. An SS-based prediction scheme was incorporated in HEVC by Conti *et al.* [19]. Following the idea of a pseudo-video sequence (PVS), initially proposed by Olsson *et al.* [23], an alternative approach was proposed by Liu et al., who converted a plenoptic image into sub-aperture images, which were treated as frames of a PVS [17]. The HEVC video

Manuscript received January 29, 2019; revised July 30, 2019; accepted January 15, 2020. Date of publication January 29, 2020; date of current version February 10, 2020. The work in this paper was funded from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 676401, European Training Network on Full Parallax Imaging. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisimachos P. Kondi. *(Corresponding author: Waqas Ahmad.)*

Waqas Ahmad, Mårten Sjöström, and Roger Olsson are with the Department of Information Systems and Technology, Mid Sweden University–Sundsvall, 851 70 Sundsvall, Sweden (e-mail: waqas.ahmad@miun.se).

Suren Vagharshakyan, Atanas Gotchev, and Robert Bregovic are with the Faculty of Information Technology and Communication Sciences, Tampere University, 33100 Tampere, Finland.

Digital Object Identifier 10.1109/TIP.2020.2969087

encoder was used to encode the PVS, and the scheme was selected as the best proposal in the international conference on multimedia and expo (ICME) grand challenge. The representation of an input plenoptic image suitable for the HEVC video encoder has shown a high compression efficiency compared to introducing additional tools in the HEVC image compression standard.

In the grand challenge organized by international conference on image processing (ICIP) 2017 [12], plenoptic images were provided in the form of sub-aperture images, and all submitted compression schemes used the sub-aperture representation of plenoptic images for compression. Ahmad et al. proposed interpreting sub-aperture images as a frame of multi-view sequences and performed compression using multi-view extension of high-efficiency video coding (MV-HEVC) [13]. A two-dimensional prediction and rate allocation schemes were proposed to improve the compression efficiency. Tabus *et al.* [24] exploited the disparity information of input plenoptic images to increase compression efficiency. The disparity map was quantized into several regions, and the displacement of each region of the side view relative to the central view was estimated. A set of sparse views, a disparity map corresponding to the central view and region displacements of the side views were encoded. A pixel-level correlator was developed to further refine the side views from corresponding neighbor views. The compression scheme proposed by Zhao and Chen [25] categorizes the sub-aperture images into two groups: selected views and dropped views. The selected views were treated as PVS and coded using a video encoder. The dropped views were approximated as a weighted sum of the decoded version of selected views. Jia et al. proposed a specific ordering of sub-aperture images and compressed them as a PVS [26]. The decoded version of the PVS was again converted into a plenoptic image and the residual information was estimated and transmitted in order to enhance the visual quality.

In addtion to compression proposals for plenoptic images in the grand challenges, Li et al. proposed a memory-optimized 2D hierarchical coding structure for plenoptic image compression [27]. The sub-aperture images were divided into four quadrants, and predictions among images were contained within each quadrant in order to optimize the reference picture buffer. Li *et al.* [28] proposed a scalable coding scheme for the compression of plenoptic images captured with a plenoptic 2.0 camera. A sub-sampled set of microlens images and disparity information of missing microlens images were compressed and used to predict the input plenoptic image. Later on, the predicted plenoptic image was used to compress the original plenoptic image with an HEVC inter-prediction scheme. Bakir *et al.* [29] presented a plenoptic image compression scheme in which the input sub-aperture images were divided into two groups. First group was encoded using an HEVC encoder, and the second group was estimated using linear approximation of previously encoded sub-aperture images. On the decoder side, an additional deep-learning-based scheme was used to improve the reconstruction quality of sub-aperture images. Few researchers have developed compression solutions for multi-camera based LF data. Hawary et al. proposed a scalable compression scheme that mainly relies on the sparsity in the angular Fourier domain of the captured LF [30]. A sparse set of views were compressed as a PVS and were used to predict the remaining views. Jiang *et al.* [14] proposed a compression solution based on the homography information between the side views and the central view. A joint optimization problem was set up in which the homographies that minimized the low-rank approximation error were estimated. Ahmad et al. proposed interpreting LF data captured with the multi-camera system as frames of multiple PVSs and compressed using MV-HEVC [31]. In this way, the 2D correlation present among the views of LF data was exploited by using temporal and inter-view prediction tools available in MV-HEVC. Komatsu *et al.* [32] introduced a computationally efficient scalable coding scheme for multi-camera based LF data. A set of binary images were chosen to record the common structure among all views, and the differences among the views were represented with additional weight images. The number of binary images was provided as a free parameter in a scalable coding framework that controls the trade-off between quality and computational complexity. Alves *et al.* [33] analyzed the redundancy in plenoptic images and multi-camera–based LF data using 4D DCT transform.

B. Motivations and Contributions

The grand challenges for plenoptic image compression [20], [21] and the availability of plenoptic image datasets [34]–[36] have resulted in numerous plenoptic image compression schemes. However, LF data captured with multi-camera systems have received less attention from the research community. The wide range of post processing applications of LF data makes it challenging to benchmark LF compression schemes. At present, both objective and subjective assessment methods are used to evaluate LF compression schemes. The ICME 2016 grand challenge results [37] shows that schemes [17] with good compression efficiency at low bit rates also perform better in subjective quality assessment. Morever, it can be observed that most of the previously mentioned compression schemes [24], [25], [30], [38] use a subset of views to generate the remaining views. In this way, the reconstruction algorithms were integrated into the compression framework. In the perspective of JPEG Pleno's call for LF compression proposals, this study presents the shearlet-transform–based prediction (STBP) tool suitable for multi-camera LF data compression, especially at low bit rates. The proposed scheme uses an epipolar plane image (EPI) representation for a subset of input LF views and predicts the remaining views by applying shearlet transform in the frequency domain. In [39] it was reported that a shearlet-transform–based reconstruction scheme for LF data performs significantly better than other state-of-the-art LF reconstruction schemes.The pre-study in [38] provides the initial motivation to develop a shearlet-transform– based LF compression scheme. This paper extends the pre-study in [38] with the following key contributions:

- The theory, methodology and analysis of the abstract presented in [38] is explained in detail enabling the full reproduction of our proposed scheme. The paper presents

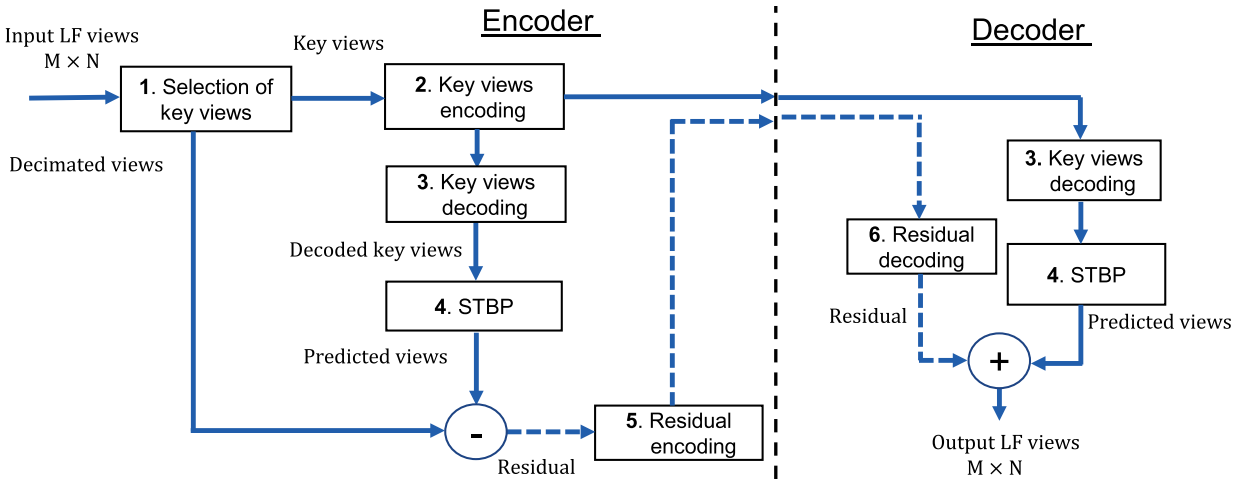

Fig. 1. Block diagram of the proposed compression scheme. On the encoder side, the proposed scheme categorizes the input LF views into key views and decimated views. Key views are compressed, and their corresponding decoded views are used to predict the decimated views by applying the STBP scheme. Residual information is calculated between decimated views and predicted views, and they are compressed separately along with key views. On the decoder side, a similar procedure is applied to predict the decimated views by using key views. The residual bitstream is decoded and added with predicted views to improve the visual quality.

an experimental setup for testing the reconstruction methods within the proposed compression scheme.

