

Plenoptic Image Coding using Macropixel-based Intra Prediction

Xin Jin, *Senior Member, IEEE*, Haixu Han and Qionghai Dai, *Senior Member, IEEE*

Abstract— The plenoptic image in a super high resolution is composed of a number of macropixels recording both spatial and angular light radiance. Based on the analysis of spatial correlations of macropixel structure, this paper proposes a macropixel-based intra prediction method for plenoptic image coding. After applying an invertible image reshaping method to the plenoptic image, the macropixel structures are aligned with the coding unit grids of a block-based video coding standard. The reshaped and regularized image is compressed by the video encoder comprising the proposed macropixel-based intra prediction, which includes three modes: multi-block weighted prediction mode (MWP), co-located single-block prediction mode (CSP), and boundary matching based prediction mode (BMP). In the MWP mode and BMP mode, the predictions are generated by minimizing spatial Euclidean distance and boundary error among the reference samples, respectively, which can fully exploit spatial correlations among the pixels beneath the neighboring microlens. The proposed approach outperforms HEVC by an average of 47.0% bitrate reduction. Compared with other state-of-the-art methods, like pseudo-video based on tiling and arrangement method (PVTA), intra block copy (IBC) mode, and locally linear embedding (LLE) based prediction, it can also achieve 45.0%, 27.7% and 22.7% bitrate savings on average, respectively.

Index Terms— Plenoptic image coding, macropixel-based intra prediction, light field coding, HEVC/H.265

I. INTRODUCTION

Light field cameras have attracted great attention in recent years with the investigation and commercialization in hand-held light field cameras, so-called plenoptic cameras, like *Lytro* [1] and *Raytrix* [2]. Contrary to conventional cameras, plenoptic cameras based on microlens arrays record not only the spatial light intensities but also the light propagation directions using a single exposure for the three-dimensional (3D) scene. Due to the unique light gathering capability, plenoptic imaging has become a prospective imaging approach in providing functionalities like refocusing, changing viewing perspectives and retrieving depth information. After post-processing and calibration, the captured plenoptic images can be applied to a variety of applications such as fatigue-free 3D visualization

[3][4], 3D television [5], saliency detection [6] and object recognition [7] for an improved quality or a lower system complexity, which have attracted great interest both from academy and industry.

Since a plenoptic image is a super-high-definition image in which each pixel preserves the fidelity of spatial and angular information with a full color space, its intensity distribution is quite different from the image captured by conventional cameras, which desires efficient compression methods to reduce the spatial redundancy and maintain the fidelity simultaneously.

Recently, plenoptic image compression has attracted a wide attention from both industry and academy. JPEG initiated a new project called JPEG Pleno [8][9] to standardize the next generation image coding methods for the plenoptic images. MPEG included light field coding into MPEG-I Visual standardization project [10][11]. ICME 2016 [12] and ICIP 2017 [13] also organized competitions to explore efficient compression methods.

The existing plenoptic image compression approaches can be mainly classified into two categories: approaches that compress the plenoptic image directly and approaches that compress the pseudo-video generated from the plenoptic image. Approaches that compress the plenoptic image directly encode the plenoptic image or the rearranged plenoptic images via spatial predictive schemes, disparity-based predictive schemes and transform-based coding schemes to exploit the spatial redundancy among the macropixels. For the spatial predictive schemes, coding tools like displacement intra prediction [14-16] and self-similarity compensated prediction [17-18] were proposed, in which the coding unit was predicted by a matched block in the reconstructed region. In [19], a light field image coding solution based on bi-prediction self-similarity estimation and compensation was proposed, where two predictors are jointly estimated by the locally optimal rate-constrained algorithm. Although an average of 51.5% bitrate reduction can be achieved relative to high efficiency video coding standard (HEVC)[20], huge computational overhead is introduced, which shows quite limited performance in considering compression efficiency and computational complexity jointly. Intra block copy (IBC) mode [21] adopted in HEVC screen content coding extension, which is regarded as motion compensation within the current picture, can also efficiently improve the compression performance. In [22], a predictive mode based on the locally linear embedding (LLE), which estimates the coding unit using a linear combination of k -nearest

Manuscript received on Nov. 7th, 2017. This work was supported in part by NSFC under Grant 61771275 and Shenzhen Project under Grant JCYJ20170307153135771, China.

X. Jin and Haixu Han are with Shenzhen Key Lab. of Broadband Network & Multimedia, Graduate School at Shenzhen, Tsinghua Univ., China. (email: jin.xin@sz.tsinghua.edu.cn, haixu.han@qq.com, phone & Fax: +86 755 2603 6770)

Q. Dai is with TNLIST and Department of Automation, Tsinghua Univ., China (email: qhdai@tsinghua.edu.cn, phone & Fax: +86 10 62773606)

neighboring patches, was proposed. It was further extended by combining with self-similarity prediction method to improve the coding efficiency in [23]. And in [24], an adaptive block differential prediction tool in JPEG2000 was designed by an entropy analysis process aiming to reduce the inherent redundancy of plenoptic images. Besides, the plenoptic data can be arranged to be a new image by lenslet array slicing [25] or placing different angular views side-by-side [26]. They can be compressed by conventional codec or spatial-predictive methods. In the disparity-based predictive schemes, the disparity in the plenoptic image, which can be regarded as the spatial shift between the projects of a 3D point in two macropixels, is exploited to perform disparity compensation for prediction units (PUs) [27][28]. However, the spatial predictive schemes and the disparity-based predictive schemes introduced overhead bits consumed by motion/disparity vectors and did not fully exploit the optical imaging correlations among macropixels yet. Transform-based schemes use the discrete cosine transform (DCT) [29] or discrete wavelet transform (DWT) [30]. Their reported compression efficiency is higher than JPEG, while lower than HEVC intra coding. R. Monteiro *et al.* [31] proposed a two-stage block-wise high order prediction model, which predicts each block by applying a geometric transformation. It outperforms HEVC, while the compression efficiency is desired to be improved under the constraint of the computational complexity.

The approaches in the second category compress the pseudo-video generated from the plenoptic image. A low resolution pseudo video sequence consisting of extracted micro-images or subaperture images [32] is generated and compressed using the temporal prediction tools in the existing video coding standards, such as JPEG2000 [33], H.264 [34] or HEVC [20]. Since the subaperture images generally correspond to the images from different viewpoints, some methods in this category focused on the scanning topologies. S. Zhao *et al.* [35] proposed horizontal zigzag/U-shape scan and F. Dai *et al.* [36] proposed line mapping and rotational mapping to reorder the subaperture images in the pseudo video sequence for a higher coding performance. Due to the sub-views correlation, S. Zhao *et al.* [37] utilized the selected compressed views to approximate a certain view by designing linear approximation prior. While, in [38], the selected views were coded and other views were reconstructed by sparse prediction. Also, in order to exploit the geometric relation among subaperture images, some homography transformation based methods [39-41] were proposed. C. Perra *et al.* [42-44] partitioned the plenoptic image into tiles and treated each tile as a frame of a pseudo-temporal sequence for encoding using JPEG2000 and HEVC. In [45], a flexible light field compression architecture was proposed by dividing the subaperture images into different groups. Besides, the data formats and quality assessments for light field compression were investigated by the same authors in [26][46]. Exploiting the perspective information from the plenoptic images, multi-view sequences can be generated and be compressed using multiview video coding (MVC) [47-49] or 2-D hierarchical coding structure [50] to exploit the temporal and interview correlations among adjacent subaperture images. Due to introducing the temporal/interview coding tools, these schemes can provide higher efficiency compared with the methods in the first category. However, as an overhead, the

approaches suffered from a huge increase in the computational complexity, which was not applicable to real applications, especially those preferring lower latency.

