Li Y.; Sjöström, M.; Olsson, R; Jennehag, U., "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Transactions on Image Processing*, 2015

http://dx.doi.org/10.1109/TIP.2015.2498406

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

Scalable coding of plenoptic images by using a sparse set and disparities

Yun Li, Mårten Sjöström, *Member, IEEE*, Roger Olsson, *Member, IEEE*, and Ulf Jennehag

***Abstract*—One of the light field capturing techniques is the focused plenoptic capturing. By placing a microlens array in front of the photosensor, the focused plenoptic cameras capture both spatial and angular information of a scene in each microlens image and across microlens images. The capturing results in significant amount of redundant information, and the captured image is usually of a large resolution. A coding scheme that removes the redundancy before coding can be of advantage for efficient compression, transmission and rendering. In this paper, we propose a lossy coding scheme to efficiently represent plenoptic images. The format contains a sparse image set and its associated disparities. The reconstruction is performed by disparity-based interpolation and inpainting, and the reconstructed image is later employed as a prediction reference for the coding of the full plenoptic image. As an outcome of the representation, the proposed scheme inherits a scalable structure with three layers. The results show that plenoptic images are compressed efficiently with over 60 percent bit rate reduction compared to HEVC intra, and with over 20 percent compared to HEVC block copying mode.**

***Index Terms*—Plenoptic, light field, HEVC, compression,**

I. Introduction

A sampling of the light field with the directions and the intensities of outgoing radiances from a scene is captured by plenoptic cameras. The capability of image refocusing and multi-view imaging during post-production is enabled by the capturing process. However, a densely sampled plenoptic image contains repetitive patterns with a large resolution. The image can possibly be represented by a subset of its microlens images plus disparity information. The question is if a plenoptic image can be encoded efficiently by using such a representation with a proper sampling factor, and if scalability with respect to transmission and rendering is attainable.

The plenoptic function $I = P(x, y, z, \theta, \phi, \omega, t)$ [1] has seven dimensions and captures the intensities I of light rays at any viewing positions x, y, z, any directions θ, ϕ, any wavelengths ω, and any time t. Representing the color by RGB channels and for a static scene, the plenoptic function is reduced to five dimensions without ω and t. If we further assume regions are free of occluders, the plenoptic function can be simplified into four dimensions as a light field [2] [3], which is represented by a two-plane representation. The four dimensions (x, y), and (r, t) locate the coordinates of radiance passing through the two planes, respectively. There are currently four techniques for capturing a light field image, i.e., by using multi-camera arrays [4], moving cameras [5], coded apertures [6], and microlens arrays [7]. In the capturing with microlens arrays, two capturing techniques are further derived, which are standard plenoptic capturing [7] and plenoptic 2.0 [8]. Cameras with plenoptic 2.0 techniques [8] are also referred to as focused plenoptic cameras.

The concept of plenoptic capturing was first introduced by Gabriel Lippmann in 1908 [9]. A commercially available product is the Lytro camera from Lytro, Inc. founded by Ng et al. [7] in 2001. The first generation of Lytro cameras are standard plenoptic systems [10], and by our visual inspection of the captured images, so are the new generation of Lytro cameras, Illum. Because the focal plane of the microlens is on the camera image sensor plane for the standard plenoptic capturing, the camera only captures angular information in each microlens image, also called Elemental Image (EI), for a single point in the 3D space. This results in a low spatial resolution of rendered views in theory. Focused plenoptic cameras capture, however, both angular and spatial information in each EI and across EIs by putting the focal plane of microlenses away from the image sensor plane. Thus, it provides a trade-off between spatial and angular information for the capturing. The details of focused plenoptic cameras are discussed in Section II.

Plenoptic cameras have gradually gained popularity in the consumer market due to its portability and usability. By a fairly simple re-sampling of the captured plenoptic datasets, refocusing and multi-view imaging can be acquired. We refer to plenoptic images as the image captured by focused plenoptic cameras in the context of this paper. In addition, a densely sampled plenoptic image implies that adjacent EIs are highly correlated.

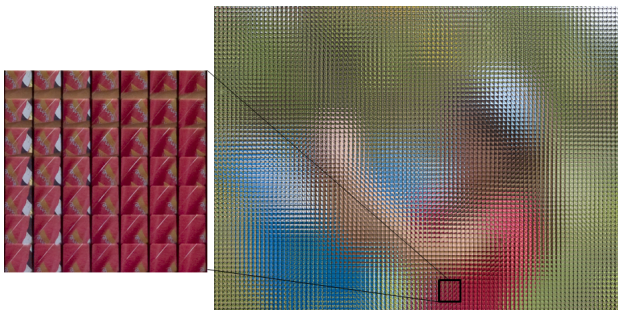


Fig. 1: Focused plenoptic image Laura [11].

A. Motivation

The plenoptic images retain both angular information and spatial information of a scene. The image consists of a grid of EIs whose contents are similar to their neighbours, see

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

Fig. 1. Therefore, one problem with respect to coding is that the image exhibits repetitive patterns, and a large amount of redundancy exists. In a densely sampled plenoptic image, the disparities between adjacent EIs are small, and one EI can be approximated by a shift followed by an interpolation from its neighbors. This implies that a full plenoptic image is possible to be reconstructed from a sparse sample set of its EIs. Thus, a coding approach that removes the redundancy before encoding might be advantageous.

Another problem associated with the coding of plenoptic image lies in that not all EIs are always needed for rendering. Transmitting such a big image frame in the network will likely introduce transmission latency at receiver sides and waste network resources. Furthermore, in the decoder, an increasing of decoding computational complexity will follow. Therefore, a scalable representation of the plenoptic image is desired, so that a quick transmission, decoding and rendering can be performed from a base layer.

As mentioned above, we are motivated by the two problems to devise an approach that can 1) remove the redundancy before coding, 2) encode plenoptic contents efficiently, and 3) provide coding scalability, which is defined in Section IV-C.

B. Previous work

There are other techniques to capture a light field image as mentioned. Coding approaches with respect to light field images in general can be applied to plenoptic images. Previous coding works on light field image compression can be mainly classified into three categories: vector quantization, predictive coding and progressive coding [12]. For the Vector Quantization (VQ) [2] approach, light field images are partitioned into small blocks, which are represented as vectors. A small subset of the vectors is trained to approximate the entire vector space. In predictive coding, an early work [13] arranges light field images into a grid, images within the grid are recursively predicted from a few intra coded images. The prediction efficiency is further improved by using homography [14]. As to the progressive coding, Discrete Wavelet Transform (DWT) is usually applied to achieve a finer granularity of scalability [12] [15] [16]. Shape Adaptive Discrete Wavelet Transform (SA-DWT) was employed in a wavelet scheme with disparity-compensated lifting and shape adaptation [16] to preserve the boundaries of objects in light field images. As light field images can be considered as 4-D contents, 4-D wavelets were used in [17] for the compression.

In addition, there are approaches that do not distinctively lie in any of the categories mentioned above. For example, the performance of light field compression by using distributed coding is evaluated in [18] [19]. The paper in [20] presents a layer approach that segments objects in ray space and applies wavelets for the compression of each segment. Furthermore, Principle Component Analysis (PCA) was also utilized in [21] [22] for de-correlating light field data. In [17], a model-based coding approach was proposed, which represents objects by voxels and exploits geometry information for prediction. In general, hybrid multi-view encoders such as MVC [23] and MV-HEVC [24] can also be applied to efficiently compress light field images.