- We encoded key views by using an MV-HEVC–based compression scheme that allows each frame to take two-dimensional (2D) predictions from available neighboring frames. In section IV-A, it is reported that even for 5×5 key views the MV-HEVC based compression scheme provides 0.4 dB better performance than HEVC based pseudo video sequence compression scheme which is used in [25], [29], [38]. An analysis of variable quality allocation among key views on STBP scheme is presented in section IV-B.
- The STBP error was encoded and incorporated in the proposed compression scheme in order to improve compression efficiency at high bit rates. The state-of-the-art methods were investigated to compress the STBP error and the impact of adding residual information at different bit rates is explained in section IV-D. It is demonstrated that the residual coding of STBP error improves the PSNR by 2 dB.
- The proposed scheme was tested on nine LF images selected from three different publically available datasets with original spatial resolution instead of down-sampled LF images as in [38].

C. Scope and Limitation

The proposed scheme can be used for plenoptic images, but it provides limited compression efficiency compared to LF captured with a multiple camera system (MCS). The main reason for this limitation is the significantly lower range of disparities contained in plenoptic images,e.g., the Lytro Generation 1 and Illum cameras, thus reduces the need for the elaborate view interpolation method presented in this work. Furthermore, the sub-aperture image (SAI) representation of plenoptic images mimics the views captured by the MCS. However, the generation of SAIs involves several pre-processing steps, such as image demosaicing, that do not take into account the characteristics of the plenoptic image as analyzed in [40]. The SAIs also suffer from vignetting noise, misalignment, and a lack of proper spatial sampling. Hence, in the proposed study, we have considered datasets captured with a multiple camera system [1], [41], [42].

D. Paper Outline

In Section II, the main elements of the proposed compression scheme are discussed. The sub-sections explain the selection and encoding of key views, the STBP and the residual coding scheme. Section III explains the test conditions and the experimental setup. In Section IV, the experiment results are reported and discussed, including the encoding of key views, the effects of compression process on shearlet transform, the prediction of decimated views, and the residual coding and complexity analysis of proposed compression scheme. Finally, Section V presents the major conclusions of this work.

II. Proposed Compression Scheme

The block diagram of the proposed compression scheme is presented in Fig. 1. The LF with $M \times N$ views is given as an input to the compression scheme. The input views are divided into two categories, i.e. key views and decimated views. The STBP is applied to the decoded version of key views in order to predict the decimated views. Moreover, the quality of predicted views is enhanced by incorporating the residual information of the decimated views. The details of each block of Fig. 1 is as follows:

1. Selection of key views: A set of sparse views (hereafter, key views) is selected from input LF by following the procedure explained in Section II-A. The remaining views are marked as decimated views, and they are used to compute the residual information.

2. Key views encoding: The MV-HEVC based compression scheme is used to compress the key views [43]. The compression scheme takes the multi-view PVSs and uses tools

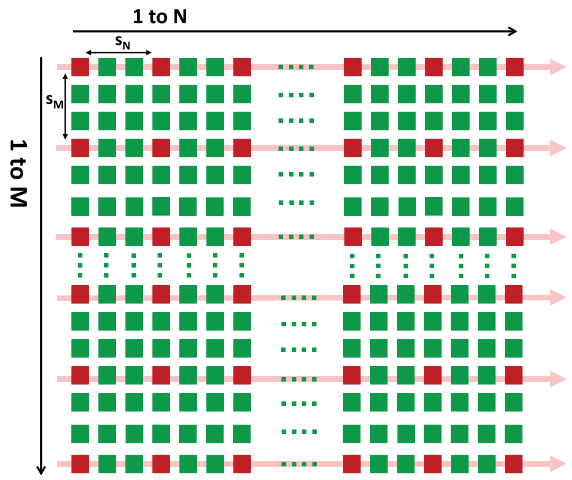

Fig. 2. The input LF with $M \times N$ views was uniformly sub-sampled with factor S in both directions to obtain $M_s \times N_s$ key views.

available in MV-HEVC to exploit the 2D correlation present in LF data.

3. Decoding of key views: The MV-HEVC based encoder as show in block 2 of Fig. 1 maintains the decoded frames in order to perform inter- and bi-predictive coding. This block uses the existing, built-in decoding of the MV-HEVC encoder.

4. Shearlet-transform–based prediction (STBP) scheme: The decimated views are recreated by predicting them from the decoded key views using the STBP. Section II-B contains a detailed description of the STBP.

5. Residual encoding: The residual information is computed by taking the difference between decimated views and predicted views. In the proposed method, residual information is converted into a single PVS and compressed along with key views. The residual compression scheme is explained in Section II-C.

6. Residual decoding: The bitstream corresponding to the residual information of decimated views is decoded using the base layer of MV-HEVC.

A. Key Views–Selection and Encoding

The captured LF with $M \times N$ views is uniformly decimated by factor S in both horizontal and vertical directions, resulting in a sparse set of $M_s \times N_s$ views also referred to as key views. In the next stage, the encoding of key views is performed and in our proposed method, the key views are interpreted as a set of M_s pseudo videos with each having N_s frames as shown in Fig. 2. The key frames are compressed using state-of-the-art MV-HEVC by following the method proposed in [43]. Exploiting the tools available in MV-HEVC, a two-dimensional prediction scheme, shown in Fig. 3, is used to classify the views as frames. The compression scheme makes use of four parameters of MV-HEVC in order to assign a specific prediction level and rate-allocation to each frame. The parameters picture order count (POC) and view ID (VID) uniquely identify the position of each frame in the MV-HEVC framework. Similarly, the decoding order (DO) and view order index (VOI) represent the decoding order of each frame on the horizontal (POC) and vertical (VID) axes. The POC and VID axes are assigned with different predictor levels.

TABLE I
LIST OF IMPORTANT NOTATIONS

Symbol	Description
M_s	Number of PVSs
N_s	Number of frames in each PVS
b_{POC}	POC assigned to base frame
b_{VID}	VID assigned to base frame
Q_{b}	Quantization parameter assigned to base frame
n_{POC}	Picture order count of each frame
v_{VID}	VID of each frame
k_{DO}	Decoding order of each frame
i_{VOI}	View order index of each frame
s_{POC}	Assigned prediction level in POC axis
t_{VID}	Assigned prediction level in VID axis

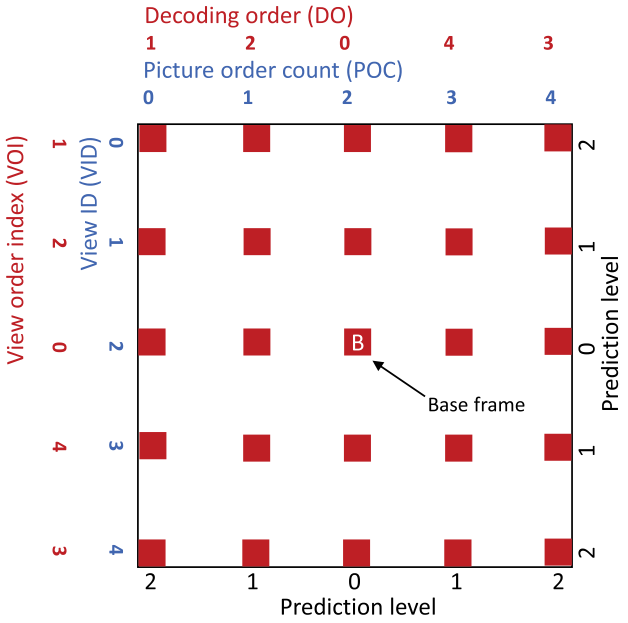

Fig. 3. The rate allocation scheme for 5×5 key views. The central view with POC $= 2$ and VID $= 2$ is chosen as base frame.

