High-Quality Shearlet Transform based Light Field Compression Under Very-Low Bitrates

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

***Abstract*—Light field (LF) acquisition devices capture spatial and angular information of the scene. In contrast with traditional cameras, the additional angular information enables novel post-processing applications such as 3D scene reconstruction, refocusing 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 are divided into two groups, i.e. key views and decimated views. The key views are compressed using multi-view extension of High Efficiency Video Coding (MV-HEVC) scheme and decimated views are predicted using the shearlet transform based prediction (STBP) scheme. Additionally, the residual information of predicted views is also encoded and sent along with the coded stream of key views. The proposed scheme is evaluated over benchmark multi-camera based LF dataset and it is demonstrated that incorporating the residual information into compression scheme increases the overall PSNR by 2 dB. The proposed compression scheme performs significantly better in low bit-rates compared to anchor schemes whose compression efficiency is better in high bit-rate scenarios. The sensitivity of the human vision system towards compression artifacts specifically in low bit-rates favors the proposed compression scheme over the anchor schemes.**

***Index Terms*—Light field, Plenoptic, Compression, Multi-camera, MV-HEVC, Shearlet transform**

I. Introduction

The spatial and angular information of the scene has attracted significant attention in various 3D capturing [1], [2], processing [3] and rendering applications [4]–[7]. The idea to capture angular information along with spatial information was initially proposed by G. Lippmann in 1908 [8]. With the advancement in computing technology, LF was captured using multiple traditional cameras [9]. Each camera captures a single perspective of the scene and 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 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 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 the LF data. However, such encoders do not take into account the correlation present in LF data and 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 to efficiently compress the 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 on both types of captured data [13], [14].

In 2016, a handful of compression schemes were presented as a response to the grand challenge on plenoptic image compression [15]–[19]. Majority of the presented schemes introduced novel tools in standard HEVC image encoder to compress a plenoptic image. Li et al. [16], [20] proposed a bi-prediction mode capability within HEVC image compression framework for compression of plenoptic images. In addition to 33 intra prediction modes in HEVC, each block was allowed to take prediction from already encoded blocks. A similar approach was proposed by Monteiro et al. in which two novel tools were added into the HEVC image compression scheme. Each block can take prediction from other blocks by using the self-similarity (SS) and local linear embedding (LLE) operators [18]. A SS-only prediction scheme was incorporated in HEVC by Conti et al. [19]. Following the idea of pseudo-video sequence (PVS) as initially proposed by Olsson et al. [21], an alternative approach was proposed by Liu et al. in which a plenoptic image was converted into sub-aperture images and treated as frames of a PVS [17]. The HEVC video encoder was used to encode the PVS, and the scheme was selected as the best proposal in the 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 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 to interpret sub-aperture images as a frame of multi-view sequences and performed compression using MV-HEVC [13]. A two-dimensional prediction and rate allocation schemes were proposed to improve the compression efficiency. Tabus et al. [22] exploited the disparity information of input plenoptic image to increase compression efficiency. The disparity map was quantized into several regions and displacement of each region of side view relative to central view was estimated. A set of sparse views, disparity map corresponding to central view and region displacements of side views were encoded. A pixel-level correlator was developed to further refine the side views from corresponding neighbor

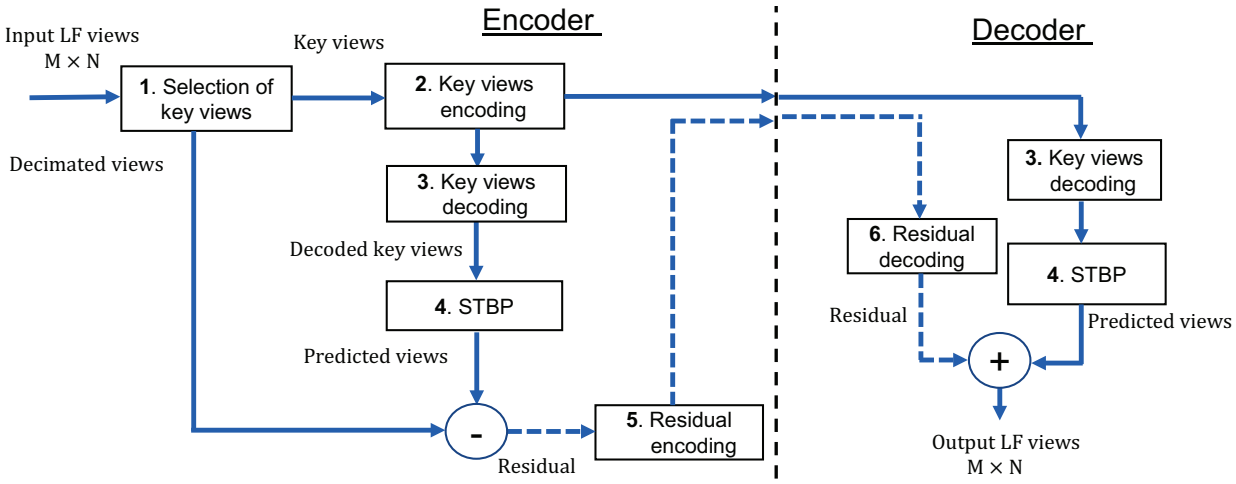


Fig. 1. Block diagram of proposed compression scheme. At encoder side, the proposed scheme categorizes the input LF views into key views and decimated views. Key views are compressed and their corresponding decoded views are used to predict the decimated views by applying the shearlet transform based prediction scheme. Residual information is calculated between decimated views and predicted views and compressed separately along with key views. At decoder side, 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.

views. The compression scheme proposed by Zhao et al. [23] categorizes the sub-aperture images into two groups, i.e. 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 [24]. The decoded version of PVS was again converted into a plenoptic image and the residual information was estimated and transmitted in order to enhance the visual quality.