Considering the compression efficiency and the encoding complexity jointly, we proposed a macropixel-based intra prediction mode called boundary matching based prediction mode (BMP) [52] and two modes called multi-block weighted prediction mode (MWP) and co-located single-block prediction mode (CSP) [53], respectively, in our previous works. In this paper, targeting a better trade-off between the compression efficiency and the computational complexity, we extended the works by: designing a complete adaptive coding solution for luminance and chrominance components with rate-distortion-optimization flow, mode signaling and entropy coding to benefit the compression efficiency from the three modes simultaneously; providing analyses on cross correlations and imaging system architecture to discover the fundamental reason for quality improvement; discussing the generality of the proposed approach for different microlens arrangements; and conducting comprehensive experiments and analyses to demonstrate the attractive performance in improving the compression efficiency together with a good tradeoff in computational complexity. To further improve the compression efficiency, the light-field-lossless image reshaping method proposed by us in [51] is also applied to align the macropixel structures in the plenoptic image with the block coding unit grids in the hybrid coding architecture.

The remainder of this paper is organized as follows. The characteristics of the plenoptic image and the image reshaping method are introduced in Section II. The framework and the details of the proposed macropixel-based intra prediction are described in Section III. The compression efficiency and computational complexity are evaluated in Section IV to demonstrate the effectiveness of the proposed algorithm. Section V concludes the paper.

II. PLENOPTIC IMAGE CHARACTERISTICS AND IMAGE RESHAPING

In this section, the plenoptic image structure is described with discussions on its characteristics. Then, the image reshaping method proposed by us in [51] is described with the analysis of the correlation variation among adjacent PUs.

A. Plenoptic image structure and characteristics

The distinct feature of a standard plenoptic camera [54] relative to the conventional imaging system is inserting a microlens array into the light path between the main lens and the image sensor as shown in Fig. 1. The light beams coming from the object with different incident angles go through the main lens and converge at the microlens. Afterwards, they are diverged and captured by the image sensor as a group of pixels, called a macropixel. Hence, using the imaging architecture of *Lytro Illum* as an instance in Fig. 1, unlike traditional images, the output of the sensor is composed of a number of macropixels in hexagonal shape, which record both spatial and angular light information. Hexagonal microlens array has been demonstrated to be the one with the largest fill-factor among the existing lens shapes and arrangements [56][57]. After light field decoding, including demosaicing, devignetting, rotation and scaling [55], the lenslet image is generated to be applied in many computer

vision applications for refocusing, depth estimation and multiview extraction.

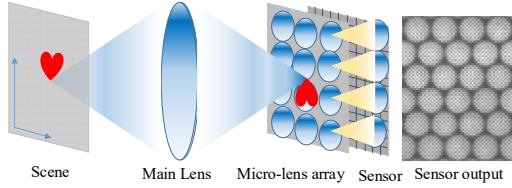


Fig. 1. Image formation process of a standard plenoptic camera using *Lytro Illum* as an instance.

It is observed that if partitioning the lenslet image by the coding unit grid, like the 16×16 block used in H.264/HEVC, not every macropixel can be entirely contained by one coding unit. Using the lenslet image captured by *Lytro Illum* as an instance in Fig. 2, the red grid shown in it, corresponding to 16×16 coding unit grid, partitions several incomplete macropixels into one block, which results in low correlations among the adjacent blocks and complex textures inside a block.

In order to verify this, an analysis of the cross-correlation between the current 16×16 block and its adjacent four 16×16 blocks located left to, left-above, above and right-above the current block is performed. The average of cross correlations is retrieved using images “Fountain_&_Vincent” and “Friends” [58] with quite different spatial features, shown in Fig. 11(d) and (e), as instances. As shown in Fig. 3, except the blocks on the left, the correlations between the current block and other adjacent blocks are limited where the average are only 0.25/0.18, respectively. It is obvious that such low correlations among the adjacent blocks will directly affect the efficiency of spatial coding tools in video coding standards. Thus, a light-field-lossless invertible image reshaping method proposed by us in [51] is applied to the work of this paper.

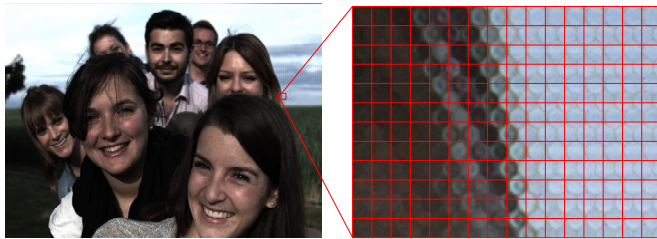


Fig. 2. Partitioning the lenslet image “Friends” [58] by 16×16 coding unit grid.

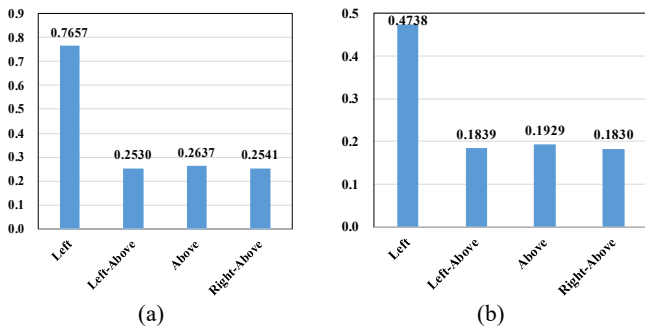


Fig. 3. Cross correlation among the adjacent 16×16 blocks: (a) Fountain_&_Vincent; (b) Friends.