Fig. 3 presents an example of 5×5 key views, and the central frame with POC $= 2$ and VID $= 2$ is taken as a base frame and assigned with a prediction level of 0. The remaining frames are assigned with either a prediction level of 1 or 2. In the rate-allocation process, the frames with low prediction level are assigned with high quality and the quality is decreased at each successive prediction level. In this way, better quality frames are used for the prediction of other frames in order to improve the overall compression efficiency.

Algorithm 1, explains the rate-allocation scheme used to encode the key frames and Table I displays the parameters used in algorithm 1. The quality of each frame is controlled by varying the quantization parameter. The presented scheme estimates the position of the base frame and assign it with a base quantization parameter. An offset is added to the base quantizaton parameter, and assigned as a quantization parameter for each remaining frame. The offset is estimated by considering the frame distance (distance between current frame and base frame) and the decoding distance of each frame relative to the base frame. The rate allocation scheme iterates over all the frames and estimates the required quantization offset (Q_o) for each frame. The frames having POC $=$ 2 or

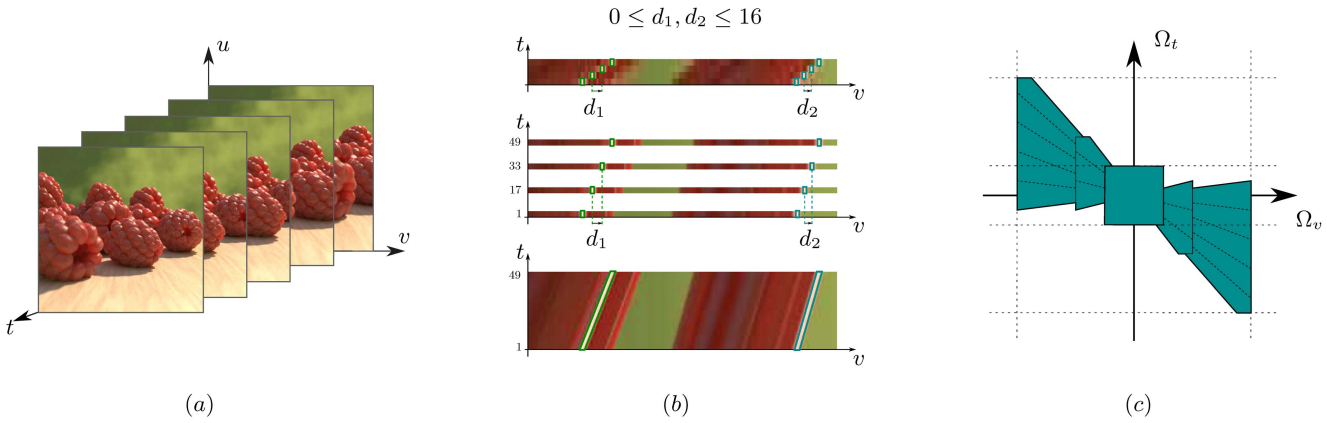

Fig. 4. (a) The parameterization of captured images, (b) the interpretation of input images as decimation of DSLF, (c) and an example of frequency plane tilling using shearlet transform required for efficient DSLF reconstruction.

TABLE II
WEIGHTS ASSIGNED TO EACH FRAME BASED ON ITS PREDICTION LEVEL

Predictor Levels (PL)	**Picture Order Count**		
View ID	$PL = 0$	$PL = 1$	$PL = 2$
$PL = 0$	Q_b	4	4
$PL = 1$	4	3	2
$PL = 2$	4	2	1

VID = 2 (in the base column or base row) are assigned a quantization offset equivalent to their prediction level (0, 1, or 2). The quantization offset for the remaining frames is calculated by using the frame distance and decoding distance with respect to the base frame (calculated in line 22). Lines 10 and 11 of algorithm 1 calculate the frame distance between the current frame and the base frame on the POC and VID axes. Similarly, from line 12 to line 20 decoding distance between the current and the base frame on both the POC and VID axes is calculated. A weight parameter W is used to control the limit of the quantization parameter for each frame. In line 11, the weighing function indexes the weights matrix for each frame by using frame prediction level on the POC (s_{POC}) and view ID axes (t_{VID}). Table II displays the parameter W assigned to each frame based on the frame's prediction level. The frames with low prediction levels are assigned with high weights compared to the frames assigned with high prediction levels. Finally, the quantization parameter ($Q_{(x,y)}$) for each frame is calculated in line 24 and returned as output by the algorithm 1.

B. Shearlet-Transform–Based Prediction (STBP) Scheme

In order to significantly reduce the LF data, a good prediction scheme is required. In our procedure, we employed an LF reconstruction algorithm utilizing the shearlet transform developed for the reconstruction of densely sampled LF (DSLF) [39], which means that disparity between adjacent views is no more than one pixel apart. This property allows one to obtain an arbitrary ray inside the viewing zone by simple local interpolation, such as linear interpolation, without involving computationally demanding global processing. The capability of STBP to reconstruct the intermediate views from a sparse set of views is exploited for LF compression.

Algorithm 1 Rate Allocation for Key Views

Input: M_s, N_s, b_{POC}, b_{VID}, Q_b, n_{POC}, v_{VID}, k_{DO}, i_{VOI}, s_{POC}, t_{VID}

1: Read POC (n_{POC}) and VID (v_{VID}) of each frame.
2: Read DO (k_{DO}) and VOI (i_{VOI}) of each frame.
3: Read prediction level in POC (s_{POC}) and VID axis (t_{VID}).
4: **for** x = 1:M_s **do**
5: **for** y = 1:N_s **do**
▷ Getting the assigned weight value of current frame
6: $W = Weightage(s_{\text{POC}}(x), t_{\text{VID}}(y))$
7: **if** $x == b_{\text{POC}}$ && $y == b_{\text{VID}}$ **then**
▷ Current frame lies in base ViewID or base POC
8: $Q_{\text{o}(x,y)} = max(s_{\text{POC}}(x), t_{\text{VID}}(y))$
9: **else**
10: $d_{\text{POC}} = \lfloor \frac{|n_{\text{POC}}(x) - b_{\text{POC}}|}{W} \rfloor$
11: $d_{\text{VID}} = \lfloor \frac{|v_{\text{VID}}(y) - b_{\text{VID}}|}{W} \rfloor$
12: **if** $n_{\text{POC}} \le b_{\text{POC}}$ **then**
13: $d_{\text{DO}} = \lfloor \frac{k_{\text{DO}}(x)}{W} \rfloor$
14: **else**
15: $d_{\text{DO}} = \lfloor \frac{k_{\text{DO}}(x) - b_{\text{POC}}}{W} \rfloor$
16: **end**
17: **if** $v_{\text{VID}} \le b_{\text{VID}}$ **then**
18: $d_{\text{VOI}} = \lfloor \frac{i_{\text{VOI}}(y)}{W} \rfloor$
19: **else**
20: $d_{\text{VOI}} = \lfloor \frac{i_{\text{VOI}}(y) - b_{\text{VID}}}{W} \rfloor$
21: **end**
▷ Quantization offset for current frame
22: $Q_{\text{o}(x,y)} = d_{\text{POC}} + d_{\text{VID}} + d_{\text{DO}} + d_{\text{VOI}}$
23: **end**
▷ Quantization parameter for current frame
24: $Q_{(x,y)} = Q_{\text{b}} + Q_{\text{o}(x,y)}$
25: **end**
26: **end**
27: *Output:* Q