Beside compression proposals for plenoptic images in the grand challenges, Li et al. proposed a memory-optimized 2D hierarchical coding structure for plenoptic image compression [25]. 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. [26] proposed a scalable coding scheme for compression of plenoptic images captured with 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 using HEVC inter-prediction scheme. Bakir et al. [27] presented plenoptic image compression scheme in which the input sub-aperture images were divided into two groups. First group was encoded using HEVC encoder and second group was estimated using linear approximation of already encoded sub-aperture images. At the decoder side, additional deep Learning based scheme was used to improve the reconstruction quality of sub-aperture images. Few researchers have proposed 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 [28]. A sparse set of views were compressed as a PVS and were used to predict the remaining views. Xian et al. [14] proposed a compression solution based on the homography information between side views and the central view. A joint optimization problem was set up in which those homographies were estimated that minimized the low-rank approximation error. Ahmad et al. proposed to interpret LF captured with the multi-camera system as frames of multiple PVS and compressed using MV-HEVC [29]. 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. [30] proposed a simple computational efficient scalable coding scheme for multi-camera based LF data. A a set of binary images were chosen to record the common structure among all views, and the difference among the views were represented with additional weight images. The number of binary images was provided as a free parameter in scalable coding framework that controls the trade-off between quality and computational complexity. Alves et al. [31] analyzed the redundancy in plenoptic images and multi-camera based LF data using 4D DCT transform.

The grand challenges for plenoptic image compression and the availability of plenoptic image datasets resulted in numerous compression solutions. However, LF captured with multi-camera systems have received less attention from the research community. The results in ICME grand challenge [32] reflect that it is important for a LF compression scheme to perform better in low bit-rates. It can be observed that most of the previously mentioned compression schemes [22], [23], [28], [33] used a sub-set of views to generate the remaining views. In this way, the reconstruction algorithms were integrated into the compression framework.

We propose a compression scheme for LF data captured with a multi-camera system that addresses the compression efficiency at low bit rates. The proposed scheme uses epipolar plane image (EPI) representation of a subset of input LF views and predicts the remaining views by applying shearlet transform in frequency domain. This paper explains in more

detail the initial work in [33] and improves the compression efficiency by incorporating the residual information. In Section II, main elements of the proposed compression scheme are discussed. The sub-sections explain the selection and encoding of key views, shearlet transform based prediction scheme and residual coding scheme. Section III explains the test conditions and the experimental setup. In Section IV, experiment results are reported and discussed, i.e. encoding of selected key views, effects of compression artifacts on shearlet transform, prediction of decimated views, residual coding and complexity analysis of proposed compression scheme. Finally, Section V presents the major conclusions.

II. Proposed Compression Scheme

The block diagram of the proposed compression scheme is presented in Fig. 1. The LF with MxN views is given as an input to the compression scheme. The input views are divided into two categories, i.e. keys views and decimated views. The Shearlet transform based prediction is applied to the decoded version of keys 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 is selected from input LF by following the procedure explained in Section II-A. From hereafter, we called selected sparse views as *key views*. 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 [34]. The compression scheme takes the multi-view pseudo video sequences and use tools available in MV-HEVC to exploits the 2D correlation present in LF data.

3. Decoding of key views: The MV-HEVC based encoder in block 2, 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 scheme: The decimated views are recreated by predicting them from the decoded key views using the STBP. Detailed description of the STBP is explained in Section II-B.

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

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

A. Key views selection and Encoding

The captured LF with MxN views is uniformly decimated by factor s in both horizontal and vertical directions, resulting into a sparse set of M_sxN_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.

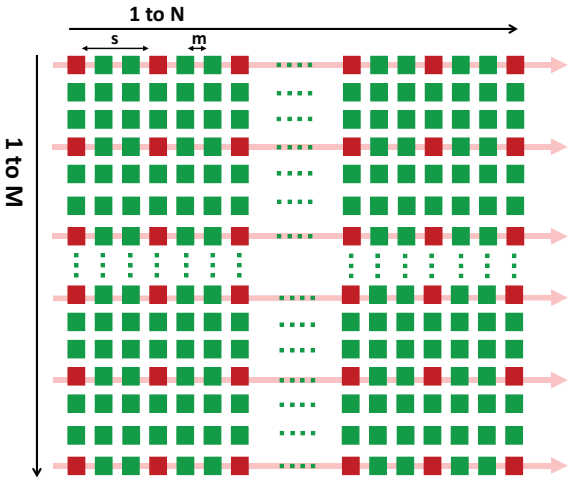


Fig. 2. The input LF with MXN views is uniformly sub-sampled with factor s in both directions to get M_sXN_s key views.

The key views are compressed using state-of-the-art MV-HEVC by following the method proposed in [34]. 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 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, Decoding Order (DO) and View Order Index (VOI) represents the decoding order of each frame in the horizontal and vertical axis. The POC and VID axis are assigned with different predictor levels. Fig. 3 shows an example of 5x5 key views, the central frame with POC=2 and VID=2 is taken as a base frame and assigned with prediction level 0. The remaining frames are assigned with either prediction level 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 prediction of other frames in order to improve the overall compression efficiency.