B. Invertible image reshaping method

The invertible image reshaping method proposed by us in [51] is applied to the preprocessed lenslet image that generated according to [9]. It realigns the macropixels in the preprocessed lenslet image to the coding unit grids for a block-based coding standard friendly structure. The image reshaping method includes two steps: macropixel alignment and adaptive interpolation. Macropixel alignment is to reshape and regularize the lenslet image to guarantee that the centers of macropixels are aligned in the vertical/horizontal direction and each macropixel can be fully contained by a non-overlapped 16×16 coding unit. For the entire macropixel alignment process, it just needs some parameters like vertical spacing between the neighboring macropixel rows, vertical/horizontal spatial offset of incomplete macropixels on the boundary, etc., which can be derived from light field decoding. A vertical coordinate transformation is applied to the lenslet image to separate the macropixel rows vertically, i.e. converting the macropixel structure in Fig. 4(a) to be that in Fig. 4(b). The pixels which are separated from the original macropixels (the S pixels marked in magenta and green in Fig. 4(b)) are moved back to their original locations, as shown in Fig. 4(c), followed by which a horizontal coordinate transformation is applied to the odd or the even macropixel rows to align the macropixels vertically to generate the structure like that shown in Fig. 4(d). After that, a boundary processing is applied to the incomplete macropixels along the image boundaries by checking whether the pixels in the incomplete macropixels are valid in rendering the 2D perspective views in the subaperture image stack or not [9]. If they are valid, pixel padding will be performed. Otherwise, the pixels will be discarded. The boundary processing also guarantees the width and the height of the reshaped lenslet image is the multiple of 8.

After macropixel alignment, the pixels at the four corners and the bottom/right boundary of each coding block, colored in white in Fig. 4(d), are generated without exact intensity values. To maximize the continuity of adjacent macropixels, adaptive interpolation is applied to fill the intensity values of those pixels based on the relative distance to the nearest neighborhoods [51].

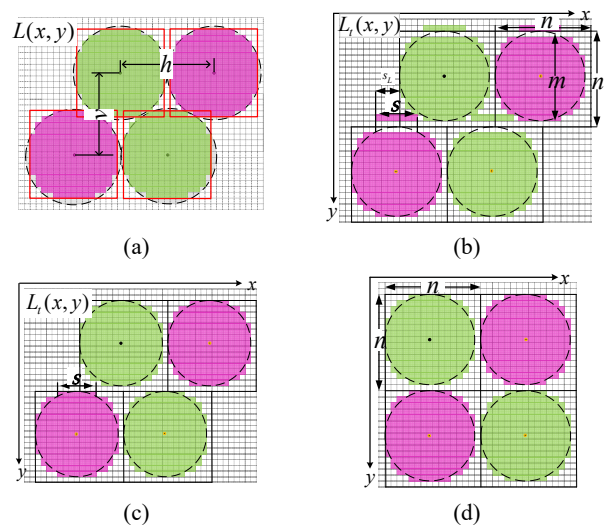


Fig. 4. Macropixel structures: (a) in a lenslet image; (b) after vertical coordinate transformation; (c) after moving S pixels back; (d) after macropixel alignment.

Finally, the reshaped lenslet image is generated with macropixel structures aligned to coding unit grids. Fig. 5 shows the reshaped portion of “Friends”, corresponding to the portion in Fig. 2, with coding unit grids aligned macropixel structures. Based on the transmitted parameters in Table I [51], inverse reshaping process can reconstruct the lenslet image structure, as shown in Fig. 4(a), exactly without sacrificing the quality of the generated light field. After image reshaping, the size of the regularized lenslet image will be 14.2% larger than the original, while, its macropixel structure is compatible with the coding unit grids in the block-based video coding standards.

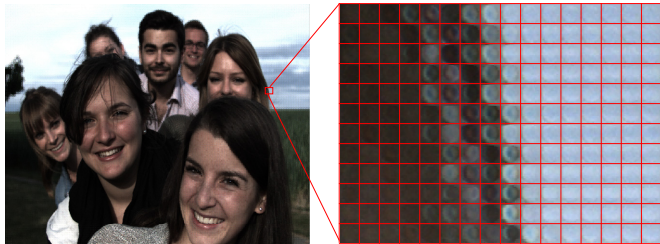


Fig. 5. The reshaped image and enlarged portion partitioned by 16×16 coding unit grid.

TABLE I. Parameters Transmitted for Inverse Reshaping [51]

Param.	Semantics	Range	Bits
m	Radius of each macropixel	[2,32]	5
v	Vertical spacing between the neighboring two macropixel rows	[2,32]	5
y_0	The maximum vertical pixel number in the incomplete macropixel row at the image upper boundary	[1,31]	5
S	The number of pixels to be moved back to each macropixel	[1,31]	5
S_L	the left pixel offset of the to-be-moved pixels relative to the boundary of the coding unit grid	[1,16]	4
$x_{0,o}$	Incomplete pixels counting from the left boundary of the odd macropixel rows	[2,32]	5
$x_{0,e}$	Incomplete pixels counting from the left boundary of the even macropixel rows	[2,32]	5
P_h	The number of pixel columns padded to the right image boundary	[0,7]	3
P_v	The number of pixel rows padded to the bottom image boundary	[0,7]	3
f_{HP}	Flag representing whether the pixels on the right boundary of the non-transformed macropixel row are valid for light field generation	1, valid 0, else	1
f_{VP}	Flag representing whether the pixels on the bottom boundary of image are valid for light field generation	1, valid 0, else	1

After applying image reshaping to the lenslet image, the inherent correlations among the adjacent macropixels can be easily exploited by the block-based video coding standard. Similar to the correlation analysis performed in Fig. 3, the cross correlations and the ratio of strong correlations are compared for the lenslet image and the reshaped lenslet image in Fig. 6 using “Fountain_&_Vincent” and “Friends” as well. It can be found that the average of cross-correlation among the adjacent 16×16 blocks can be greatly improved by image reshaping, as shown in Fig. 6 (a) and (b), especially for the block left-above, above, and right above the current block. Also, the proportion of the blocks with cross-correlation higher than 0.9 in the reshaped image is also much larger than that in the lenslet image, as shown in Fig. 6(c) and (d). It indicates that image reshaping improves the spatial correlations among the coding units, which will further

benefit the spatial coding tools in reducing the spatial redundancy. Its effectiveness in improving the compression efficiency will be further demonstrated in Section IV.

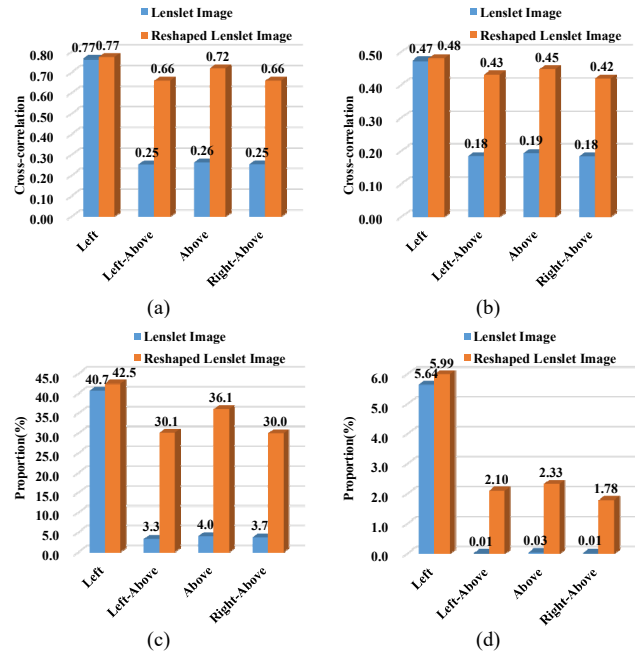


Fig. 6. The average of cross-correlation of: (a) Fountain_&_Vincent; and (b) Friends. The proportion of blocks with cross-correlation higher than 0.9 of: (c) Fountain_&_Vincent; and (d) Friends.