The full parallax 4D LF is described using two plane parameterization [9],

$$L(u, v, s, t), \tag{1}$$

where (u, v) plane represents the image plane coordinates for each view, and (s, t) are coordinates of the capturing plane as shown in Fig. 4(a). By fixing the (u, s) or (v, t) parameters horizontal and vertical EPIs [44] are formed as follows

$$E^H(v, t) = L(u_0, v, s_0, t) \tag{2}$$

$$E^V(u, s) = L(u, v_0, s, t_0) \tag{3}$$

Fixing the parameter $s = s_0$ and $u = u_0$ in (1) results in generating a horizontal EPI image, as presented in (2). Similarly, fixing $t = t_0$ and $v = v_0$ in (1) generates a vertical EPI image, as presented in (3). In general, it is assumed to have sufficient sampling over the image plane (u, v), such that cameras provide enough resolution to capture the finest details of the scene. In the proposed approach, as an LF reconstruction tool, we undertook EPI reconstruction using shearlet transform presented in [39]. Intermediate view reconstructions in the 4D full parallax case can be interpreted as multiple 3D horizontal and vertical parallax DSLF reconstructions. Each 3D parallax DSLF can be obtained by reconstructing each densely sampled EPI from coarsely sampled EPIs, as illustrated in Fig. 4(b) for the horizontal parallax case.

Due to strict structure of DSLF in the spatial and frequency domains, the reconstruction of DSLF from available measurements can be considered as a sparse regularization problem for inpainting. The regularization tool in our case is the shearlet transform, since it is a directional sensitive transform based on the shear operator, allowing one to construct desirable frequency domain tilling, as shown in Fig. 4(c). Let us assume that measurements g is obtained using M measurement matrix applied on densely sampled ground truth EPI f such that

$$g = M \odot f, \tag{4}$$

where $\odot$ represents element-wise multiplication. As demonstrated in [39] a good approximation f_n of f can be obtained using the iterative procedure

$$f_{n+1} = S^*(T_{\lambda_n}(S(f_n + \alpha_n(g - M \odot f)))), \tag{5}$$

where T_λ is a hard thresholding operator, S and S^* are direct and inverse shearlet transforms respectively. Algorithm 2 explains the shearlet transform based reconstruction process for a single EPI image. The inputs of the algorithm include number of iterations, EPI image, mask (indicates key views pixels), shearlet analysis and synthesis filters. In the analysis part, the Fourier transform of the EPI image is multiplied with each shearlet analysis filter, and corresponding coefficients are computed by taking the inverse Fourier transform of the product. The best coefficients are selected by applying a hard threshold. In the synthesis part, the Fourier transforms of each selected coefficient are computed and multiplied with the corresponding shearlet synthesis filter. The summation is computed for all the responses, and the inverse Fourier transform is applied to estimate the reconstructed EPI image. The difference (scaled by parameter α) between the reconstructed and the original EPI image is computed and added with the reconstructed EPI image. The reconstructed image is then used for the next iteration, and algorithm steps are repeated for N iterations. More details about the construction of transforms, parameters and iterative procedure can be found in [39]. It is important to note that the construction and computation of S and S^* transforms are mainly based on the $d_{range} = d_{max} - d_{min}$ - range of disparity values in available measurements g; thus, estimation of d_{min} and d_{max} are assumed as prior knowledge.

Algorithm 2 Shearlet-Transform–Based Prediction Scheme

Input: g, Given EPI image
N, number of iterations
S, Shearlet analysis filters in frequency domain
S^*, Shearlet synthesis filters in frequency domain
q, Number of shearlet filters
M, Mask of EPI image
α, Acceleration parameter for rate of convergence
λ, Set of threshold values for each iteration

1: $f_1 = g$ ▷ Initially set f_1 as the original EPI image
2: **for** n = 1:N **do**
3: $F_n = \mathscr{F}\{f_n\}$ ▷ Fourier transform of f_n
4: **for** i = 1:q **do** ▷ Perform shearlet analysis
5: $C(i) = \mathscr{F}^{-1}(F_n \times S_i)$
▷ Apply threshold to select best coefficients
6: $C^*(i) = \begin{cases} C(i), & \text{if } |C(i)| \geq \lambda_n \\ 0, & \text{if } |C(i)| < \lambda_n \end{cases}$
7: **end**
8: $F_0 = 0$
9: **for** j = 1:q **do** ▷ Perform shearlet synthesis
10: $F_j = F_{j-1} + \mathscr{F}\{C^*(j)\} \times S_j^*$
11: **end**
12: $f_n = \mathscr{F}^{-1}(F_j)$ ▷ Reconstructed EPI image
13: $f_{n+1} = f_n + \alpha_n(g - Mf_n)$
14: **end**
15: *Output:* f_{n+1}

C. Residual Encoding

The STBP scheme provides significant compression efficiency at low bit rates [38]. However, at high bit rates, proposed prediction scheme has an inherited reconstruction error, and it requires additional residual information to improve the visual quality. In the proposed compression scheme, the residual information is also encoded and sent along with the bitstream of key views. The residual is computed by taking the difference between decimated views and predicted views. Algorithm 3 explains the process of generating residual PVS. The algorithm iterates over each view of input LF (I) and reconstructed LF (R), and it estimates the error signal (E). In line 6, the minimum error value is added with the error signal in order to have non-negative values in residual PVS (P_{residual}). Hence, for each view, the minimum error values are also transmitted with the bitstream. In the proposed compression scheme, the residual information of each view is interpreted as a frame of PVS and compressed in the base layer of MV-HEVC using the intra-prediction mode.

Algorithm 3 Residual Sequence Generation

Input: I, Original LF views
R, Predicted LF views

$f = 0$
for m = 1:M **do**
  for n = 1:N **do**
    $f = f + 1$
    $E(m,n) = I(m,n) - R(m,n)$
    ▷ Making values of residual sequence non-negative
    $P_{\text{residual}}(1,f) = E(m,n) - \min(\min(E(m,n)))$
  end
end
Output: P_{residual}

TABLE III
LF IMAGES USED FOR EXPERIMENTATION

Name	M	N	S_M	S_N	W	H
Chess	17	17	4	4	1400	800
Lego Buildozer	17	17	4	4	1536	1152
Eucalyptus Flowers	17	17	4	4	1280	1536
Amethyst	17	17	4	4	768	1024
Bunny	17	17	4	4	1024	1024
Jelly Beans	17	17	4	4	1024	512
Greek	9	9	2	2	512	512
Sideboard	9	9	2	2	512	512
Set2	11	33	2	4	1920	1080

III. TEST ARRANGEMENT AND EVALUATION CRITERIA

The experimentation was performed on nine LF images selected from three different publicly available datasets [1], [41], [42]. Table III displays the details of selected LF images in which M and N represents the angular resolution and W and H represents the spatial resolution of each LF image. The input views were sub-sampled using horizontal and vertical sampling period S_M and S_N in order to form a sparse set of key views. Each LF image contains views in RGB format, and its equivalent YUV444 format was used as a reference input signal. The reference input signal was further converted into YUV420 format and given as an input to the proposed and anchor compression schemes. The shearlet transform has filtering artifacts on the image borders, as described in [39]. Instead of extra padding, the comparison was made with the anchor schemes by excluding 21 pixels from each side of the image. The mean PSNR ($PSNR_{mean}$) in Y component of all the views was used as a quality metric to evaluate the compression efficiency of the proposed scheme as explained in (6).

$$\text{PSNR}_{mean} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\text{PSNR}(m,n) \tag{6}$$

The PSNR of a specific view (at view position m and n) is estimated by:

$$\text{PSNR(m, n)} = 10\log_{10}\frac{255^2}{MSE(m,n)} \tag{7}$$

where the mean square error between the views is estimated by:

$$MSE(m,n) = \frac{1}{T}\sum_{x=b+1}^{W-b}\sum_{y=b+1}^{H-b}[I(x,y) - I'(x,y)]^2 \tag{8}$$

where b represents the border pixels excluded from each side of the image; W and H indicate the width and height of each view, respectively; T denotes the number of pixels of each view considered for comparison ($T = (W - 2 * b)(H - 2 * b)$); and $I(x,y)$ and $I'(x,y)$ represent the value of pixels in the original and reconstructed views, respectively. The BD-PSNR [45] metric was also used to compare the compression results.