Algorithm 1, explains the rate allocation scheme used to encode the key views. The algorithm inputs are: number of pseudo video sequences (M_s), number of frames in each pseudo video sequence (N_s), base view POC (b_{POC}), View ID (b_{VID}) and quantization parameter (Q_b). 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 VID=2 (lies in base column or base row) are assigned quantization offset equivalent to their prediction level (0, 1 or 2). The quantization offset for remaining frames is calculated by using the frame distance and decoding distance with respect to the base frame (calculated in the line 22). Line 10 and 11 of algorithm 1 calculates frame distance between current frame and the base frame in POC and ViewID axis. Similarly, from line 12-20 decoding distance between the

TABLE I
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

current and the base frame in both POC and ViewID axis is calculated. A weight parameter W is also used to control the limit of quantization offset. Table I shows the parameter W assigned to each frame based on 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.

Algorithm 1 Rate allocation for Key views

Input: M_s, N_s, b_{POC}, b_{VID}, Q_b

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: $Q_{o(x,y)} = max(s_{\text{POC}}(x), t_{\text{VID}}(y))$
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** $n_{\text{POC}} \leq b_{\text{POC}}$ **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** $v_{\text{VID}} \leq b_{\text{VID}}$ **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: $Q_{o(x,y)} = d_{\text{POC}} + d_{\text{VID}} + d_{\text{DO}} + d_{\text{VOI}}$
23: **end**
▷ Quantization parameter for current frame
24: $Q_{(x,y)} = Q_b + Q_{o(x,y)}$
25: **end**
26: **end**
27: *Output:* Q

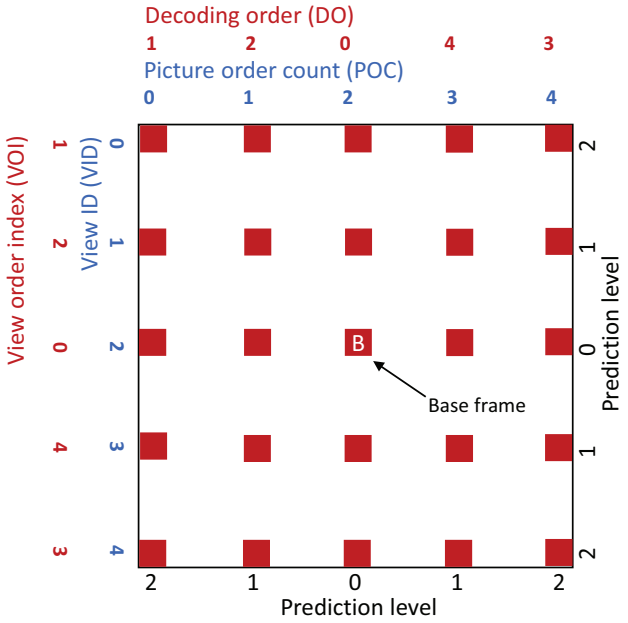


Fig. 3. The rate allocation scheme for 5x5 key views. The central view with POC=2 and VID=2 is chosen as base frame.

B. Shearlet transform based prediction scheme

In order to significantly reduce the LF data, a good prediction scheme is required. In our procedure, we consider a light field reconstruction algorithm utilizing shearlet transform developed for reconstruction of densely sampled light filed [35]. A densely sampled light field means that disparity between adjacent views is no more than 1 pixel apart. This property allows for obtaining 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.

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

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

where (u, v) plane is representing image plane coordinates for each view, and (s, t) are coordinates of capturing plane as shown in Fig. 4(a). By fixing (u, s) and (v, t) parameters horizontal and vertical epipolar plane images [36] are formed as follows

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

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

In general, it is assumed to have sufficient sampling over image plane (u, v), such that cameras provide enough resolution to capture the finest details of the scene.

In proposed approach, as LF reconstruction tool, we consider EPI reconstruction using shearlet transform presented in [35]. Intermediate views reconstruction in 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 sample EPI from

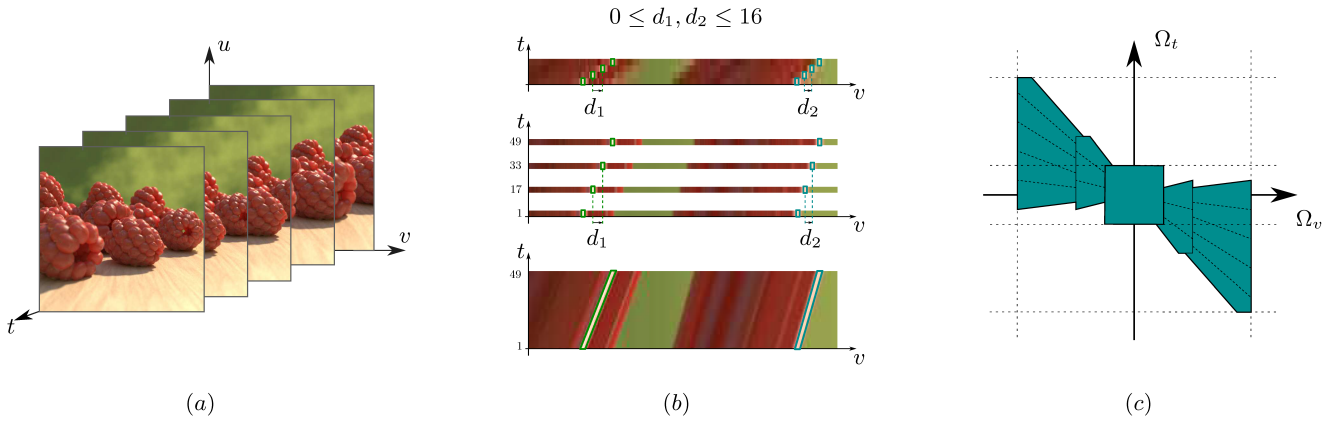


Fig. 4. (a) Parameterization of captured images, (b) interpretation of input images as decimation of densely sampled light field, (c) example of frequency plane tilling using shearlet transform required for efficient densely sampled light filed reconstruction.

coarsely sampled EPIs, as illustrated in Fig. 4(b) for horizontal parallax case.