III. PROPOSED MACROPIXEL-BASED INTRA PREDICTION

The overall block-diagram of the proposed lenslet image compression system and the encoding architecture with the proposed macropixel-based Intra prediction based on HEVC are depicted in Fig. 7(a) and (b), respectively. The proposed lenslet image compression system mainly consists of three modules (the blocks in gray in Fig. 7(a)):

- 1) *Image Reshaping*, as described in Section II B, is first applied to the preprocessed lenslet image $L(x, y)$ [9] to reshape and regularize macropixels structures to be aligned with coding unit grids in the block-based hybrid encoder, as shown in Fig. 4;
- 2) The reshaped lenslet image is fed to the encoder and decoder, denoted by *Codec*, comprising the proposed macropixel-based intra prediction for compression;
- 3) *Inverse Reshaping* is applied to the decompressed lenslet image according to the transmitted parameters in Table I to recover the macropixel structure as that in $L(x, y)$ for output.

Fig. 7(b) depicts the proposed encoder architecture, which is based on HEVC and comprises the proposed macropixel-based intra prediction module. In HEVC, a CU is associated with the partitioned PUs and the transform units (TUs). Each PU rooting at the CU level is designed to carry information related to the prediction mode such as the type, sizes, and partition patterns for better prediction [60]. Considering the proposed compression method targets encoding the lenslet image by exploiting the spatial correlations among the macropixels, the proposed macropixel-based Intra prediction modes are added into the rate distortion optimization (RDO) process of the encoder as additional candidate intra prediction modes for each

PU. The proposed method including multi-block weighted prediction (MWP) mode, co-located single-block prediction (CSP) mode, and boundary matching based prediction (BMP) mode will be introduced in the following.

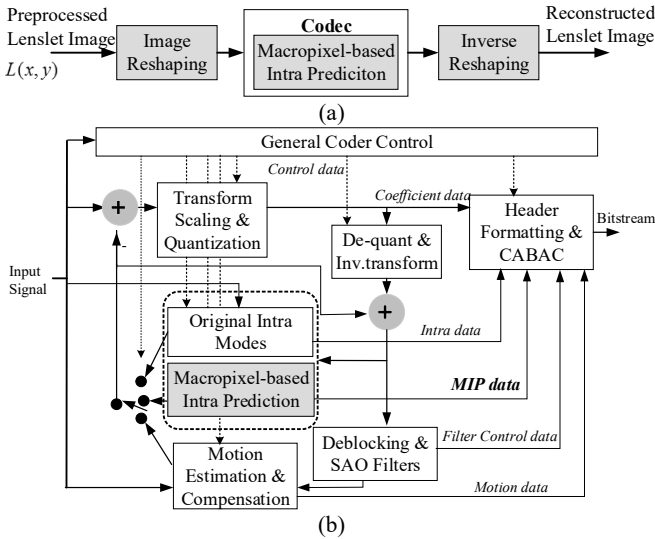


Fig. 7. (a) The proposed lenslet image compression system; (b) the proposed encoder architecture.

A. Multi-block weighted prediction mode

The MWP mode is to predict the block by the combination of co-located reference blocks in the adjacent reconstructed macropixels. Compared with the existing spatial predictive coding schemes targeting lenslet images, like displacement intra prediction [14-16] and self-similarity compensated prediction [3][18], MWP avoids complexity expensive searching step with better exploration of the macropixel correlations.

1) *Partition Types*: Following the definition in HEVC intra coding, the PU division allowed for MWP is still $2N \times 2N$ and $N \times N$. While, different from that defined in the standard [20], an intra CU can be split into four PUs at all CU sizes for the proposed modes. Thus, MWP supports the PU size from 4×4 to 32×32 . Considering the macropixels in the lenslet image record both spatial and angular light information, to preserve the fidelity of the light field, full color space is exploited, in which the resolution of chroma is the same with that of the luma. The chroma PUs go through the same process of intra prediction. If the corresponding luma PU selects the proposed mode, the selected mode will be set as one of the candidate modes, the *derived* mode, for the chroma PU.

2) *Reference Blocks*: For the current PU, four reference blocks co-located in the spatially adjacent reconstructed macropixels are selected. Consistent with the PU size of the current block, the reference block size ranges from 4×4 to 32×32 samples. When the PU size is the integer multiples of 16, from 16×16 to 32×32 , four adjacent reference blocks located left to, left-above, above and right-above the current PU, as shown in Fig. 8 (a) and (b), in the same size are selected. As the current PU size is 4×4 or 8×8 , four co-located blocks in spatially adjacent reconstructed 16×16 blocks are utilized as reference blocks, as shown in Fig. 8 (c) and (d). As some of the reference

blocks are unavailable, i.e. they have not been reconstructed yet, only available reference blocks are used.

As mentioned above, the proposed method allows the CU to be split into four PUs at all the CU sizes. Thus, if the current CU size is 32×32 and the PU size is 16×16 , the right-above reference block is unavailable for the PU located at the bottom-right corner, as the red block shown in Fig. 9. For the case, we copy the reconstructed 16×16 block that is left to the current PU as the right-above reference block to solve the reference missing problem.

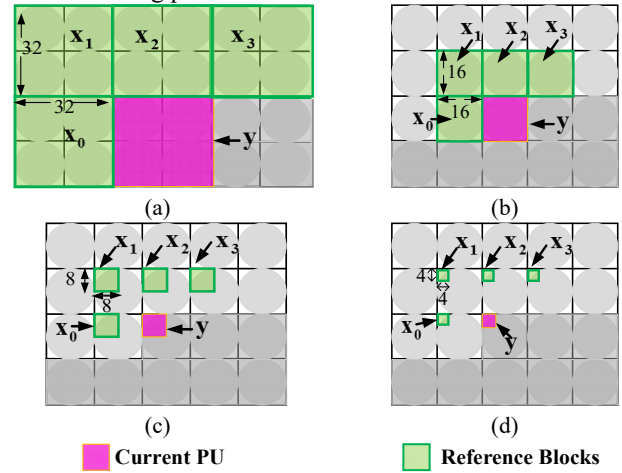


Fig. 8. The relationship between the current prediction unit (the block in magenta) and the reference blocks (those in green) as the PU size equals to: (a) 32×32 ; (b) 16×16 ; (c) 8×8 ; and (d) 4×4 .