In order to apply the above mentioned approaches for plenoptic coding, EIs must first be separated out from the original captured image. However, the geometry information for locating the position of each EI is not always available. Therefore, to avoid the separation process, the Self-Similarity (SS) modes [25] were introduced into HEVC and H.264 for plenoptic images [26] [25] and videos [27]. SS modes predict an image block from its neighboring reconstructed blocks. The process is essentially a single hypothesis prediction. Furthermore, HEVC range extension has recently incorporated the Block Copying (BC) mode [28] for coding of screen contents. The BC mode is similar to SS mode with a single hypothesis prediction. However, it has a limitation on the search areas for prediction references. We have also proposed an efficient displacement intra prediction scheme [29] for plenoptic images by using more than one hypothesis, which is effective in reducing the prediction error.

If camera geometry is known, multi-view encoders with hierarchical coding structures in general can be used for coding of plenoptic images. Nevertheless, an obvious drawback of using multi-view encoders directly is that each EI must be padded to the size of a power of two [30] for feeding into the encoder. Because an EI is very small, the padding will result in an unnegligible amount of extra data to be encoded. In addition, the coding performance depends on the coding structure. Our previously proposed displacement intra prediction scheme [29] can efficiently exploit the inter-EIs correlation without considering the coding structure. But, the displacement intra does not provide any scalability for transmission and rendering. In [31], a layered-based approach for light field images typically captured by camera arrays has been proposed. It explores the plenoptic sampling function, performs a non-uniformly spaced layer extraction, and conducts the rendering with a probabilistic interpolation approach. However, this approach is mostly suitable for camera captured light fields. In [32], a scalable approach has been proposed for focused plenoptic images by using the rendered views as prediction references. However, different image processing techniques can be applied on rendered views. As a consequence, the rendered views may not be a good reference. In addition, in this scheme, the coding bit rate for the rendered views (the reference) is not included in the final bit stream. We, therefore, further proposed a coding system that utilizes a sparse set and disparities to address the problem [33] of scalability. Nevertheless, in this coding scheme, disparities are encoded losslessly. As a result, the bit rate allocated to the disparities may be costly when coding plenoptic images at low bit rates, and it is unlikely to reduce the depth bit rate significantly if temporal prediction is considered for videos. Additionally, it is of interest to evaluate the parameter space of the scheme and the scalability with respect to rendering.

C. Proposed method

In this paper, we introduce a scalable coding approach for plenoptic images by using a sparse set of EIs and disparities, and the disparities are lossy encoded. Approximated camera geometry is assumed to be known, and EIs can be separated

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

from the plenopitc image. We start by estimating disparities for EIs, and then uniformly retain a sparse set of EIs. Based on the sparse set and disparities, a full plenoptic image is reconstructed by using prediction with interpolation and, for those unpredictable areas, with inpainting. The reconstructed plenoptic image is utilized to predict the original full image by using a modified HEVC encoder. The proposed scheme has a three-layer structure. From the first to the second layer, spatial resolution scalability is provided, and from the second to the third, quality scalability is enabled.

The novelties of this paper are as follows: 1) We encode plenoptic images by using a sparse set of EIs and their associated disparities. The proposed scheme is implemented into HEVC. Compared to our previous work [33], the proposed coding scheme utilizes lossy encoded disparities for plenoptic image reconstruction. The lossy coding can reduce the bit rate allocated to the disparities while possibly retaining the visual quality as compared to the lossless coding; 2) The scalability of the proposed coding scheme is theoretically described and empirically analyzed; 3) The quality of reconstructed parts of full plenoptic images is visually inspected and analyzed; 4) The parameter space for the sparse sampling factor is explored to determine the best sampling factor; 5) We evaluate the proposed system with a high quality lossy coding of disparities, i.e. quality of 60 to 70 dB in PSNR.

The overall aim of the work is to improve the compression efficiency for plenoptic contents. The work is limited to the compression for densely sampled focused plenoptic images. The goal is to investigate the rate-distortion performance for the decoded plenoptic images at the third layer and the quality of the plenoptic image reconstruction at the second layer of the scalable structure.

D. Outline

The paper is organized as follows. The focused plenoptic camera is presented in Section 2. We illustrate our previously proposed displacement intra prediction scheme in Section 3 and the proposed scheme in Section 4. Experimental setup and evaluation criteria are presented in Section 5, and Section 6 shows the results and analysis. Section 7 concludes this paper.

II. Focused plenoptic system

As presented in [8], a focused plenoptic camera is typically in the form shown in Fig. 2. The microlens array is placed such that the microlens is focusing on a plane in front of the photosensor. The main lens system brings a 3D scene into focus at the main lens image plane.

1) Capturing: Plenoptic capturing is a sampling of the light field in its four dimensions. The sampling density is related to the camera parameters a and b, and the distance of the objects to the camera in a 3D scene. A more densely sampled plenoptic image is referred to as more adjacent EIs capture the angular information for a spatially located point in the scene. As a result, adjacent EIs have a higher correlation. In Fig. 2, putting the main lens image plane farther away from the microlens array, i.e., a larger a, increases the sampling density. This can be shown by using a simple ray tracing, as

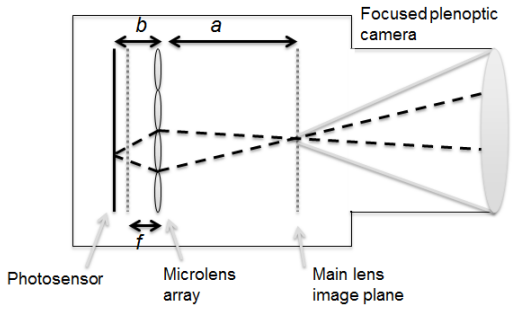


Fig. 2: Focused plenoptic camera [8].

more lenses capture the same 3D point when a is increased. In addition, given a fixed a and b, it can also be shown with the ray tracing that moving the object of the scene farther away from the camera increases the sampling density. As an example, parts of plenoptic images from Plane and Toy [34] with different sampling density are shown in Fig. 3.

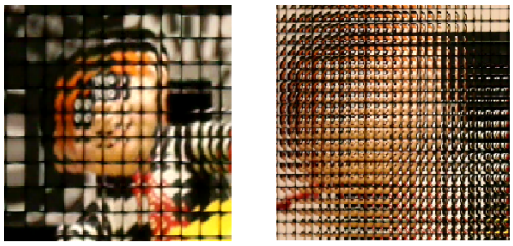


Fig. 3: Focused plenoptic image: sparsely sampled (left) and densely sampled (right) [34].

2) Rendering: Views with different perspectives can be rendered from a plenoptic image. A view is rendered by combining patches from EIs [8]. Fig. 4 describes such a rendering process, where a view is rendered by combining patches from each EI $I_{E(x,y)}(r,t)$ in the captured image $C(x,y,r,t)$. $x \in [1,N]$, $y \in [1,M]$, $r \in [1,N_t]$, and $t \in [1,M_t]$, where N, M, N_t, and M_t are the size of each dimension, e.g., in Fig. 4, $N = M = 4$, and N_t and M_t are the resolution of an EI in each dimension. However, by using a fixed patch size, artifacts will likely appear on parts of a rendered view, because the patch size is dependent on the depth of the scene [8]. Since the depth can be translated into disparities between EIs, it is feasible to perform a depth dependent rendering by using estimated disparities from EIs. With the disparity information, all patches of various sizes are magnified and combined to form a rendered view.