Due to very strict structure of DSLF in spatial and frequency domain, 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, which allows to construct desirable frequency domain tilling Fig. 4(c). Let's assume that measurements g obtained using M measurement matrix applied on ground truth densely sampled EPI f such that

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

where $\odot$ representing element-wise multiplication. As shown in [35] 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 hard thresholding operator, S and S^* are direct and inverse shearlet transform respectively. The algorithm 2 explains the shearlet transform based reconstruction process for single EPI image. The inputs of the algorithm are number of iterations, EPI image, mask (indicates key views pixels), shearlet analysis and synthesis filters. In 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 synthesis part, the Fourier transforms of each selected coefficient is computed and multiplied with the corresponding shearlet synthesis filter. The summation is computed for all the responses and 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 next iteration and algorithm steps are repeated for N iterations. More details about construction of transforms, parameters and iterative procedure can be find in [35]. It is important to notice that construction and computation of S and S^* transforms are mainly based on $d_{range} = d_{max} - d_{min}$ - 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

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

C. Residual Encoding

The STBP scheme provides significant compression efficiency in low bit-rates. However, in high bit-rates proposed prediction scheme has an inherit 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 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 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

Algorithm 3 Residual sequence generation

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

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

III. Test arrangement and evaluation criteria

The experimentation was performed on the light field dataset provided by Stanford University [37]. Table II shows the selected LF images from the Stanford dataset. Each LF image contains 17x17 views in RGB format and its equivalent YUV444 format was used as a reference input signal. The reference input signal was further converted into the YUV420 format and given as an input to the proposed and anchor compression schemes. The shearlet transform has filtering artifacts on image border as described in [35]. 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).

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

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

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

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 image, W and H indicate the width and height of each view respectively. T represents the number of pixels of each view considered for comparison ($T = (W-2*b)(H-2*b)$). $I(x,y)$ and $I'(x,y)$ represent the value of pixel in original view and reconstructed view. The BD-PSNR [38] metric is also used to compare the compression results. The compression efficiency of proposed scheme was evaluated against the state-of-art compression scheme [29] and two benchmark HEVC [39] and X265 [40] anchor schemes. The LF views were converted into a single PVS and given as input to the benchmark anchor schemes. The first frame was encoded as intra-frame, second as P-frame and all the remaining frames were encoded as B-frames.

TABLE II
Selected LF images from Stanford dataset

S/N	Image Name	Resolution (WxH)
1	Chess	1400x800
2	Lego Buildozer	1536x1152
3	Eucalyptus Flowers	1280x1536
4	Amethyst	768x1024
5	Bunny	1024x1024
6	Jelly Beans	1024x512

IV. Results and analysis

A. Encoding of key views

The initial step of the proposed compression scheme is to compress the key views by using MV-HEVC as explained in Section II-A. Alternatively, the key views were also converted into a single PVS and compressed using HEVC encoding scheme. Two LF images from Standford dataset, namely Chess and Eucalyptus Flowers were compressed on four different bit-rates in order to test varying bit-rate scenarios. The RD comparison between the proposed scheme and HEVC scheme is presented in Fig. 5. It can be seen that the proposed scheme provides better compression efficiency compared to 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 neighbouring views. Moreover, by allocating better quality to frames that were used for prediction of other frames improves compression efficiency. The compression efficiency of the proposed scheme will improve with the increase in the number of key views since more frames will take prediction 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 of quality among images as a consequence of compression process can affect the reconstruction process employing EPI images. An experiment was performed in order to study the effect of variable rate allocation on the STBP process. The truck image from Standford

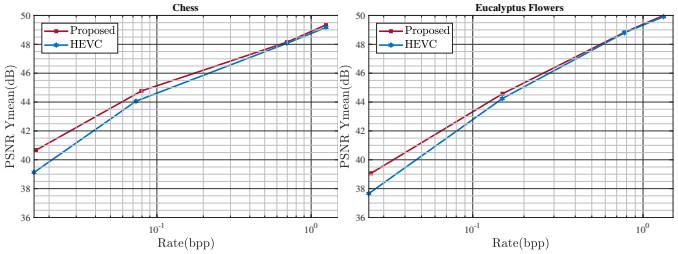


Fig. 5. Rate-distortion analysis between proposed compression scheme and HEVC scheme for 5x5 key views.

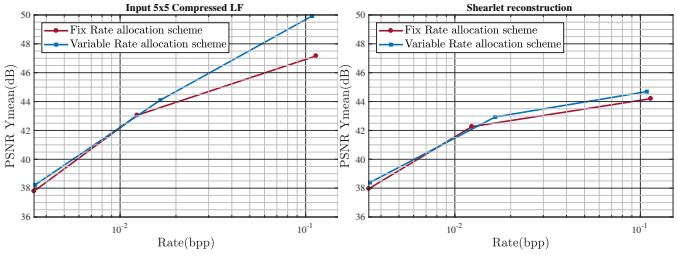


Fig. 6. Rate-distortion analysis of variable and fixed quality rate allocation schemes. a) the input 5x5 key views are compressed using fixed and variable rate allocation schemes. b) Corresponding shearlet reconstruction of 17x17 views for fixed and variable rate allocation schemes.

dataset was used and a subset of 5x5 key views was extracted from 17x17 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 fix quantization parameter was used to have the same quality among images. In the second encoding scheme, variable quantization parameters were used for all 25 views to have variable quality. The compression was performed on three different bit-rates for both fix and variable rate allocation schemes. Fig. 6 (a) and (b) show the RD curves for 5x5 key views and reconstruction results of the STBP respectively. The combined compression efficiency of the variable rate allocation scheme on 5x5 key views and then shearlet reconstruction outperforms the fixed rate allocation scheme. Hence, the key views compressed with variable rate-allocation scheme can be used as input of STBP.