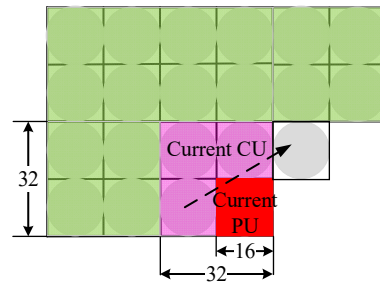


Fig. 9. The right-above reference block generated by copying the reference block left to the current PU as PU size equals to 16×16 and the CU size is 32×32 .

3) *Sample Prediction*: The current PU, denoted by y in Fig. 8, is linearly predicted by

$$y' = w_0 x_0 + w_1 x_1 + w_2 x_2 + w_3 x_3, \quad (1)$$

where y' represents the prediction of y ; x_i is a reference block, i.e. a green block shown in Fig. 8; and w_i is the weighting parameter corresponding to x_i . As x_i is unavailable, w_i equals to zero. The summation of w_i equals to 1. Picking out the available reference blocks from the four candidate reference blocks, their weighting parameters are derived by minimizing the Euclidean distance between the current block and the reference blocks as

$$\begin{aligned} & \text{minimize} \quad \|Xw - y\|_2^2 \\ & \text{subject to} \quad \mathbf{1}^T w = 1, \\ & \quad \quad \quad w \geq 0 \end{aligned} \quad (2)$$

where w is the vector of weights, in which each entry is w_i ; y is the vectorized sample values of the current PU; X is the matrix of reference blocks in which each column is a vectorized x_i .

For each PU, Eq. (2) is solved by the logarithmic barrier method [61] to derive the weighting parameters. After converting Eq. (2) to the equality constrained problem using barrier method, a Newton's method is applied to derive the current weights at each iteration. The initial weights are assigned empirically according to the correlation analysis among macropixels. The maximum number of iterations is set to achieve a tradeoff between the computational complexity and compression efficiency. If the logarithmic barrier method can converge before the maximum iterations, the derived weights will be utilized. Otherwise, the initial values will be assigned to the weights.

Using the derived weights, the predicted block can be generated by Eq. (1). Since the weights are floating point value in the range of 0 and 1, the multiplication and the generated floating-point values of the predicted samples are not friendly to encoding. Thus, the weights are scaled to the range from 0 to 127 and rounded to the closest integer. The scaling range is chosen based on the consideration of granularity and the representation efficiency. Then, the predicted sample generation is updated to be:

$$y' = (w_0x_0 + w_1x_1 + w_2x_2 + w_3x_3 + 64) \gg 7, \quad (3)$$

where \gg denotes a bit shift operation to the right.

Using MWP mode, the residual between the current PU and the predicted samples calculated by Eq. (3) will be coded if the mode is selected by RDO. Also, the weights will be coded into the bitstream. The details in mode decision and weights coding are introduced in the Subsection D in the following.

B. Co-located single-block prediction mode

MWP mode can provide good prediction for the current PU by exploiting spatial correlation among the pixels under the adjacent microlens. However, it also introduces overhead bits cost by the weights, which may affect the coding efficiency especially at low coding bitrate. Thus, based on the strong correlations among the adjacent blocks presented in the reshaped lenslet image, as demonstrated in Fig. 6, a co-located single-block prediction (CSP) is proposed to predict the current PU by:

$$y' = x_i. \quad (4)$$

x_i is a reference block selected from the four co-located reference blocks, the blocks in the green as shown in Fig. 8.

Four CSP modes are added to use reference block left to, top-left to, above and top-right to the current PU individually as the prediction. The mode signaling method is introduced in Subsection D also.

C. Boundary matching based prediction mode

CSP can predict the current PU easily with fewer overhead bits relative to MWP mode, while in some case the prediction is relatively coarse especially when the adjacent macropixels are imaging the object boundaries. Thus, to exploit the correlation among the neighboring macropixels and to reduce the overhead bits simultaneously, a boundary matching based prediction (BMP) mode is proposed.

Similar to MWP, BMP also uses block-based linear weighted prediction to predict the current PU. The reference blocks are the same with those shown in Fig. 8. Distinctively, considering the correlation of intensity values between the current PU and reference blocks can be reflected by their spatial boundary pixels to some extent, BMP uses boundary samples in the reference blocks, instead of all the samples in the reference blocks, and the reconstructed samples around the current PU, instead of the original samples in the current PU, to derive weighting parameters. The samples used are those colored in green and magenta in Fig. 10.

Thus, column i in X in Eq. (2) is updated by vectorizing the top sample row and the left sample column in x_i , denoted by x_i' in Fig. 10, if x_i is available. y is updated by vectorizing the reconstructed sample row/column above/left-to the current PU, as shown in Fig. 10. Solving Eq. (2) by logarithmic barrier method [61], the weights are derived to generate the predicted samples using Eq. (3), in which x_i is the reference block as defined in Fig. 8.

Since the reconstructed boundary samples are available both at the encoder and the decoder, BMP does not need to encode the weighting parameters, which shows its advantage relative to MWP. Although Eq. (2) needs to be solved at the decoder side also, the complexity increment introduced is still acceptable. The complexity results are provided in the next section.

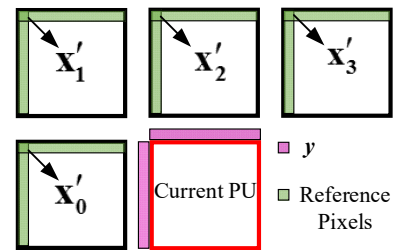


Fig. 10. Boundary pixels used by BMP in Eq. (2).

D. Mode selection and coding

The proposed three types of macropixel-based intra prediction try to exploit the spatial correlations among the macropixels with tradeoff in complexity and overhead bits. MWP can generate an accurate prediction based on the optimization results between the current block and the reference blocks. While, solving the optimization problem will introduce complexity overhead to the encoder and the overhead bits cost by coding the weights will affect the compression efficiency especially at low bit rate. CSP shows low complexity overhead introduced together with low overhead bits, while the prediction may be a bit coarse as the adjacent macropixels are imaging the object boundaries. BMP is in between, which reduces overhead bits relative to MWP and may solve the problem of CSP for the blocks around object boundaries. However, its prediction may not be as accurate as that of MWP since the weights are determined by limited number of reference samples. Also, it introduces some computational complexity overhead to the decoder to derive the weights.