Image refocusing is to integrate and average the angular information of a spatial point in the 3D scene. In the operation with respect to EIs, they are overlapped with each other by using an assigned disparity as shown in Fig. 5, and the color intensity for the overlapped pixels is averaged with the number of overlapping. This operation will bring into focus the objects in a depth plane that corresponds to the assigned disparity.

III. Displacement intra prediction

For the paper to be self-contained and for a better clarity of the proposed scheme, our previously proposed displacement

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

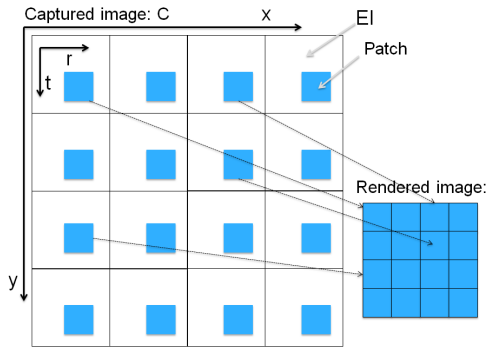


Fig. 4: Captured plenoptic image and all-in-focus rendered image by using a constant patch size.

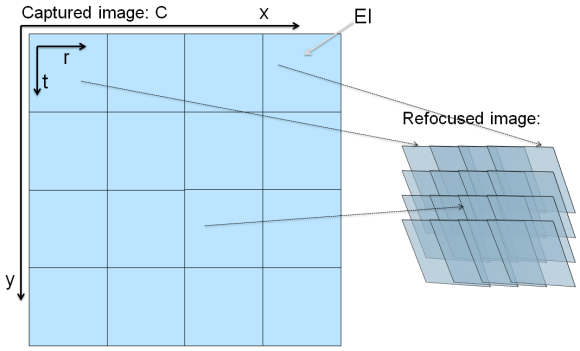


Fig. 5: Captured plenoptic image and refocused image [8] (the EI overlapping is shown in a rotated 3D manner). The overlapped pixels of EIs are averaged for the rendered image.

intra prediction [29] is briefly illustrated here. The displacement intra prediction scheme can perform a bi-directional prediction in spatial domain for coding of plenoptic images and is referred to as B-coder.

As shown in Fig. 6(a), two parts of the image are assumed as two reference pictures available in the reference picture list L_0 and L_1. A current coding block is predicted from the best matching reference block, which can be the best matching block in list L_0, the best in list L_1, or $\frac{(P_0+P_1)}{2}$. P_0 and P_1 are two blocks obtained from L_0 and L_1, respectively. The best is measured in terms of minimum rate-distortion. In addition, the original HEVC directional intra prediction is also evaluated in the Rate-Distortion Optimization (RDO) process. As a result, the best prediction mode is selected for coding the current block.

The displacement intra B-coder has been integrated into HEVC framework with a maximum of two hypotheses. The scheme efficiently reduces inter-EIs redundancy without knowing lens geometry. A detailed description of the original HEVC intra and the displacement intra can be referred to [30] and [29], respectively.

IV. Proposed Method

The proposed scheme is to provide the coding with three-layer scalability and enable an efficient coding. At the first layer, a sparse sampled set of EIs is retained along with

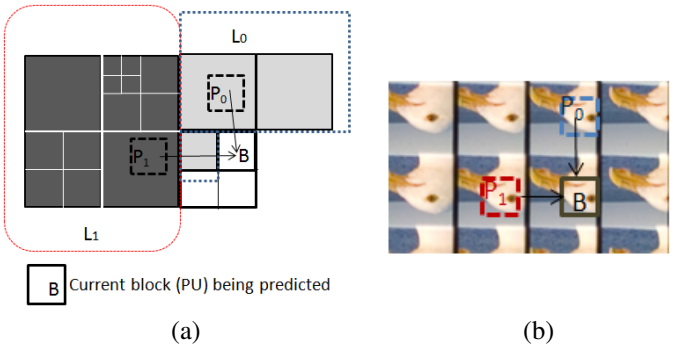


Fig. 6: Bi-prediction within an image. (a) Two parts in color light gray and dark gray are assumed as two reference pictures and available in the reference list L_0 and L_1; (b) an illustration of the prediction on a light field image.

the estimated disparities. A full plenoptic image can then be reconstructed at the second layer. The original image is encoded by using the reconstructed image as a prediction reference at the third layer.

Fig. 7 and Fig. 8 present the overview diagrams of the proposed coding scheme, the details of each block in the diagram are explained in the following subsections.

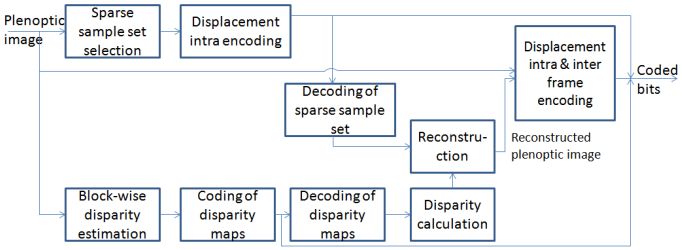


Fig. 7: The proposed plenoptic image encoding system.

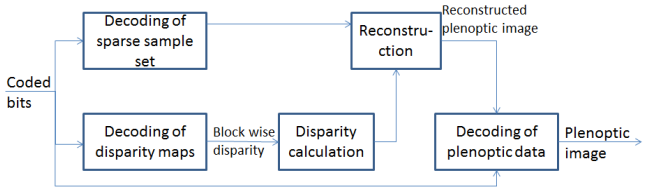


Fig. 8: The proposed plenoptic image decoding system.

A. Encoding

Sparse sample set selection: A plenoptic image is sampled into a sparse plenoptic image set as illustrated in Fig. 9. Assume (x, y) are the coordinates of an EI $I_{E(x,y)}(r, t)$ within the plenoptic image $C(x, y, r, t)$ in Fig. 9. A sparsely sampled image $C_s(x_s, y_s, r, t)$ is obtained with a sampling factor s such that $x_s \in [1, N/s]$, $y_s \in [1, M/s]$, and $C_s(x_s, y_s, r, t) = C(x_s \cdot s, y_s \cdot s, r, t)$. The sampling process on a captured plenoptic image is illustrated in Fig. 10.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

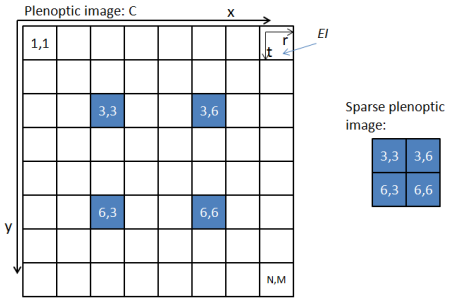


Fig. 9: A 8 by 8 plenoptic image sparsely sampled by a factor of $s = 3$.