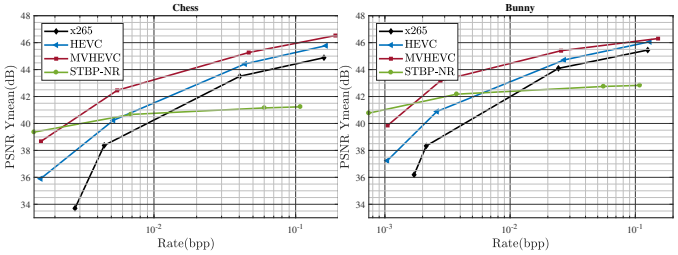


Fig. 7. Rate-distortion analysis between shearlet transform based prediction scheme and anchor schemes.

C. Prediction of decimated views using shearlet transform

Fig.7 shows a RD comparison between the STBP scheme (without incorporating residual information) and anchor schemes. The performance of anchor schemes is better in high bit-rates compared to the proposed scheme. However, in 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, high quality of residual information enables the anchor schemes to achieve efficient compression. On the contrary, shearlet transform relies on key views to predict the intermediate views without incorporating residual information, 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.

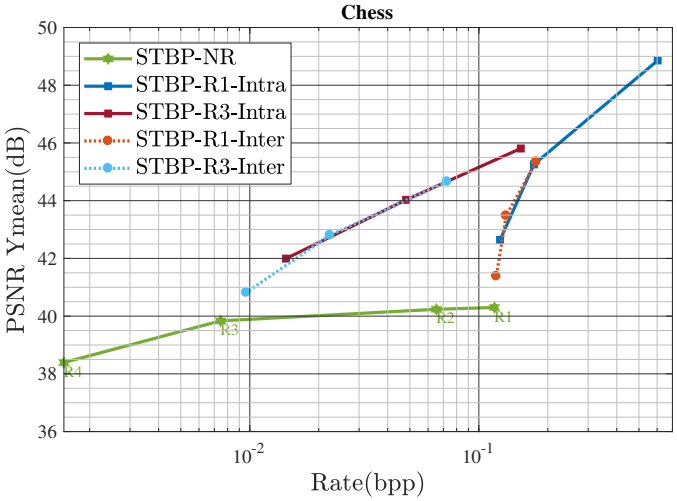


Fig. 8. Rate-distortion analysis between HEVC intra-prediction and inter-prediction coding.

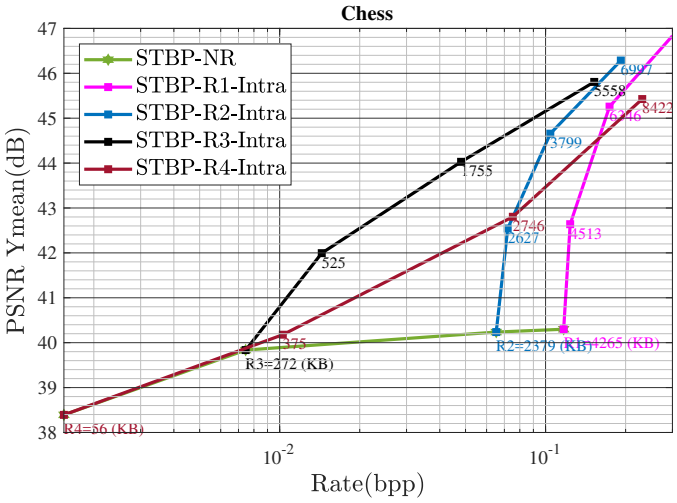


Fig. 9. Rate-distortion analysis of shearlet transform based prediction scheme with residual information added to predicted views at rate R1, R2, R3 and R4.