Thus, to fully exploit the advantages of the three types of prediction simultaneously, they are added to RDO process of intra prediction in HEVC and becomes candidate intra prediction modes with the other 35 intra modes defined in

HEVC [62]. The mode with the lowest RD cost will be selected as the coding mode for the current PU. To reduce the complexity of RDO, if the PU in size of $2N \times 2N$ selects one of the proposed modes as the best mode, the PUs in size of $N \times N$ will use the same proposed mode with updated weights, if needed, as CU size is $2N \times 2N$. Since the lenslet image always uses full color space, YUV 4:4:4, to preserve the fidelity, the chroma PUs are predicted and coded using five defined modes in the standard, as listed in Table II. Intra_Derived mode inherits the intra prediction mode from the corresponding luminance PU directly. As our proposed mode is selected by the luminance PU, it will work as Intra_Derived mode during RDO of chroma PU. The reference pixels as described in each mode will be used according to the chroma PU size. The weights will be reoptimized and coded for PU of U and V component individually. Table II shows the updated mode specification in luma intra prediction and in chroma intra prediction.

TABLE II. SPECIFICATION OF LUMA AND CHROMA INTRA PREDICTION MODES

Luma Mode Index	Luma intra prediction mode	Chroma Mode Index	Chroma intra prediction mode
0~34	Original luma intra mode	0	Planar
35	CSP(left)	1	Angular(26)
36	CSP(above)	2	Angular(10)
37	CSP(left-above)	3	DC
38	CSP(right-above)	4	Derived (the proposed mode if selected by Luma PU)
39	MWP		
40	BMP		

Similar to that defined in HEVC, three most probable modes are selected based on the modes of the PUs left to and above the current PU. If the selected mode of the current PU is an element in the set of the most probable modes, the index in the set is transmitted to the decoder. Otherwise, a 6-bit fixed length code is used to signal the mode. While, in chroma mode coding, it adopts the same coding methods in HEVC.

For the weights value in MWP, a coding approach based on most probable weights is applied, which is similar to the way of most probable mode coding. Since the summation of the weights is 127, only $M-1$ weights are coded, where M is the number of available reference blocks. Also, the set of the three most probable weights is established, in which the elements are selected from the PU left to and above the current PU. The default candidates in the set are assigned with the weight value 0, 1 and 127. When the weight values of the PU above and left to the current PU are the same, the value and two closest weight values are selected to construct the set of the most probable weights. For the weight in the most probable weight set, its index in the set is transmitted. For that outside of the set, 7-bit fixed length code is used for signaling.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we demonstrate the effectiveness of the proposed algorithm. First, the test conditions are introduced in detail. Then, experimental results including the compression efficiency comparison, computational complexity analysis and mode selection statistics are provided.

A. Test conditions

To measure the compression performance of the proposed method, twelve plenoptic images including six representative images downloaded from the JPEG Pleno dataset [58] and six images captured by us are tested. The lenslet images with resolution 7728×5368 are captured by *Lytro Illum* cameras, which are decoded from the raw files using “Light Field Toolbox for Matlab” [59]. The sample images are shown in Fig. 11. Since the images come from several plenoptic cameras and the optical parameters of the cameras vary slightly due to the manufacturing technologies, testing them can demonstrate the robustness of the proposed approach to the lenslet images captured by different plenoptic cameras. The end-to-end processing workflow recommended by JPEG Pleno [8][9], including demosaicing, devignetting, slicing and rendering, is applied to generate 15×15 subaperture images with spatial resolution 625×434 . The demosaicing process that converts the raw Bayer-pattern to RGB color image uses conventional linear demosaicing method [55] with default parameters in [59]. Devignetting is used to correct vignetting effect by dividing the raw image by the white image. Considering the proposed method targets compressing the lenslet image with high quality, gamma correction for light field is not applied to guarantee that the objective evaluation reflects the real performance.

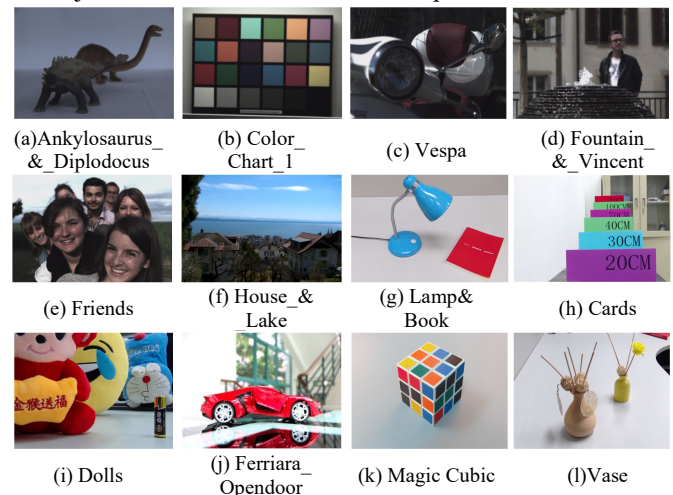


Fig. 11. Tested plenoptic images. (a)-(f): Images from JPEG Pleno database [58]; (g)-(l): images captured by our own *Lytro Illum* cameras.

The proposed macropixel-based intra prediction is implemented into the reference software of HEVC Format Range Extension (RExt) [63] profile, HM-16.9SCM8.0 [64], as additional intra prediction modes according to Fig. 7. After converting the lenslet images from RGB to YUV4:4:4 color space, the tested images are coded by “All Intra” setting as defined in [65] under RExt configurations using QP values of 26, 32, 38 and 44. The RD performance is measured in terms of BD-Bitrate [66]. The bitrate in BD-Bitrate is defined by bit-per-pixel (bpp) that is calculated via dividing the number of bits obtained by the total number of pixels in the input plenoptic images. The PSNR in BD-Bitrate is computed as the mean of all PSNR values for each individual view as defined in [9]. It is given by:

$$PSNR = \frac{1}{kl} \sum_{k=2}^{k-1} \sum_{l=2}^{l-1} PSNR(k, l), \quad (5)$$

in which $PSNR(k, l)$ is the conventional objective metric computed for the k -th and l -th individual view according to:

$$PSNR(k, l) = 10 \log_{10} \frac{255^2}{MSE(k, l)}, \quad (6)$$

$$\text{and } MSE(k, l) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - R(i, j)]^2. \quad (7)$$

m and n are the dimensions of a rendered individual view in units of pixels. $I(i, j)$ and $R(i, j)$ are the values of the pixels in position of (i, j) in the view rendered from the decoded lenslet image and that rendered from the non-compressed reference lenslet image. The rendering method recommended by light field compression evaluation method in [9][67] is used with the default rendering configurations [9].