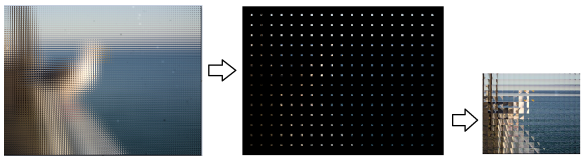


Fig. 10: An example of the sampling.

Displacement intra encoding: The sparsely sampled image can be encoded by state-of-the-art image encoders. In this work, we employ the displacement intra B-coder [29], mentioned in the previous section, for the encoding.

Decoding of sparse sample set: The coded image is decoded. This decoded sparse sample set of images are used for a later reconstruction.

Block-wise disparity estimation: The disparity estimation is performed on the original plenoptic images. As an entire EI is considered as a block, the disparity between the current EI and the EI at its right side is estimated as the horizontal disparity, and the current EI and the EI at its bottom side as the vertical disparity. The estimation is performed by minimizing the Mean Square Error (MSE) between the two neighboring EIs, e.g., for estimating the horizontal block-wise disparity $D_h(x, y)$, the disparity map D_h is obtained by

$$\begin{aligned} &\underset{D_h(x,y)}{argmin}(\underset{r\in[1,N_t],t\in[1,M_t]}{\text{MSE}}(I_{E(x,y)}(r + D_h(x, y), t), \\ &I_{E(x+1,y)}(r, t)), \end{aligned} \tag{1}$$

where $x \in [1, N - 1]$ and $y \in [1, M]$. For measuring the vertical block-wise disparity map D_v, it is by

$$\begin{aligned} &\underset{D_v(x,y)}{argmin}(\underset{r\in[1,N_t],t\in[1,M_t]}{\text{MSE}}(I_{E(x,y)}(r, t + D_v(x, y)), \\ &I_{E(x,y+1)}(r, t)), \end{aligned} \tag{2}$$

where $x \in [1, N]$ and $y \in [1, M-1]$. The pixels shifted outside of the EI are discarded without taking into calculation.

The results from the estimation are two disparity maps for the horizontal and the vertical directions. Therefore, two disparity maps, D_h of resolution 7 by 8 and D_v of 8 by 7, are produced for the plenoptic image illustrated in Fig. 9.

Coding of disparity maps: The two block-wise disparity maps are encoded by using HEVC inter-frame prediction, i.e., one disparity map is encoded as intra-coded frame, from which another is predicted by using HEVC inter-frame prediction. These maps are encoded in high quality to ensure an accurate plenoptic reconstruction.

Decoding of disparity maps: The coded block-wise disparity maps are decoded.

Disparity calculation: For a later reconstruction, the disparities between all EIs outside the sparse set to each EI in the sparse set with a range of r must be estimated. We refer these disparities to as sparse-set-centered disparities. An EI in the sparse set is located at each (x_s, y_s) within the plenoptic image illustrated in Fig. 11.

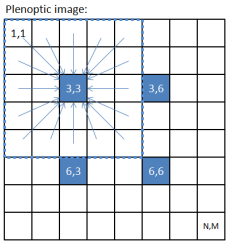


Fig. 11: Disparity calculation from all EIs outside the sparse set in the blue box to an EI centered at $(3, 3)$ in the sparse set with a range $r = 2$.

Because the block-wise disparities have already been acquired, the sparse-set-centered disparities can simply be calculated by an addition horizontally and vertically from the block-wise disparities. It is shown that $D_h(x, y)$ is the block-wise horizontal disparity for the EI at the coordinate (x, y) to its neighbor at the right side, and $D_v(x, y)$ is the block-wise vertical disparity to its neighbor at the bottom side. The horizontal and the vertical sparse-set-centered disparities for the EI at (x, y) to the EI at (x_s, y_s) are calculated by:

$$\begin{aligned} &D_{hs}((x, y), (x_s, y_s)) = \\ &\begin{cases} \sum_{i=x}^{x_s-1} D_h(i, y), & x_s > x \\ \sum_{i=x_s}^{x-1} -D_h(i, y), & x > x_s, \end{cases} \\ &D_{vs}((x, y), (x_s, y_s)) = \\ &\begin{cases} \sum_{i=y}^{y_s-1} D_v(x, i), & y_s > y \\ \sum_{i=y_s}^{y-1} -D_v(x, i), & y > y_s. \end{cases} \end{aligned} \tag{3}$$

The sparse-set-centered disparities, D_{hs} and D_{vs}, provide the disparity vector to reconstruct the unknown EIs from the known EIs in the sparse set.

Reconstruction: A full plenoptic image is reconstructed from the decoded sparse plenoptic image set. As Fig. 12 shows, the EIs from the decoded sparse image set are placed into their original coordinates within the full plenoptic image. Based on the decoded and calculated sparse-set-centered disparities, the unknown EIs are obtained from the known EIs by a disparity shift. If multiple known EIs are available for an unknown EI within the range r, they are averaged. As an example, in Fig. 12, the EI at coordinate $(1, 1)$ is extrapolated from EI at $(3, 3)$, and EI at $(5, 4)$ is interpolated from four known EIs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

After the interpolation or extrapolation process, there are still areas missing in each of the reconstructed EIs. Inpainting approaches in general can be used to fill the missing areas. In our work, a fluid dynamic inpainting approach [35] is employed to inpaint the missing areas. This inpainting method assumes the isophotes in the image as flows. The missing data is filled by solving the Navier-stokes equation. An example of the reconstruction of an image is illustrated in Fig. 13.

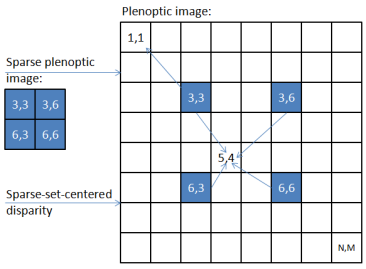


Fig. 12: Reconstruction of plenoptic images with $r = 2$. EI at (1,1) is extrapolated, and EI at (5, 4) is interpolated.

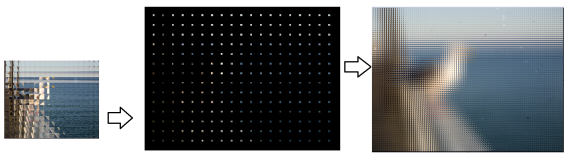


Fig. 13: An example of the reconstruction.

Displacement intra & inter frame encoding: The HEVC encoder is modified for the prediction of plenoptic images in coding. During the initialization of the modified encoder, the reconstructed plenoptic image from the *Reconstruction* process is loaded into the reference picture list in HEVC and available for inter frame prediction. During encoding, both intra prediction and inter prediction are performed, the best coding mode with the RDO for each coding block is chosen for the prediction. Prediction residues are quantized, transformed and entropy encoded as in HEVC. The intra prediction used here is the displacement intra B-coder. Fig. 14 illustrates the prediction for the modified encoder.

The final coded bit streams consist of three components: 1) coding of sparse image sets from *Displacement intra encoding*, 2) coding of disparities from *coding of disparity maps*, and 3) coding of full plenoptic data from *Displacement intra & inter frame encoding*.

B. Decoding

Decoding of sparse sample set: The sparse plenoptic image set is decoded.

Decoding of disparity maps: The coded block-wise disparity maps are decoded.