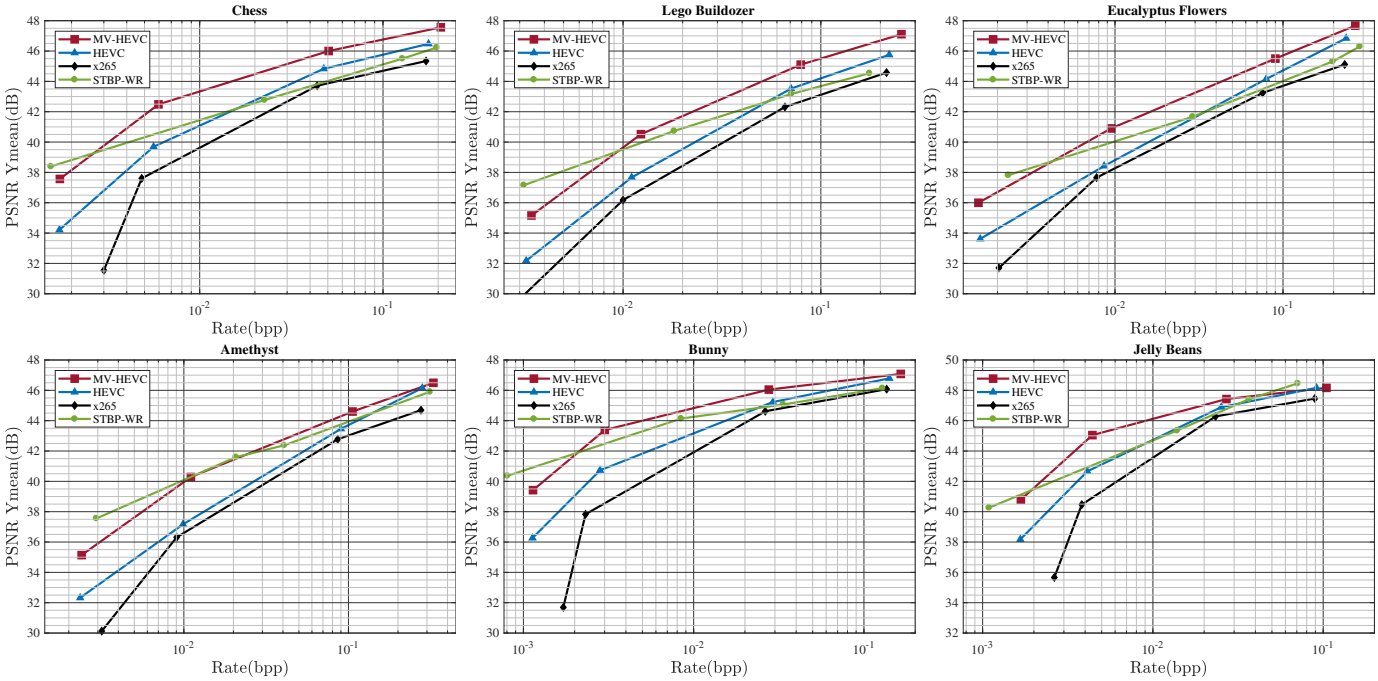


Fig. 10. Rate-distortion analysis of proposed compression scheme (STBP-WR) with two benchmark anchor schemes (HEVC and X265) and a state-of-art compression scheme (MV-HEVC).

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 that works similar to HEVC for such an input. In order to exploit the correlation present in residual information using tools available in HEVC, the residual sequence was encoded with intra-prediction 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 given input to the STBP were encoded on four different bit-rates and that also describes the rate of STBP-NR (since no residual information is added). Hereafter, we called these four rates as 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 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 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 shows a similar compression efficiency between inter-prediction and intra-prediction schemes. The inter-prediction makes use of motion estimation and compensation for predicting the current frame which is not beneficial for residual PVS since it doesn't possess properties like natural images. The similar compression efficiency between two schemes reflects less correlation among frames of residual PVS. Hence, we proposed to use HEVC intra-prediction mode to encode the residual PVS. The HEVC intra-prediction scheme has 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 shows the enhancement in visual quality obtained due to the addition of residual information. THE STBP-NR represents the reference STBP scheme, evaluated on 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 is added with predicted views. Similarly, for the other three rates (R2, R3, and R4) the residual information corresponding to each rate is encoded with different quantization parameters and its decoded version is added with the corresponding predicted views. The response of residual coding is not same for all four bit-rates. Fig. 9 shows significant compression efficiency at higher rates compared to lower rates. The performance of the STBP in low-bit rate is much better (as shown in 7) and adding residual with Intra coding method doesn't improve visual quality relative to the added size of coded residual. The prediction efficiency of the STBP starts decreasing from R3 and it gets very low at high bit-rates (R2 and R1). In other words, the input quality of key views at high bit-rate has less influence on the STBP. For examples, the coded size of key views at R2 is 2379 KB and increasing the quality of key views by allocating extra 1886 KB (at R1=4265 KB) has less impact on visual quality (around 0.1 dB increase). In comparison, adding 248 KB of residual information at R2 has a significant impact on overall visual quality (around 2 dB increase). It can be concluded that adding residual information at high bit-rates improves the compression efficiency of the proposed scheme.