TABLE III. CODING CONFIGURATIONS OF TESTING CASES

Testing cases	Coding tools configurations
HEVC	HEVC RExt Profile, bypassing the gray blocks in Fig. 7
IBC	“HEVC” + IBC mode proposed in [21]
LLE	“HEVC” + LLE method proposed in [22]
PVTA	“HEVC” + PVTA method proposed in [42] with Low Delay configurations
IR	“HEVC” + gray blocks in Fig.7 except the proposed intra prediction modes, i.e. “HEVC” + image reshaping (IR) proposed by us in [51]
MWP	“HEVC” + the proposed MWP mode
CSP	“HEVC” + the proposed 4 CSP modes
BMP	“HEVC” + the proposed BMP mode
ThreeModes	“HEVC” + MWP + CSP + BMP
IR+IBC	“HEVC” + IR + IBC
IR+LLE	“HEVC” + IR + LLE
IR+MWP	“HEVC” + IR + MWP
IR+CSP	“HEVC” + IR + CSP
IR+BMP	“HEVC” + IR + BMP
IR+MWP+CSP	“HEVC” + IR + MWP + CSP
IR+CSP+BMP	“HEVC” + IR + BMP + CSP
IR+MWP+BMP	“HEVC” + IR + MWP + BMP
Proposed	“HEVC” + IR + MWP + CSP + BMP

B. Experimental results

1) Comparison among the combinations of the proposed coding tools

The efficiency of the proposed intra prediction modes and that of the combinations of the coding tools are evaluated in this subsection. First, the efficiency of each prediction mode and that of image reshaping are listed in Table IV using *HEVC* as the benchmark. As shown in the table, the image reshaping method, testing case *IR*, can achieve 13.0% bitrate reduction on average compared with *HEVC* because of making the macropixel structure be friendly to the block based coding architecture. All the three proposed macropixel-based intra prediction modes *MWP*, *CSP* and *BMP* can improve the compression efficiency obviously, in which applying *CSP* individually achieves the highest bitrate reduction of 20.4%. Applying the three modes together, *Three Modes*, can further improve the coding efficiency, although it may not that significant relative to *CSP*.

Secondly, the compression efficiency of the combinations of the coding tools are evaluated in Table V. It can be found that applying the proposed intra prediction mode to the reshaped

image can provide much higher improvement in the compression efficiency, which is even larger than directly adding the efficiency improvement achieved by each tool individually. For an instance, “IR+MWP vs. HEVC” in Table V is much larger than adding “IR vs. HEVC” with “MWP vs. HEVC” in Table IV. Also, the bitrate savings achieved by all the combinations are much bigger than using the coding tools individually. Comparing the effectiveness of all the combinations, the testing case that performs the best is the *Proposed* approach which integrates the four coding tools together. An obvious bitrate reduction, 47.0% on average, can be achieved. The second-best case is *IR+MWP+CSP* which outperforms HEVC by 46.5% on average. Randomly selecting images “Vase”, “Magic Cubic” and “Ankylosaurus_ & Diplodocus” as instances, the RD performance of each combination at different QPs are shown in Fig. 15 (a) and (b). They demonstrate that the proposed coding tools improve the compression efficiency obviously at all tested bitrates.

TABLE IV. BD-BR COMPARISON FOR THE PROPOSED PREDICTION MODES AND IMAGE RESHAPING

Image Name	IR vs. HEVC	MWP vs. HEVC	CSP vs. HEVC	BMP vs. HEVC	Three Modes vs. HEVC
Ankylosaurus_ & Diplodocus	-30.0%	-33.1%	-42.5%	-21.0%	-44.6%
Color_Chart_1	-15.6%	-42.1%	-47.0%	-28.1%	-48.0%
House_&lake	-34.0%	-40.4%	-51.0%	-25.5%	-51.5%
Fountain_ & Vincent	-1.5%	-22.1%	-22.0%	-15.1%	-23.1%
Friends	-6.4%	-5.3%	-6.3%	-5.6%	-7.8%
Vespa	-1.6%	-17.3%	-19.1%	-13.8%	-20.5%
Lamp&Book	-15.2%	-0.6%	-1.1%	-1.5%	-2.0%
Cards	-8.6%	-9.9%	-11.0%	-4.8%	-11.7%
Dolls	-12.5%	-8.5%	-7.1%	-7.2%	-10.2%
Ferriara_ Opendoor	-1.7%	-12.7%	-12.5%	-9.8%	-14.1%
MagicCubic	-13.9%	-6.3%	-13.6%	-6.4%	-15.0%
Vase	-14.4%	-11.6%	-12.1%	-6.5%	-13.8%
Average	-13.0%	-17.5%	-20.4%	-12.1%	-20.8%

2) Comparison among different coding methods

To demonstrate the effectiveness of the proposed algorithm, six testing cases as *IBC* [21], *LLE* [22], *IR+IBC*, *IR+LLE*, *PVTA* [42] and *Proposed* are tested. Among the six testing cases, *PVTA* is a pseudo video coding approach which generates the pseudo video by tiling the lenslet image and compresses the video using the “Low Delay” configuration. The compression results using *HEVC* as benchmark are shown in Table VI and the compression improvements achieved by the proposed approach relative to the other methods are listed in Table VII. It can be found that *IBC* [21], *LLE* [22] and *PVTA* [42] can improve the compression efficiency of lenslet image obviously although the improvement achieved by *PVTA* fluctuates heavily with the change in the content. It is also interesting to see that by cooperating with the proposed image reshaping method, *IR+LLE* and *IR+IBC* can further improve the coding efficiency relative to *IBC* and *LLE*, as shown in Table VI. While, even under such improvement, the proposed approach can still outperform them significantly. The proposed approach can achieve a maximum of 80.5% bitrate reduction with an average of 47.0% relative to *HEVC*. It outperforms *IBC/LLE/PVTA* by

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) < 9

27.7%/22.7%/45.0% bitrate reduction on average, which is very beneficial to lenslet data storage and transmission. Similarly taking images “Vase”, “Magic Cubic” and “Ankylosaurus_&_Diplodocus” as instances, the RD performance at different QPs of *Proposed*, *IBC*, *LLE*, *PVTA* and *HEVC* are shown in Fig. 15 (c). They demonstrate that the proposed method performs much better than other coding methods at all compression ratios.

TABLE V. BD-BR COMPARISON AMONG THE COMBINATIONS OF THE PROPOSED PREDICTION MODES

Image Name	IR+MWP vs. HEVC	IR+CSP vs. HEVC	IR+BMP vs. HEVC	IR+MWP+CSP vs. HEVC	IR+CSP+BMP vs. HEVC	IR+MWP+BMP vs. HEVC	Proposed vs. HEVC
Ankylosaurus_&_Diplodocus	-59.2%	-66.7%	-60.4%	-70.8%	-65.8%	-65.1%	-71.0%
Color_Chart_1	-75.9%	-75.8%	-69.3%	-80.9%	-76.2%	-78.1%	-80.5%
House_&lake	-67.5%	-71.0%	-68.0%	-72.2%	-71.7%	-71.4%	-72.6%
Fountain_&_Vincent	-40.5%	-37.8%	-31.4%	-42.0%	-37.8%	-41.0%	-41.8%
Friends	-19.9%	-17.6%	-16.4%	-22.1%	-19.1%	-21.0%	-22.7%
Vespa	-38.7%	-34.3%	-28.6%	-41.5%	-35.1%	-39.5%	-41.7%
Lamp&Book	-24.1%	-22.0%	-22.1%	-26.1%	-25.0%	-26.3%	-27.5%
Cards	-43.2%	-34.5%	-32.2%	-45.7%	-38.2%	-43.8%	-45.5%
Dolls	-33.7%	-27.6%	-24.0%	-35.2%	-28.7%	-34.8%	-35.8%
Ferriara_Opendoor	-28.1%	-23.7%	-19.2%	-29.2%	-24.8%	-28.5%	-29.3%
MagicCubic	-40.8%	-44.5%	-42.9%	-49.6%	-47.5%	-47.3%	-51.9%
Vase	-41.9%	-33.1%	-36.1%	-43.3%	-38.5%	-42.7%	-43.6%
Average	-42.8%	-40.7%	-37.5%	-46.5%	-42.3%	-45.0%	-47.0%