Disparity calculation and *Reconstruction:* These procedures are identical to the ones in the encoding. The two sparse-set-centered disparity maps are calculated from the decoded block-wise disparity maps as described in the Encoding system. With

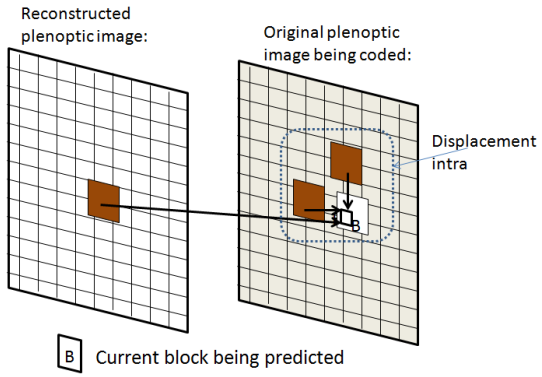


Fig. 14: The prediction process for the modified HEVC encoder.

the sparse image set and the disparity maps, a full plenoptic image is reconstructed.

Decoding of plenoptic data: As an inverse process of *Displacement intra & inter frame encoding*, a plenoptic image is decoded by using the reconstructed plenoptic image as a prediction reference.

C. Scalability

The proposed scheme is scalable and can be considered as having three layers. The first layer is the sparse image set, which is in fact a sparsely sampled plenoptic image. Rendered views can be obtained directly from this image. The amount of angular and spatial information in the image depends on the sampling factor s. A smaller s implies more angular and spatial intensities can be achieved for the rendering. The second layer is the reconstructed full plenoptic image if the disparity maps are available. The reconstruction quality depends on the factor s and how well the disparity estimation, interpolation, and inpainting are performed. The third layer is the residues from the prediction by using the reconstructed plenoptic image. When these residues and their associated information are present, the original plenoptic image can be decoded with a given coded quality in terms of PSNR. It must be clarified here that the scalability from the first layer to the second layer is the resolution/spatial scalability, and that from the second layer to the third layer is the quality/SNR scalability.

This scalability property is beneficial if the network resource is limited, because the image in the first layer is much smaller than the original full image and is sufficient for producing a 2D view if the sampling factor s is appropriate. In addition, in a differentiated network, the disparity maps and the sparse image set can be set to a high priority for transmission. If the data for the third layer are lost during transmission, a full plenpotic image is still possible to be reconstructed in the second layer for rendering.

D. Computational complexity

The computational complexity of HEVC coding has been analyzed empirically in [36]. The complexity of the *Displacement intra encoding* is equivalent to HEVC B frame coding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

with one reference picture in each of the reference picture lists. For the *Reconstruction* process, it involves interpolation and inpainting, the computational complexity depends on how fast they can perform. An interpolation is an operation of averaging multiple pixels (with a maximum of four in our experiment). For the Navier-stokes inpainting, it is shown that large missing areas in an image were inpainted in a magnitude of seconds by using a standard PC [35]. The inpainting is a parameter that can be changed in the scheme, and a detailed analysis of the complexity can be found in [35]. As to the *Displacement intra & inter frame encoding*, it is equivalent to HEVC B frame coding with two reference pictures in each of the reference picture lists. Consequently, it can be seen that the overall computational complexity of both encoding and decoding of the proposed scheme is higher than using the displacement intra B-coder or the HEVC intra only. However, if only the first layer is needed for transmission and rendering, the coding complexity is lower, which depends on the sampling factor s.

V. Experimental setup and evaluation criteria

Light field images Seagull, Fredo, and Laura [11] were used in the test. These plenoptic images are densely sampled with a different depth distribution and scene. The original images have a resolution of 7240 by 5236, and the EI is of 75 by 75 with a rectangular shape. Because vignetting appears on the EIs at the corner of the plenopitc images, we cut the EIs into size of 64 by 64 from approximately the center position of each EI and attached them together to form a processed plenoptic image. Note that, in general, the size is chosen according to camera settings, and may not be a power of two. The processed version of the image can be seen in Fig. 15 for Seagull. It has a resolution of 6080 by 4544. All images were transformed into YUV 4:2:0 format.

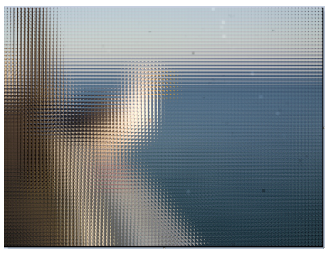


Fig. 15: Processed plenoptic image: Seagull.

HEVC Test Model (HM) reference software version 11 was used for the coding of the plenoptic image and the block-wise disparity maps. The sparse plenoptic image set was encoded by using the displacement intra B-coder. The Quantization Parameters (QP) were 22, 27, 32, and 37. The coding configurations were set as the "All Intra-Main" and the "Low-delay B-Main" setting in JCTVC-L1100 [37] for the HEVC original intra and the displacement intra, respectively. The block-wise disparity maps were encoded by using HEVC inter frame prediction with the Coding Tree Unit (CTU) of size 16, QP 20 and "Low-delay B-Main". We also modified the HEVC encoder and integrated the displacement intra into the proposed scheme for the process *Displacement intra & inter frame encoding*. The current QP used for this process was the same as for the coding of the sparse plenoptic image set, and the coding setting was the "Low-delay B-Main".

The objective quality was assessed on the Y component with PSNR, and the bit rate, bits per pixel (bpp), was calculated from the coded bit stream for all YUV components. The rate distortion curve is plotted for PSNR vs. bpp, and the BD-PSNR [38] was also computed. The results are compared to original HEVC intra, the displacement intra B-coder, and the Block Copying (BC) mode of HEVC range extension version 13. The configurations of the B-coder were defined as "Low-delay B-Main" [37] with a search range of 192 for the displacement vector, and the BC mode as "ALL Intra" [39].

The following aspects of the proposed coding scheme are of our interest: 1) We investigate the performance of the scheme by changing s from 2 to 5 and setting $r = s - 1$ in order to determine the best sampling factor s. 2) With respect to the best sampling factor, the scalability of the scheme is analyzed: 2.a) For the first layer, the sparse image set is a plenoptic image of lower resolution and encoded by the displacement intra *image B-coder*, whose coding efficiency has been investigated in [29]. For the second layer, we also compute the PSNR of the reconstructed image vs. the bit rate of the sparse sample set plus the disparity maps. This is to evaluate the objective reconstruction quality. However, the second layer involves pixels displacement, interpolation, and inpainting processes. Therefore, the reconstructed plenoptic image and its corresponding rendered image are partly shown for a visual inspection. The rendered image is obtained by using the all-in-focus rendering approach discussed in Section II. The patches are taken from the center of each EI with a fixed size of 8, which allows an artifact free rendering for the presented parts.

VI. Results and analysis

The results of the parameter space of sampling factors and the scalability of the scheme are discussed in the following subsections.