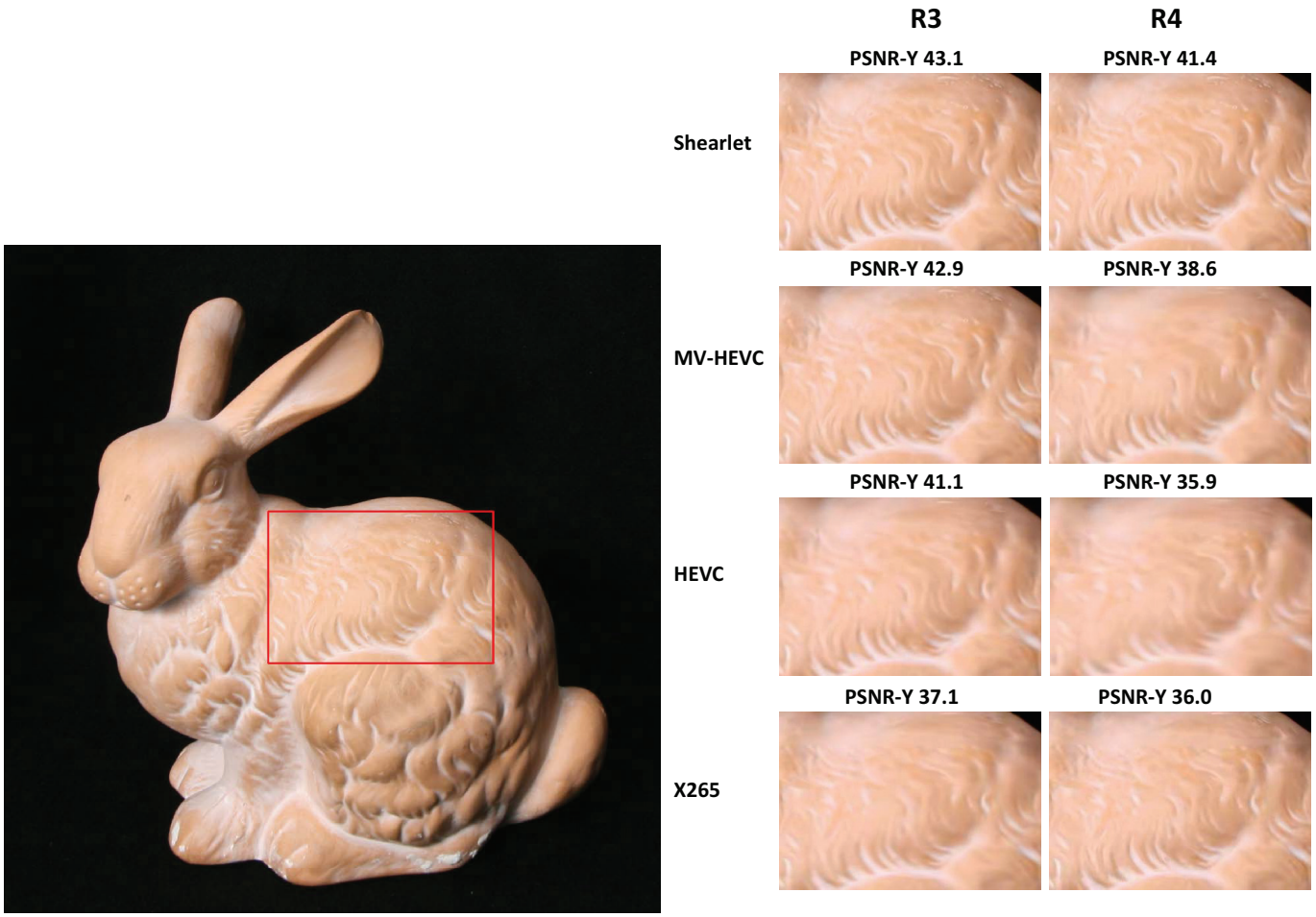


Fig. 11. Subjective analysis of the proposed compression scheme for a view from Bunny image. a) Shows the original view with highlighted subregion. b) Shows the rendered subregion compressed with proposed scheme, MV-HEVC, HEVC and X265 compression schemes. Two low-bitrates (R3 and R4) are considered and the PSNR of each view is also mentioned for each compression scheme.

Fig. 10 shows the comparison of the proposed compression scheme with two benchmark schemes and a state-of-the-art scheme [29]. 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 in low and medium bit-rates relative to the high bit-rates scenarios. In comparison to the state-of-the-art scheme, the proposed scheme has better compression in low bit-rates and the scheme presented in [29] perform better in medium and high bit-rates. The sensitivity of the human vision system towards compression artifacts specifically in low bit-rates [41] favors the proposed compression scheme over other presented compression schemes.

E. Subjective Analysis

A sub-region of Bunny image taken from view (6,9) is shown in Fig. 11 for all the compression schemes at rate R3 and R4. The HEVC and X265 based compression schemes show notable blurriness in the decoded view compared to MV-HEVC based scheme. At both rates, it can be seen that the proposed scheme retains most of the information compared to all the other compression schemes.

F. Computational Complexity

The computational complexity of the proposed compression scheme is dependent on usage of residual information. In the case when residual information is not used in proposed scheme the encoder will compress only key views using MV-HEVC. At the decoder side, the key views will be decoded using MV-HEVC decoder and the STBP will be used to predict the decimated views. In such compression scheme, a significant complexity is reduced at the encoding side because only key views (8.5% of input LF) will be compressed. At 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 to use the STBP process at encoder side and the residual information will be coded using MV-HEVC single layer intra-prediction mode. At the decoding side, the key processes will be decoding of key views by using MV-HEVC, STBP for prediction of decimated views and decoding of the residual bitstream. Hence, the enhancement in the visual quality of predicted views is obtained at the increased computational cost.

V. Conclusion

In this paper, we present a novel compression solution for LF data captured with the multi-camera system. The input LF views were divided into two categories, i.e. key views and decimated views. The key views were encoded using MV-HEVC and decimated views were predicted using the shearlet transform based prediction scheme. Additionally, the residual information was also coded in order to further enhance the visual quality of the predicted views. The proposed compression scheme perform better in low bit-rates compared to anchor schemes whose compression efficiency is better in high bit-rate. The sensitivity of the human vision system towards compression artifacts in 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 case, at the encoder side, key views will be coded and sent to decoding side where shearlet transform based prediction scheme will predict the remaining decimated views. The proposed compression scheme can benefit applications where fewer resources are available at the encoding side. The proposed scheme can be further improve by introducing coding tools that exploits the correlation present in residual information. In future, we will investigate alternative compression of the residual information.

Acknowledgment

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.

References

[1] M. Ziegler, R. op het Veld, J. Keinert, and F. Zilly, "Acquisition system for dense lightfield of large scenes," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2017*. IEEE, 2017, pp. 1–4.

[2] M. Martínez-Corral, J. Barreiro, A. Llavador, E. Sánchez-Ortiga, J. Sola-Pikabea, G. Scrofani, and G. Saavedra, "Integral imaging with fourier-plane recording," in *Three-Dimensional Imaging, Visualization, and Display 2017*, vol. 10219. International Society for Optics and Photonics, 2017.

[3] A. Ansari, A. Dorado, G. Saavedra, and M. M. Corral, "Plenoptic image watermarking to preserve copyright," in *Three-Dimensional Imaging, Visualization, and Display 2017*, vol. 10219. International Society for Optics and Photonics, 2017.

[4] S. Hong, A. Ansari, G. Saavedra, and M. Martinez-Corral, "Full-parallax 3d display from stereo-hybrid 3d camera system," *Optics and Lasers in Engineering*, vol. 103, pp. 46–54, 2018.

[5] P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, and T. Balogh, "Towards adaptive light field video streaming," *COMSOC MMTC Communications-Frontiers*, 2017.