The compression efficiency of proposed scheme was evaluated against the state-of-the-art compression scheme [31] and two benchmark HEVC [46] and X265 [47] anchor schemes. The video coding schemes use highly optimized tools to exploit temporal correlation,(i.e., motion vector scaling, advance motion vector prediction, differential motion vector encoding, etc). The interpretation of LF views as a PVS enables video coding schemes to exploit the inter-view correlation present in LF data. The LF views were converted into a single PVS and given as an input to the benchmark anchor schemes [46], [47]. The first frame was encoded as an intra-frame, the second as a P-frame, and all the remaining frames were encoded as B-frames.

IV. RESULTS AND ANALYSIS

A. Encoding of Key Views

The initial step of the proposed compression scheme involved compressing the key views by using MV-HEVC as explained in Section II-A. The key views were also converted into a single PVS and compressed using an HEVC encoding scheme. Two LF images from the Stanford dataset, namely Chess and Eucalyptus flowers were compressed at four different bit rates in order to test varying bit-rate scenarios. The rate distortion (RD) comparison between the proposed scheme and the HEVC scheme is presented in Fig. 5. It can be seen that the proposed scheme provides better compression efficiency compared to the benchmark HEVC scheme with an average BD-PSNR gain of 0.4 DB. The proposed scheme enables each frame to exploit two-dimensional inter-view correlation from neighboring views. Moreover, allocating better quality to frames used for the prediction of other frames improves compression efficiency. The compression efficiency of the proposed scheme improves with the increased number of key views since more frames take predictions from better quality frames.

B. Compression Artifacts on Shearlet Transform

The shearlet transform was applied to EPI images that exhibit a special line structure. The line in the EPI image corresponds to points/regions visible in the perspective views captured by each camera. The variation in quality among images as a consequence of the compression process can affect the reconstruction process when employing EPI images.

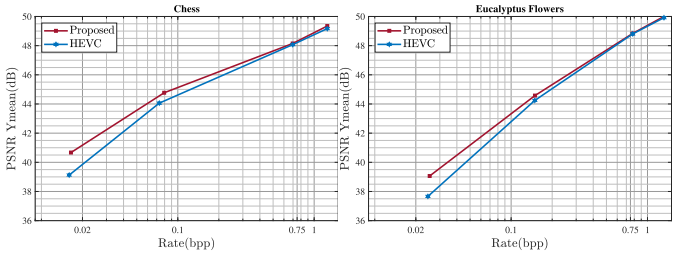

Fig. 5. Rate-distortion analysis between the proposed compression scheme and the HEVC scheme for 5×5 key views.

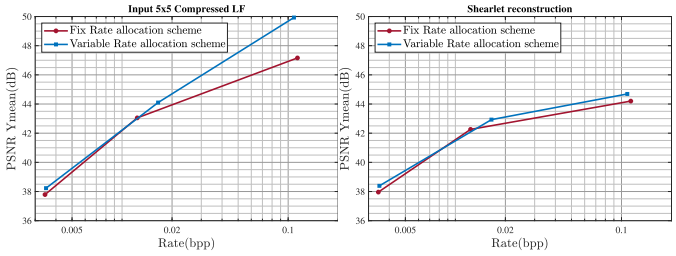

Fig. 6. Rate-distortion analysis of variable and fixed quality rate-allocation schemes. a) The input 5×5 key views were compressed using fixed and variable rate-allocation schemes. b) The corresponding shearlet reconstruction of 17×17 views for fixed and variable rate allocation schemes.

An experiment was performed in order to study the effect of variable rate-allocation on the STBP process. The Truck image from Stanford dataset was used and a subset of 5×5 key views was extracted from 17×17 input views by following the procedure explained in Section II-A. The encoding of key views was performed using HEVC with two different rate-allocation schemes. In the first encoding scheme, a fixed quantization parameter was used to ensure the similar quality among images. In the second encoding scheme, variable quantization parameters were used for all 25 views to yield variable quality. The compression was performed on three different bit rates for both fixed and variable rate-allocation schemes. Fig. 6 (a) and (b) display the RD curves for 5×5 key views and the reconstruction results of the STBP, respectively. The combined compression efficiency of the variable rate-allocation scheme on 5×5 key views and the shearlet reconstruction outperform the fixed rate allocation scheme. Hence, the key views compressed with variable rate-allocation scheme can be used as an input of STBP.

C. Prediction of Decimated Views Using Shearlet Transform

Fig.7 displays an RD comparison between the STBP (without incorporating residual information) and anchor schemes. The performance of anchor schemes is better at high bit rates compared to the proposed scheme. However, at the low bit rates, the compression efficiency of the proposed scheme is higher compared to anchor schemes. The difference in the behavior of compression schemes is a consequence of their utilization of input information. In the high bit-rate scenario, the high quality of residual information enables the anchor schemes to achieve efficient compression. Conversely, shearlet transform relies on key views to predict the intermediate views without incorporating residual information;

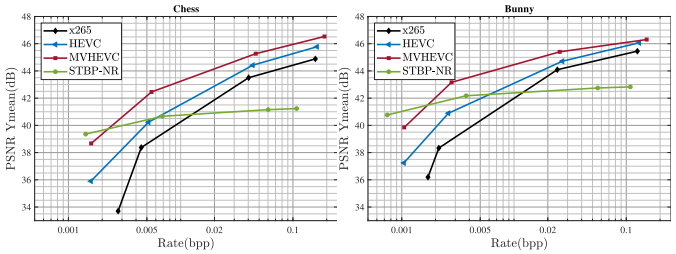

Fig. 7. Rate-distortion analysis between STBP and anchor schemes.

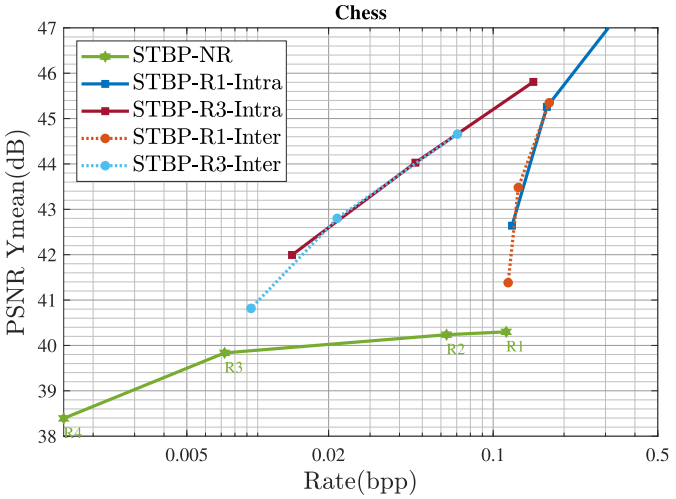

Fig. 8. Rate-distortion analysis of HEVC intra- and inter-prediction coding.