A. Sampling factors

The BD-PSNR/rate in Table I shows the largest bit rate reduction compared to HEVC intra was achieved with the sampling parameter $s = 2$ for the proposed scheme. The bit rate reductions compared to HEVC intra are 64.82, 60.90, and 48.87 percent for Seagull, Fredo and Laura, respectively. The performance of the proposed scheme declines with the increase of s. This indicates that an accurate reconstruction of plenoptic images and a precise prediction of the coding of full plenoptic data are essential to improve the coding performance. In addition, Table I illustrates that bit rate reductions of 2.89, 3.05, and 2.29 percent were achieved for Seagull, Fredo and Laura, respectively, compared to the displacement intra. This shows that the majority of the bit rate reduction is achieved by using the displacement intra prediction, which utilizes the reference blocks from the spatial domain. It is further shown

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

that the proposed scheme surpasses HEVC BC mode with over 20 percent bit rate reduction for all the tested images. Although the proposed scheme only achieved an improvement of around 3 percent in bit rate saving compared to the displacement intra, it provides a scalable coding structure for the coding, transmission and rendering, and the results from each layer of the structure will be discussed in Section VI-B.

TABLE I: BD-PSNR/rate: compared to HEVC intra

Image	Coding methods	BD-PSNR (dB)	BD-rate (%)
Seagull	Proposed (s=2)	+4.68	-64.82
	Proposed (s=3)	+4.69	-62.86
	Proposed (s=4)	+4.47	-62.62
	Proposed (s=5)	+4.42	-62.14
	Displacement intra	+4.41	-61.93
	HEVC BC mode	+2.30	-36.36
Fredo	Proposed (s=2)	+5.39	-60.90
	Proposed (s=3)	+5.31	-60.56
	Proposed (s=4)	+5.29	-60.43
	Proposed (s=5)	+5.18	-59.61
	Displacement intra	+4.90	-57.85
	HEVC BC mode	+2.71	-36.17
Laura	Proposed (s=2)	+4.19	-48.87
	Proposed (s=3)	+4.02	-47.21
	Proposed (s=4)	+3.98	-46.90
	Proposed (s=5)	+3.94	-46.54
	Displacement intra	+3.94	-46.58
	HEVC BC mode	+1.87	-24.25

The results in Fig. 16, Fig. 17, and 18 further confirm that the proposed scheme with $s = 2$ reduces coding bit rate significantly more than HEVC intra and HEVC BC mode. It also performs better than the displacement intra B-coder for all tested QPs. The results are consistent for the tested images.

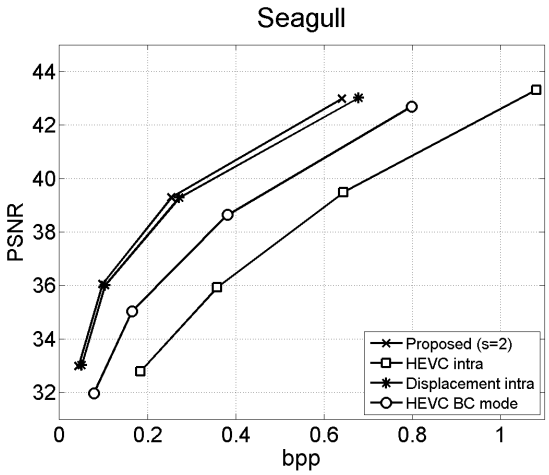


Fig. 16: Rate-distortion curves for Seagull.

Table II, Table III, and Table IV show the coding bit rates for each coding component of the proposed scheme with $s = 2$. It is illustrated that the coding of full plenoptic data contributes to most of the coded bit stream, especially at the higher bit rates, while the disparity maps add least overhead to the bit stream. This also suggests that for an overall improvement of the coding scheme, it is important to reduce the bit rate for the coding of full plenoptic data.

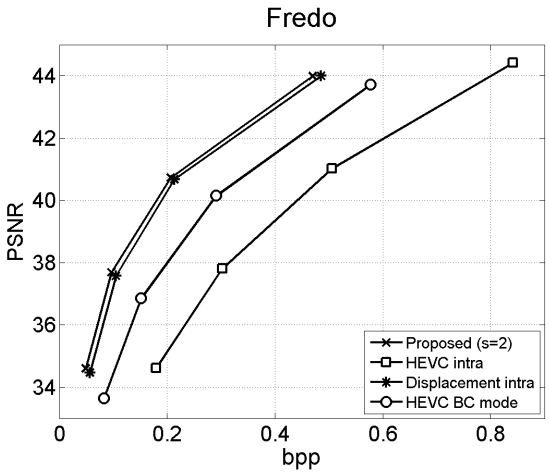


Fig. 17: Rate-distortion curves for Fredo.

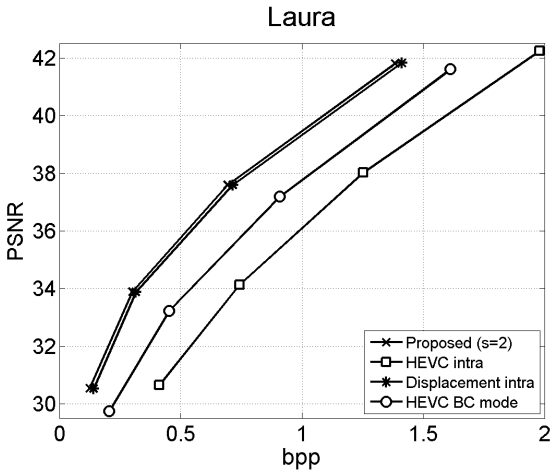


Fig. 18: Rate-distortion curves for Laura.

B. Scalability

The first layer:

Because the sparse image set in the case of $s = 2$ has only half of the resolution of the original image in each dimension, a direct comparison between the first layer to the second and the third layers is impossible. The sampling process results in a loss of angular and spatial information in the sparse image set in general.

The second layer: Fig. 19, Fig. 20, and Fig. 21 additionally illustrate the objective quality of the reconstructed image obtained from the process *Reconstruction* for $s = 2$. The reconstruction quality is above 30 dB for Seagull and Fredo, and around 29 dB for Laura. The variations in PSNR values are small, in the range of 3 dB, with the QP changed from 22 to 37. As mentioned, because the reconstruction process involves pixel displacements, interpolation, etc., a visual inspection is performed to examine the actual visual quality of the reconstruction.

Fig. 23 and Fig. 24 depict parts of the reconstructed images

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

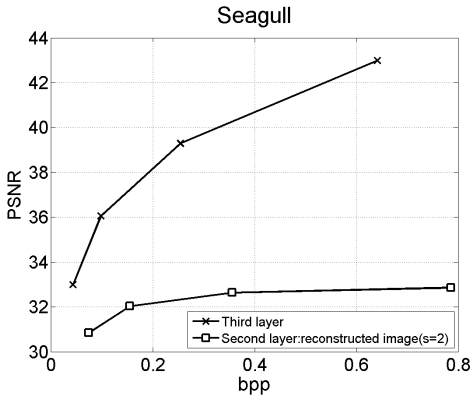


Fig. 19: Rate-distortion curves for Seagull.

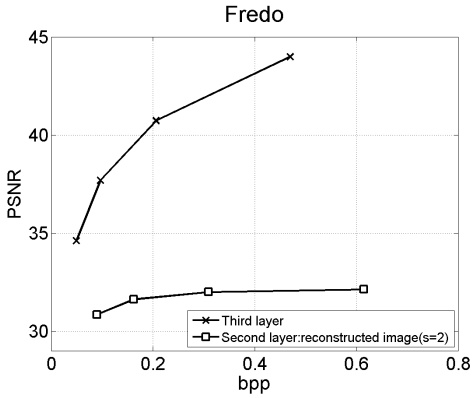


Fig. 20: Rate-distortion curves for Fredo.