[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. 15–1, 2015.

[7] D. G. Dansereau, O. Pizarro, and S. 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.

[8] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

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

[10] R. Ng, M. Levoy, B. Mathieu, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[11] C. Perwass and L. Wietzke, "Single lens 3d-camera with extended depth-of-field," in *Human Vision and Electronic Imaging XVII*, vol. 8291. International Society for Optics and Photonics, 2012, p. 829108.

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

[13] W. Ahmad, R. Olsson, and M. Sjöström, "Interpreting plenoptic images as multi-view sequences for improved compression," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4557–4561.

[14] X. Jiang, M. Pendu, R. Farrugia, S. Hemami, and C. Guillemot, "Homography-based low rank approximation of light fields for compression," in *Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 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 *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[16] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 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 *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[18] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field hevc-based image coding using locally linear embedding and self-similarity compensated prediction," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[19] C. Conti, P. Nunes, and L. Soares, "Hevc-based light field image coding with bi-predicted self-similarity compensation," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[20] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1308–1319, 2016.

[21] R. Olsson, M. Sjöström, and Y. Xu, "A combined pre-processing and h. 264-compression scheme for 3d integral images," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 513–516.

[22] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and jpeg 2000," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4567–4571.

[23] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4562–4566.

[24] C. Jia, Y. Yang, X. Zhang, X. Zhang, S. Wang, S. Wang, and S. Ma, "Optimized inter-view prediction based light field image compression with adaptive reconstruction," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4572–4576.

[25] 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 Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1107–1119, 2017.

[26] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 80–91, 2016.

[27] N. Bakir, W. Hamidouche, O. Déforges, K. Samrouth, and 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.

[28] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable light field compression scheme using sparse reconstruction and restoration," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3250–3254.

[29] W. Ahmad, M. Sjöström, and R. Olsson, "Compression scheme for sparsely sampled light field data based on pseudo multi-view sequences," in *Optics, Photonics, and Digital Technologies for Imaging Applications V*, vol. 10679. International Society for Optics and Photonics, 2018, p. 106790M.

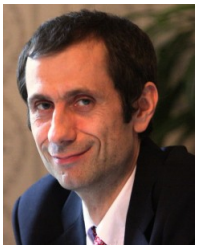
[30] K. Komatsu, K. Takahashi, and T. Fujii, "Scalable light field coding using weighted binary images," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 903–907.
[31] G. Alves, M. P. Pereira, M. B. de Carvalho, F. Pereira, C. L. Pagliari, V. Testoni, and E. A. da Silva, "A study on the 4d sparsity of jpeg pleno light fields using the discrete cosine transform," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 1148–1152.
[32] I. Viola, M. Rerabek, and T. Ebrahimi, "Comparison and evaluation of light field image coding approaches," *IEEE Journal of selected topics in signal processing*, vol. 11, no. EPFL-ARTICLE-230995, 2017.
[33] W. Ahmad, S. Vagharshakyan, M. Sjöström, A. Gotchev, R. Bregovic, and R. Olsson, "Shearlet transform based prediction scheme for light field compression," in *Data Compression Conference (DCC 2018), Snowbird, Utah, US, March 27-March 30, 2018*, 2018.
[34] W. Ahmad, R. Olsson, and M. Sjostrom, "Towards a generic compression solution for densely and sparsely sampled light field data," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 654–658.
[35] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 133–147, 2018.
[36] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.
[37] V. Vaish and A. Adams, "The new stanford light field archive, [online]," http://lightfield.stanford.edu/lfs.html, Accessed = 2018-08-01.
[38] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *ITU SG16 Doc. VCEG-M33*, 2001.
[39] HM, "HEVC reference software, [online]," https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9/, Accessed = 2018-08-01.
[40] Multicoreware, "X265 hevc encoder, [online]," https://bitbucket.org/multicoreware/x265/, Accessed = 2018-08-01.
[41] D. Lin and P. Chau, "Objective human visual system based video quality assessment metric for low bit-rate video communication systems," in *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006, pp. 320–323.

Waqas Ahmad received the MSc degree in Electronics engineering from Mohammad Ali Jinnah university in 2012. He is working toward the PhD degree at the Department of Information Systems and Technology (IST) at Mid Sweden University since 2016. His research interests are in the area of light field compression.

Suren Vagharshakyan received the MSc degree in mathematics from Yerevan State University, in 2008. He is working toward the PhD degree at the Department of Signal Processing at Tampere University of Technology since 2013. 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 cole Polytechnique Fdrale 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, Lausanne, Switzerland, from 1997 to 2001. In 2001, he joined Mid Sweden University, and 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 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 received the MSc degrees in radio and television engineering (1990) and applied mathematics (1992) and the PhD degree in telecommunications (1996) from the Technical University of Sofia, and the DSc (Tech) degree in information technologies from the Tampere University of Technology (2003). He is a professor at Tampere University of Technology. His recent work concentrates on algorithms for multisensor 3-D scene capture, transform-domain light-field reconstruction, and Fourier analysis of 3-D displays. He is a member of the IEEE.



Robert Bregovic received the MSc degree in electrical engineering from University of Zagreb, in 1998, and the DrSc(Tech) degree in information technology from Tampere University of Technology, in 2003. He has been working at 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. He is a member of the IEEE.

Roger Olsson received the M.Sc. degree in electrical engineering and the Ph.D. degree in telecommunication from Mid Sweden University, Sweden, in 1998 and 2010, respectively. He was with the video compression and distribution industry from 1997 to 2000. He was a Junior Lecturer with Mid Sweden University 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. His research interest includes plenoptic image capture, processing, and compression; plenoptic system modeling; and depth map capture and processing.