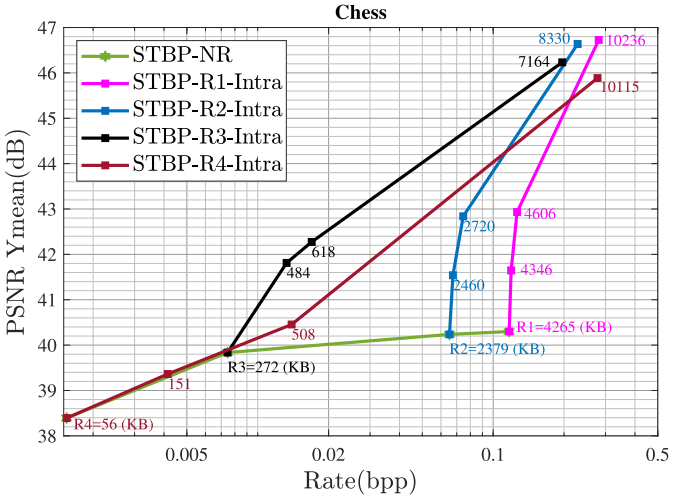

Fig. 9. Rate-distortion analysis of the STBP scheme with residual information added to the predicted views at rates R1, R2, R3, and R4.

hence, it has an inherit reconstruction error. In the low bit-rate scenario, the bit budget of the proposed compression scheme allows the encoder to provide higher quality to key views. In this way, the shearlet transform utilizes good quality key views to predict the intermediate views. On the other hand, anchor schemes distribute the bit budget among all the views that result in degradation of overall visual quality.

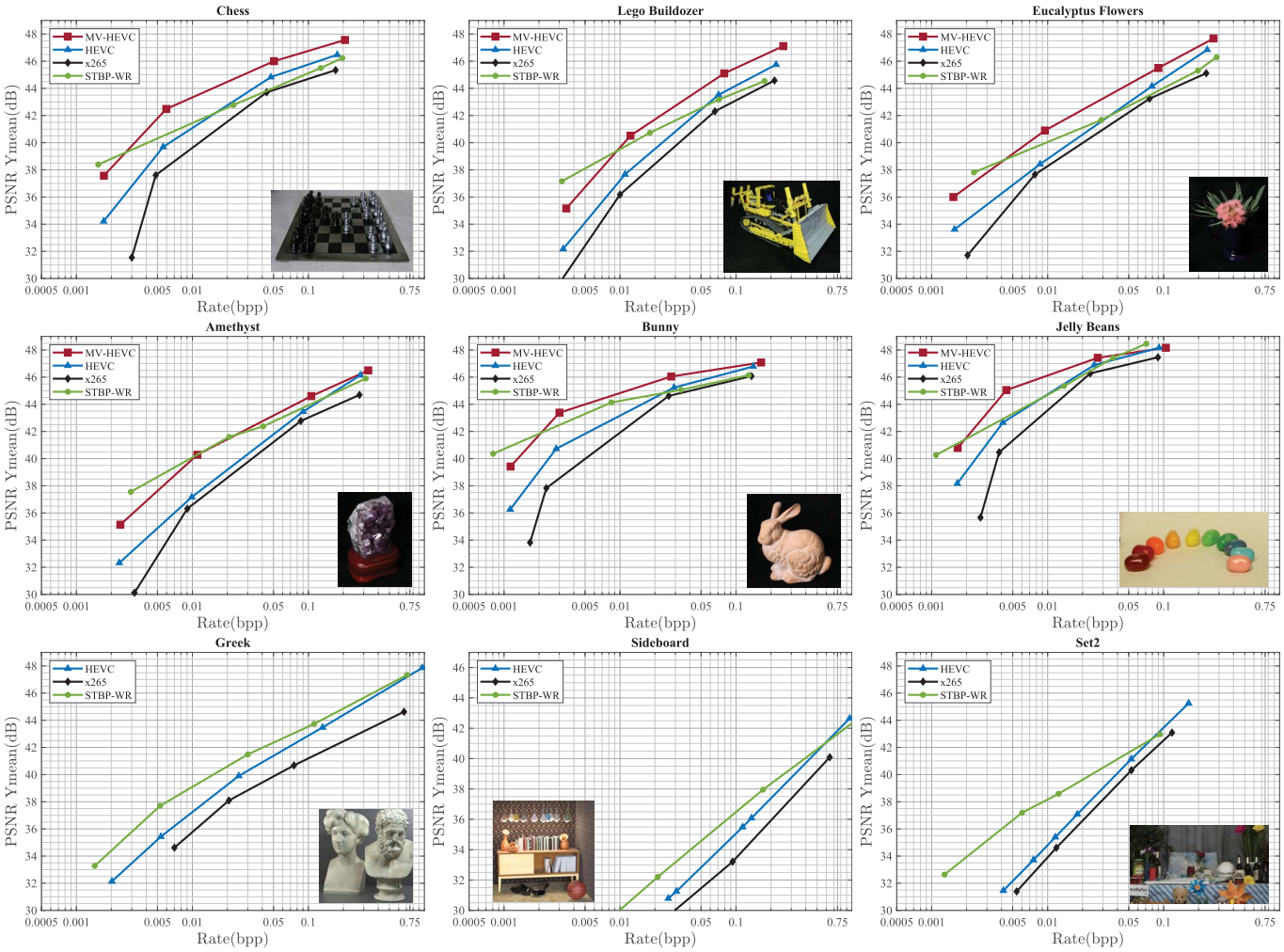

Fig. 10. Rate-distortion analysis of the proposed compression scheme (STBP-WR) with two benchmark anchor schemes (HEVC and X265) and a state-of-art compression scheme (MV-HEVC). The central view of each LF image is also shown in each sub figure.

D. Residual Encoding

The reconstruction error of the STBP scheme can be reduced by encoding the residual information of the predicted views. In the proposed compression scheme, the residual information is treated as a single PVS and given as an input to the base layer of MV-HEVC, which works similarly to HEVC for such an input. In order to exploit the correlation present in the residual information using tools available in HEVC, the residual sequence was encoded with intra- and inter-prediction modes, and the RD comparison is shown in Fig. 8. The STBP-NR curve represents the reference STBP scheme (without residual coding). The key views, selected as input for STBP were encoded at four different bit rates, and that also describes the rate of STBP-NR (since no residual information was added). Hereafter, we call these four rates R1, R2, R3, and R4. The residual information corresponding to rates R1 and R3 was encoded using HEVC intra-prediction (STBP-R1-Intra and STBP-R3-Intra) and the HEVC inter-prediction mode (STBP-R1-Inter and STBP-R3-Inter). In the inter-prediction mode, the encoder was allowed to use all the available prediction tools in the HEVC framework (intra-, inter- and bi-prediction modes). However, the encoder was forced to use only intra-prediction modes in HEVC intra coding. The RD curves in Fig. 8 reflect a similar compression efficiency between inter- and intra-prediction schemes. Inter-prediction makes use of motion estimation and compensation for predicting the current frame which is not beneficial for residual PVS since it does not possess the properties of natural images. The similar compression efficiency between the two schemes reflects a lower correlation among frames of residual PVS. Hence, we proposed using the HEVC intra-prediction mode to encode the residual PVS. The HEVC intra-prediction scheme has a relatively less computational cost compared to HEVC inter-prediction coding and it can be further used to obtain random access capability in the proposed compression scheme.

Fig. 9 displays the enhancement in visual quality obtained due to the addition of residual information. The STBP-NR represents the reference STBP scheme, evaluated at four bit rates (R1 = 4265 KB, R2 = 2379 KB, R3 = 272 KB, and R4 = 56 KB). The scheme STBP-R1-Intra adds the residual information with predicted views at rate R1 (4265 KB). The residual information was coded with different quantization values, and its decoded version was added with predicted