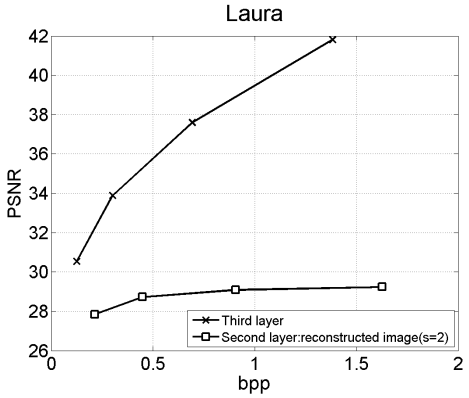


Fig. 21: Rate-distortion curves for Laura.

TABLE II: Seagull: coding bytes per picture for each component of the proposed scheme

QP	Disparities	Sparse image set	Coding of full plenoptic data	Total
22	468	661 116	1 552 115	2 213 699
27	468	298 926	577 672	877 066
32	468	129 829	206 438	336 735
37	468	61 659	85 379	147 506

TABLE III: Fredo: coding bytes per picture for each component of the proposed scheme

QP	Disparities	Sparse image set	Coding of full plenoptic data	Total
22	564	516 795	1 106 050	1 623 409
27	564	260 161	451 398	712 123
32	564	136 140	197 055	333 759
37	564	75 240	94 430	170 234

with $s = 2$ and their corresponding rendered images for Seagull. Fig. 23(a) and Fig. 24(a) are reconstructed from the high quality and the low quality coded sparse image set, respectively. Compared to the same part of the original plenoptic image and rendered image illustrated in Fig. 22, distortions other than the compression artifacts are insignificant, i.e., it is indistinguishable which EI is reconstructed in Fig. 23(a) and Fig. 24(a). However, the reconstruction quality depends on the disparity estimation, the disparity compression, the sampling factor s, the interpolation/extrapolation, and the inpainting.

The third layer: The objective quality of the third layer has been discussed and presented in the beginning of this section and is shown in Fig. 16, Fig. 17, Fig. 18, and Table I. It was illustrated that the proposed scheme with bit rates combined from all the compressed components outperforms the state-of-the-art schemes. Parts of the decoded images and its corresponding rendered views are illustrated in Fig. 25 and Fig. 26 for the high and the low quality, respectively. They are visually slightly better than the corresponding counterparts reconstructed in the second layer shown in Fig. 23 and Fig. 24.

(a) (b)

Fig. 22: (a) Parts of the original plenoptic image; (b) corresponding rendered image.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

TABLE IV: Laura: coding bytes per picture for each component of the proposed scheme

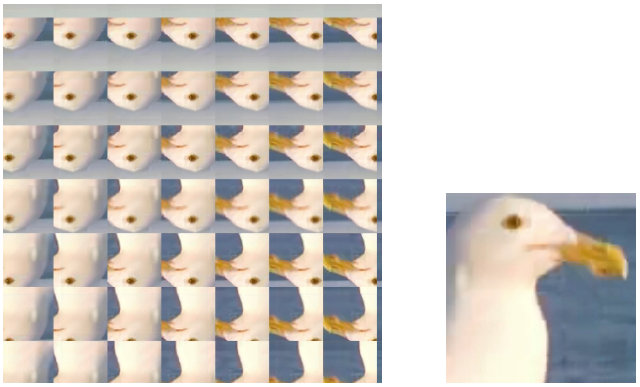
QP	Disparities	Sparse image set	Coding of full plenoptic data	Total
22	331	1 369 558	3 406 527	4 776 416
27	331	762 798	1 630 750	2 393 879
32	331	376 826	552 599	1 039 756
37	331	180 655	251 670	432 656

(a) (b)

Fig. 23: (a) Parts of the reconstructed plenoptic image at the second layer from the coded sparse set with QP 22; (b) corresponding rendered image.

VII. Conclusion

In this paper, we have proposed a scalable coding scheme for densely sampled plenoptic images. The scheme sparsely samples the image and represents a full plenoptic image by its sparse image set and associated disparities, which are encoded accordingly. A full plenoptic image is reconstructed from the decoded sparse set and disparities by using interpolation/extrapolation and inpainting. The reconstructed full image is utilized for a prediction to encode the original plenoptic image with a required PSNR. The proposed scheme is scalable with three layers such that the rendering can be performed with the sparse image set, the reconstructed plenoptic image, and the decoded plenoptic image.

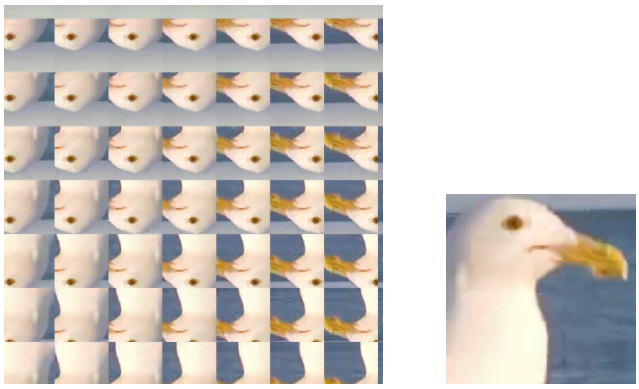


(a) (b)

Fig. 24: (a) Parts of the reconstructed plenoptic image at the second layer from the coded sparse set with QP 37; (b) corresponding rendered image.

(a) (b)

Fig. 25: (a) Parts of the decoded image at the third layer with QP 22; (b) corresponding rendered image.



(a) (b)

Fig. 26: (a) Parts of the decoded image at the third layer with QP 37; (b) corresponding rendered image.

The coding results demonstrated that plenoptic images were compressed efficiently with the proposed scheme. It outperformed HEVC BC mode with more than 2dB quality improvement or by over 20 percent bit rate reduction when measuring by using BD-PSNR/rate. It also surpassed our previously proposed displacement intra B-code by as much as 3 percent bit rate reduction. Visual inspection of the tested image showed that distortions other than compression artifacts were insignificant for the reconstructed image in the second layer of the scalable structure. However, the reconstructed quality depends on several factors, e.g., the sampling factor, interpolation, and inpainting. An accurate reconstruction in the second layer and a precise prediction in the third layer can facilitate an efficient coding of plenoptic images with a required PSNR. Although the improvement over the displacement intra B-coder is small, the scalable feature of the proposed scheme

provides a flexible reconstruction of the plenoptic image from its sparse set, which can enable the coding and transmission to adapt to the limitation of network bandwidth capacity.

VIII. Future work

To optimize the depth estimation, the interpolation and the inpainting process are our future research. In addition, a detailed analysis of the scalability with respect to network transmission and error concealment is also of our future consideration.

Acknowledgment

This work has been supported by grant 20120328 of the Knowledge Foundation, Sweden, by grant 00174636 of the EU European Regional Development Fund, Mellersta Norrland, Sweden. We also want to acknowledge Todor Georgiev for providing the light field images online.

References

[1] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. 1991, pp. 3–20, MIT Press.

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

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 1996, SIGGRAPH '96, pp. 43–54, ACM.