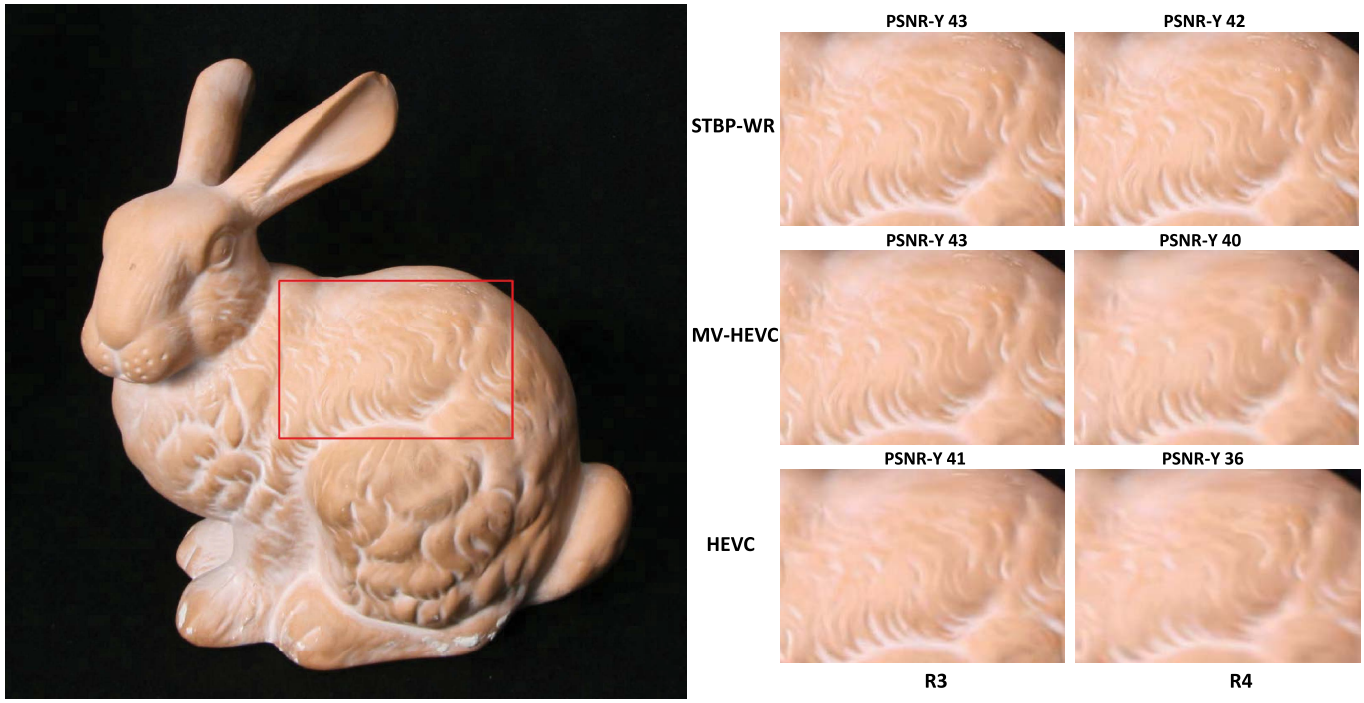

Fig. 11. Subjective analysis of the proposed compression scheme for a view from the bunny image. a) The original view with highlighted sub region. b) Shows the rendered sub-region compressed with the proposed scheme, as well as the MV-HEVC, and HEVC compression schemes. Two low bit rates (R3 and R4) were considered, and the PSNR of each view was also mentioned for each compression scheme.

views. Similarly, for the other three rates (R2, R3, and R4), the residual information corresponding to each rate was encoded with different quantization parameters, and its decoded version was added with the corresponding predicted views. The response of residual coding was not found same for all four bit-rates. Fig. 9 illustrates the more significant compression efficiency at higher rates compared to that at lower rates. The performance of STBP at low bit rates is much better (as shown in 7), and adding a residual with Intra coding method does not improve visual quality relative to the added size of the coded residual. The prediction efficiency of the STBP starts decreasing from R3 and it becomes very low at high bit rates (R2 and R1). In other words, the input quality of key views at high bit rates has less influence on STBP. For example, the coded size of key views at R2 is 2379 KB, and increasing the quality of key views by allocating an extra 1886 KB (at R1 = 4265 KB) has less impact on visual quality (around a 0.1 dB increase). In comparison, adding 248 KB of residual information at R2 has a significant impact on overall visual quality (around a 2 dB increase). It can be concluded that adding residual information at high bit rates improves the compression efficiency of the proposed scheme.

Fig. 10 illustrates the comparison of the proposed compression scheme with the two benchmark schemes and a state-of-the-art scheme [31]. It can be seen that, overall, the proposed scheme performs better compared to both anchor schemes. The compression efficiency of the proposed scheme is significantly better at low and medium bit rates relative to the high bit rate scenarios. In comparison to the state-of-the-art scheme, the proposed scheme has better compression at low bit rates, and the scheme presented in [31] performs better at medium and high bit rates. The sensitivity of the human vision system towards compression artifacts specifically at low bit rates [48], favors the proposed compression scheme over the other presented compression schemes.

E. Subjective Analysis

A sub-region of the Bunny image taken from view (6,9) is highlighted in Fig. 11 for compression rates R3 and R4. The HEVC-based compression schemes show notable artifacts in the decoded view in both selected rates. MV-HEVC-based compression scheme shows observable blurriness at rate R4. At both rates, it can be seen that the proposed scheme retains most of the information compared to other compression schemes. The compressed views from all the compresses schemes at rate R4 are available online [49].

F. Computational Complexity

The computational complexity of the proposed compression scheme is dependent on usage of residual information. In cases where residual information is not used in the proposed scheme the encoder only compresses key views using MV-HEVC. On the decoder side, the key views are decoded using the MV-HEVC decoder, and STBP is used to predict the decimated views. In such compression schemes, complexity is significantly reduced on the encoding side because only key views are compressed. On the decoding side, the MV-HEVC-based decoding of decimated views is replaced with the STBP process. The addition of residual information in the proposed compression scheme requires the use of the

STBP process on the encoder side, and the residual information is coded using the MV-HEVC single layer intra-prediction mode. On the decoding side, the key processes include decoding the key views by using MV-HEVC, STBP to predict the decimated views and to decode the residual bitstream. Hence, the enhancement in the visual quality of the predicted views is obtained at an increased computational cost.

The proposed compression scheme can be used in two ways: 1) The first way results in a complete LF compression scheme with high RD efficiency at low bit rates and acceptable RD efficiency at high bitrates. 2) In the second way, the encoder selects the proposed scheme for low bit rates and switches to a more optimal compression scheme for high bit rates. Modern encoders select optimal tools during the encoding process by using the RD optimization. The decision for encoders becomes more complex when they iterate over all available options in order to select the most optimum tool for a specific scenario. However, Fig. 10 demonstrates that STBP consistently performs better at low bit rates which can simplify the encoder's task.

V. Conclusion

In this paper, we have presented a novel compression solution for LF data captured with a multi-camera system. The input LF views were divided into two categories: key views and decimated views. The key views were encoded using MV-HEVC and decimated views were predicted using the STBP scheme. Additionally, the residual information was also coded in order to further enhance the visual quality of the predicted views. The proposed compression scheme performed better in low bit rates compared to anchor schemes which have better compression efficiency at high bit rates. The sensitivity of the human vision system towards compression artifacts at low bit rates favors the proposed compression scheme over the anchor schemes. The proposed compression scheme can be used without incorporating the residual information. In such cases, on the encoder side, key views are coded and sent to the decoding side where STBP scheme predicts the remaining decimated views. The proposed compression scheme can benefit applications where fewer resources are available on the encoding side. The proposed scheme can be further improved by introducing coding tools that exploit the correlation present in residual information. In the future, we will investigate alternative compression method for the residual information, and we also intend to perform a subjective quality assessment of the proposed scheme with other state-of-the-art LF reconstruction algorithms.

References

[1] M. Ziegler, R. Op Het Veld, J. Keinert, and F. Zilly, "Acquisition system for dense lightfield of large scenes," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jun. 2017, pp. 1–4.

[2] M. Martínez-Corral *et al.*, "Integral imaging with Fourier-plane recording," *Proc. SPIE*, vol. 10219, May 2017, Art. no. 102190B.

[3] A. Ansari, A. Dorado, G. Saavedra, and M. M. Corral, "Plenoptic image watermarking to preserve copyright," *Proc. SPIE*, vol. 10219, May 2017, Art. no. 102190A.

[4] S. Hong, A. Ansari, G. Saavedra, and M. Martínez-Corral, "Full-parallax 3D display from stereo-hybrid 3D camera system," *Opt. Lasers Eng.*, vol. 103, pp. 46–54, Apr. 2018.

[5] P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, and T. Balogh, "Towards adaptive light field video streaming," in *Proc. IEEE COMSOC MMTC Commun.-Frontiers*, Jul. 2017, pp. 1–4.

[6] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 1–20, Mar. 2015.

[7] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1027–1034.

[8] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. de Phys. Théorique et Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[9] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.

[10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, 2005.

[11] C. Perwaß and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," *Proc. SPIE*, vol. 8291, Feb. 2012, Art. no. 829108.

[12] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE Multimedia-Mag.*, vol. 23, no. 4, pp. 14–20, Oct. 2016.

[13] W. Ahmad, R. Olsson, and M. Sjostrom, "Interpreting plenoptic images as multi-view sequences for improved compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4557–4561.

[14] X. Jiang, M. Le Pendu, R. A. Farrugia, S. S. Hemami, and C. Guillemot, "Homography-based low rank approximation of light fields for compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 1313–1317.

[15] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–4.

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

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

[18] R. Monteiro *et al.*, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–4.

[19] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–4.

[20] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, "ICME 2016 grand challenge: Light-field image compression," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Seattle, WA, USA, Jul. 2016.

[21] *JPEG Pleno*, document ISO/IEC JTC 1/SC29/WG1N74014, Call for Proposals on Light Field Coding, Geneva, Switzerland, Jan. 2017.

[22] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1308–1319, Jul. 2016.

[23] R. Olsson, M. Sjöström, and Y. Xu, "A combined pre-processing and H.264-compression scheme for 3D integral images," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 513–516.

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

[25] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4562–4566.

[26] C. Jia *et al.*, "Optimized inter-view prediction based light field image compression with adaptive reconstruction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4572–4576.

[27] L. Li, Z. Li, B. Li, D. Liu, and H. Li, "Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 7, pp. 1107–1119, Oct. 2017.

[28] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 80–91, Jan. 2016.

[29] N. Bakir, W. Hamidouche, O. Deforges, K. Samrouth, and M. Khalil, "Light field image compression based on convolutional neural networks and linear approximation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1128–1132.
[30] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable light field compression scheme using sparse reconstruction and restoration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3250–3254.
[31] W. Ahmad, M. Sjöström, and R. Olsson, "Compression scheme for sparsely sampled light field data based on pseudo multi-view sequences," *Proc. SPIE*, vol. 10679, May 2018, Art. no. 106790M.
[32] K. Komatsu, K. Takahashi, and T. Fujii, "Scalable light field coding using weighted binary images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 903–907.
[33] G. Alves *et al.*, "A study on the 4D sparsity of JPEG Pleno light fields using the discrete cosine transform," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1148–1152.
[34] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, 2016, pp. 1–2.
[35] W. Ahmad, L. Palmieri, R. Koch, and M. Sjöström, "Matching light field datasets from plenoptic cameras 1.0 And 2.0," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video*, Jun. 2018, pp. 1–4.
[36] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, and M. Carli, "SMART: A light field image quality dataset," in *Proc. 7th Int. Conf. Multimedia Syst.*, 2016, p. 49.
[37] I. Viola, M. Rerabek, and T. Ebrahimi, "Comparison and evaluation of light field image coding approaches," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 7, pp. 1092–1106, Oct. 2017.
[38] W. Ahmad, S. Vagharshakyan, M. Sjöström, A. Gotchev, R. Bregovic, and R. Olsson, "Shearlet transform based prediction scheme for light field compression," in *Proc. Data Compress. Conf.*, Snowbird, UT, USA, Mar. 2018, p. 396.
[39] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 133–147, Jan. 2018.
[40] Y. Li, R. Olsson, and M. Sjöström, "An analysis of demosaicing for plenoptic capture based on ray optics," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video*, Jun. 2018, pp. 1–4.
[41] V. Vaish and A. Adams. *The New Stanford Light Field Archive*. Accessed: Jan. 8, 2018. [Online]. Available: http://lightfield.stanford.edu/lfs.html
[42] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. Asian Conf. Comput. Vis.* Springer, 2016, pp. 19–34.
[43] W. Ahmad, R. Olsson, and M. Sjostrom, "Towards a generic compression solution for densely and sparsely sampled light field data," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 654–658.
[44] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.ion*, vol. 1, no. 1, pp. 7–55, 1987.
[45] G. Bjontegaard, *Calculation of average PSNR differences between RD-curves*, ITU SG16 document VCEG-M33, 2001.
[46] HM. *HEVC Reference Software*. Accessed: Jan. 8, 2018. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9/
[47] Multicoreware. *X265 HEVC Encoder*. Accessed: Jan. 8, 2018. [Online]. Available: https://bitbucket.org/multicoreware/x265/
[48] D. Lin and P. Chau, "Objective human visual system based video quality assessment metric for low bit-rate video communication systems," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct. 2006, pp. 320–323.
[49] Supplementary Data. *Bunny LF Image Compressed at Rate R4*. Accessed: Jan. 31, 2019. [Online]. Available: https://miun.box.com/s/chtxv9pbba0eu1sqyjlvm86ov22vwzuo

Waqas Ahmad received the M.Sc. degree in electronics engineering from Mohammad Ali Jinnah University in 2012. He is currently pursuing the Ph.D. degree with the Department of Information Systems and Technology (IST), Mid Sweden University–Sundsvall. His research interests are in the area of light field acquistion, processing, and compression.

Suren Vagharshakyan received the M.Sc. degree in mathematics from Yerevan State University in 2008. He is currently pursuing the Ph.D. degree with Tampere University. His research interests are in the area of light field capture and reconstruction.

Mårten Sjöström received the M.Sc. degree in electrical engineering and applied physics from Linkping University, Sweden, in 1992, the Licentiate of Technology degree in signal processing from the KTH Royal Institute of Technology, Stockholm, Sweden, in 1998, and the Ph.D. degree in modeling of nonlinear systems from the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2001. He was an Electrical Engineer with ABB, Sweden, from 1993 to 1994, a Fellow with CERN from 1994 to 1996, and a Ph.D. Student at EPFL, from 1997 to 2001. In 2001, he joined Mid Sweden University–Sundsvall. He was appointed as an Associate Professor and a Full Professor of Signal Processing in 2008 and 2013, respectively. He has been the Head of Computer and System Sciences with Mid Sweden University–Sundsvall since 2013. He founded the Realistic 3D Research Group in 2007. His current research interests are within multidimensional signal processing and imaging and system modeling and identification.

Atanas Gotchev (Member, IEEE) received the M.Sc. degree in radio and television engineering, the M.Sc. degree in applied mathematics, and the Ph.D. degree in telecommunications from the Technical University of Sofia, in 1990, 1992, and 1996, respectively, and the D.Sc. (Tech) degree in information technologies from the Tampere University of Technology in 2003. He is currently a Professor with Tampere University. His recent work concentrates on algorithms for multisensor 3D scene capture, transform-domain light-field reconstruction, and Fourier analysis of 3D displays.

Robert Bregovic (Member, IEEE) received the M.Sc. degree in electrical engineering from the University of Zagreb, in 1998, and the D.Sc. (Tech) degree in information technology from the Tampere University of Technology, in 2003. He has been with Tampere University (former Tampere University of Technology) since 1998. His research interests include the design and implementation of digital filters and filterbanks, multirate signal processing, and topics related to acquisition, processing/modeling, and visualization of 3D content.

Roger Olsson received the M.Sc. degree in electrical engineering and the Ph.D. degree in telecommunication from Mid Sweden University–Sundsvall, Sweden, in 1998 and 2010, respectively. He was with video compression and distribution industry from 1997 to 2000. He was a Junior Lecturer with Mid Sweden University–Sundsvall from 2000 to 2004, where he taught courses in telecommunication, signals and systems, and signal and image processing. Since 2010, he has been a Researcher with Mid Sweden University–Sundsvall. His research interests include plenoptic image capture, processing, and compression; plenoptic system modeling, and depth map capture and processing.