[4] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June 2004, vol. 2, pp. II–294–II–301 Vol.2.

[5] Y. Taguchi, A. Agrawal, S. Ramalingam, and A. Veeraraghavan, "Axial light field for curved mirrors: Reflect your perspective, widen your view," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 499–506.

[6] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007.

[7] R. Ng, "Digital light field photography," *Doctoral thesis, Stanford University*, 2006.

[8] T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021106, Apr. 2010.

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

[10] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," *Proc. SPIE*, vol. 8667, pp. 86671J–86671J–10, 2013.

[11] "Todor Georgiev website," http://tgeorgiev.net/, retrieved: 08, 2013.

[12] X. Dong, D. Qionghan, and X. Wenli, "Data compression of light field using wavelet packet," *ICME '04. 2004 IEEE International Conference*, pp. 1071–1074, 2004.

[13] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 338–343, 2000.

[14] S. Kundu, "Light field compression using homography and 2D warping," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1349–1352, Mar. 2012.

[15] P. Ramanathan and B. Girod, "Rate-Distortion Analysis for Light Field Coding and Streaming," *Signal Processing: Image Communication*, , no. March 2006, pp. 462–275.

[16] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod, "Light field compression using disparity-compensated lifting and shape adaptation.," *IEEE transactions on image processing*, vol. 15, no. 4, pp. 793–806, Apr. 2006.

[17] M. Magnor and B. Girod, "Model-based coding of multiviewpoint imagery.," *SPIE conference Proceedings Visual Communications and Image Processing (VCIP), Perth, Australia*, pp. 14–22, 2000.

[18] X. Zhu, A. Aaron, and B. Girod, "Distributed Compression of Light Fields," *Stanford University, report, online in Citeseer*.

[19] N. Gehrig and P. Dragotti, "Distributed compression of multi-view images using a geometrical coding approach," *Electronic Engineering*, pp. 421–424, 2007.

[20] A. Gelman, "Multiview image compression using a layer-based representation," *Image Processing (ICIP),*, vol. 2, no. 1, pp. 1–4, 2010.

[21] D. Lelescu and F. Bossen, "Representation and coding of light field data," *Graphical Models*, vol. 66, no. 4, pp. 203–225, July 2004.

[22] K. Nishino, Y. Sato, and K. Ikeuchi, "Eigen-texture method: Appearance compression based on 3d model," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999, vol. 1, pp. –624 Vol. 1.

[23] B. A. Vetro, F. Ieee, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H . 264 / MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol. 99, no. 4, 2011.

[24] H. Schwarz, C. Bartnik, and S. Bosse, "3D video coding using advanced prediction, depth modeling, and encoder control methods," *Picture Coding Symposium*, pp. 3–6, 2012.

[25] C. Conti, L. Ducla Soares, and P. Nunes, "Influence of self-similarity on 3D holoscopic video coding performance," *Proceedings of the 18th Brazilian symposium on Multimedia and the web - WebMedia '12*, p. 131, 2012.

[26] C. Conti, P. Nunes, and L. D. Soares, "New HEVC prediction modes for 3D holoscopic video coding," *2012 19th IEEE International Conference on Image Processing*, pp. 1325–1328, Sept. 2012.

[27] C. Conti, P. Nunes, and L. D. Soares, "3D Holoscopic Video Coding Based on HEVC with Improved Spatial and Temporal Prediction," *Telecommunications-conftele, Castelo Branco, Portugal*, pp. 1–4, May 2013.

[28] D. Flynn, J. Sole, and T. Suzuki, "High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 4," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-N1005*, April 2013.

[29] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient Intra Prediction Scheme For Light Field Image Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florance, Italy*, May 2014.

[30] B. Bross, W. Han, J. Ohm, G. Sullivan, Y. Wang, and T. Wiegand, "High efficiency video coding (HEVC) text specification working draft 10," *JCT-VC Document, JCTVC-L1003*, 2013.

[31] J. Pearson, M. Brookes, and P. Dragotti, "Plenoptic layer-based modeling for image based rendering," *Image Processing, IEEE Transactions on*, vol. 22, no. 9, pp. 3405–3419, Sept 2013.

[32] C. Conti, P. Nunes, and L. Soares, "Inter-layer prediction scheme for scalable 3-d holoscopic video coding," *Signal Processing Letters, IEEE*, vol. 20, no. 8, pp. 819–822, Aug 2013.

[33] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Coding of Plenoptic Images by Using a Sparse Set and Disparities," *Proceeding of the IEEE International Conference on Multimedia and Expo (ICME), Torino, Italy*, June 2015.

[34] A. Aggoun, O. A. Fatah, J. J. Fernandez, C. Conti, P. Nunes, and L. D. Soares, "Acquisition, Processing and Coding of 3D Holoscopic Content for Immersive Video Systems," *3DTV-Conference: Vision Beyond Depth (3DTV-CON), Aberdeen, Scotland*, 2013.

[35] M. Bertalmio, A. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I–355–I–362 vol.1.

[36] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "Hevc complexity and implementation analysis," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1685–1696, Dec 2012.

[37] F. Bossen, "Common test conditions and software reference configurations," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-L1100*, 2013.

[38] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33*, 2001.

[39] D. Flynn, K. Sharman, and C. Rosewarne, "Common test conditions and software reference configurations for HEVC range extensions," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-O1006*, Nov. 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2498406, IEEE Transactions on Image Processing

Yun Li received his master of science in computer engineering and his technical licentiate degree in computer and system science from Mid Sweden University (MIUN), Sweden, in 2008 and 2013, respectively. He has been a full time researcher and Ph.D. student in the Department of Information and Communication Systems at MIUN since 2011. His research interest includes video coding, transmission, rendering and computer vision.

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 KTH, Stockholm, Sweden, in 1998, and the Ph.D. degree in modeling of nonlinear systems from EPFL, Lausanne, Switzerland, in 2001. He worked as an Electrical Engineer at ABB, Sweden, from 19931994, was a fellow at CERN from 19941996, and a Ph.D.-student at EPFL, Lausanne, Switzerland during 19972001. In 2001, he joined Mid Sweden University and was appointed Associate Professor and Full Professor in Signal Processing in 2008 and 2013, respectively. He is the head of the subject Computer and System Sciences at 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, as well as system modelling and identification.

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 worked in the video compression and distribution industry 1997-2000. He again joined Mid Sweden University as a junior lecturer 2000-2004 where he taught courses in telecommunication, signals- and systems, and signal- and image processing. Since 2010 he is employed as a researcher at Mid Sweden University where his research interest includes plenoptic image capture, processing, and compression; plenoptic system modelling; and depth map capture and processing.

Ulf Jennehag received the M.Sc. degree in electrical engineering and telecommunication from Mid Sweden University, Sweden, in 2000, the Licentiate of Technology degree in Teleinformatics from Royal Institute of Technology (KTH), Sweden, in 2005, and the Ph.D. degree in computer and system sciences form Mid Sweden University, in 2008. He worked as a post doc at Audio Department in Fraunhofer IIS, Erlangen, Germany, during 2008-2009. In 2009 he joined Mid Sweden University and are as of 2010 employed as an Assistant Professor. His current research interests are within multimedia streaming, video coding, and Internet of Things.