TABLE VI. COMPARISON AMONG DIFFERENT METHODS USING HEVC AS BENCHMARK

Image Name	IBC [21] vs. HEVC	LLE [22] vs. HEVC	IR+IBC vs. HEVC	IR+LLE vs. HEVC	PVTA [42] vs. HEVC	Proposed vs. HEVC
Ankylosaurus_&_Diplodocus	-45.1%	-43.6%	-60.6%	-60.9%	-57.6%	-71.0%
Color_Chart_1	-66.2%	-74.5%	-74.7%	-77.6%	-37.9%	-80.5%
House_&lake	-53.6%	-61.4%	-65.9%	-67.9%	-46.8%	-72.6%
Fountain_&_Vincent	-32.2%	-41.2%	-35.1%	-39.2%	-18.6%	-41.8%
Friends	-6.2%	-5.2%	-13.8%	-13.3%	20.6%	-22.7%
Vespa	-22.1%	-21.4%	-27.4%	-28.4%	0.8%	-41.7%
Lamp&Book	-5.0%	-1.2%	-21.6%	-19.1%	20.5%	-27.5%
Cards	-13.4%	-30.0%	-30.4%	-27.4%	-12.9%	-45.5%
Dolls	-19.4%	-20.9%	-26.0%	-29.6%	17.8%	-35.8%
Ferriara_Opendoor	-25.5%	-32.7%	-26.3%	-28.2%	0.9%	-29.3%
MagicCubic	-27.2%	-10.7%	-39.7%	-32.9%	9.9%	-51.9%
Vase	-13.1%	-11.8%	-25.3%	-25.8%	-13.5%	-43.6%
Average	-27.4%	-29.6%	-37.2%	-37.5%	-10.5%	-47.0%

TABLE VII. COMPARISON AMONG DIFFERENT METHODS

Image Name	Proposed vs. IBC	Proposed vs. LLE	Proposed vs. PVTA
Ankylosaurus_&_Diplodocus	-45.0%	-48.4%	-51.5%
Color_Chart_1	-40.6%	-18.4%	-70.0%
House_&lake	-40.2%	-24.2%	-37.1%
Fountain_&_Vincent	-13.8%	0.3%	-36.2%
Friends	-17.4%	-18.3%	-35.3%
Vespa	-24.8%	-25.6%	-45.0%
Lamp&Book	-23.1%	-26.3%	-42.1%
Cards	-31.4%	-19.8%	-43.9%
Dolls	-24.9%	-16.9%	-45.2%
Ferriara_Opendoor	-4.9%	6.0%	-32.2%
MagicCubic	-32.6%	-45.7%	-59.8%
Vase	-34.1%	-35.2%	-41.7%
Average	-27.7%	-22.7%	-45.0%

3) Computational complexity analysis

To evaluate the computational complexity, execution time is retrieved for the testing cases using a PC with Intel® Core™ i7-4790 CPU @ 3.60 GHz with 12GB RAM and 64-bits Windows 7 operating system. Taking the execution time of *HEVC* as the basic unit, the relative execution time ratios of the testing cases are summarized in Fig. 12. As shown in the figure, among all the testing cases, *PVTA* consumes the longest execution time because that it is an inter-frame coding technique. Among the proposed intra prediction modes, *CSP* presents the lowest complexity, which is much lower than *IBC* and *LLE*, by skipping the spatial search process. The ascending order of the computational complexity of the proposed modes is *CSP*, *MWP* and *BMP*. Although the dimension of \mathbf{y} and \mathbf{x}_i used in *BMP* is smaller than that in *MWP* in Eq. (2), logarithmic barrier method always needs more iterations for convergence, which results in

higher complexity during encoding. It is interesting to see that combinations of CSP, e.g. *IR+MWP+CSP*, *IR+CSP+BMP*, and *Proposed*, present lower computational complexity than those without *CSP*, i.e. the complexity of *IR+MWP+CSP*, *IR+CSP+BMP*, and *Proposed* is lower than *IR+MWP*, *IR+BMP*, and *IR+MWP+BMP*, respectively. The reason is that for the CU size is $2N \times 2N$, as the *CSP* is selected as the best mode for PU in size of $2N \times 2N$, it will be used as the best mode of PU in size of $N \times N$ for RD cost comparison to eliminate the computations consumed by *MWP* and *BMP* in calculating the weights. The *Proposed* scheme presents a bit higher computational complexity relative to *IBC* and *LLE* and lower complexity relative to *PVTA*. While, -27.7%/-22.7%/45.0% bitrate reduction can be achieved according to that shown in Table VII. It is found that the modes combination *IR+MWP+CSP* presents a better trade-off between the compression efficiency and the computational complexity if checking the performance between Table V and Fig. 12. Its complexity is 32.45% lower than the *Proposed* and the compression efficiency is 2.4% lower than the *Proposed*. While, its compression efficiency is still much higher than that of *IBC*, *LLE* and *PVTA* as shown in Table VIII and the complexity increment is much less. So, *IR+MWP+CSP* can be a recommended approach if the computing resources at the encoder is limited.

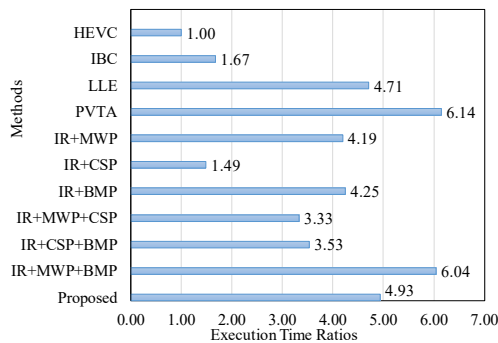


Fig. 12. Execution time ratios of different coding methods relative to HEVC.

TABLE VIII. COMPRESSION COMPARISON FOR *IR+MWP+CSP*

Image Name	<i>IR+MWP+CSP</i>		
	vs. <i>IBC</i>	vs. <i>LLE</i>	vs. <i>PVTA</i>
Ankylosaurus_ & Diplodocus	-44.4%	-47.9%	-51.5%
Color Chart 1	-41.6%	-19.4%	-70.4%
House & lake	-37.6%	-22.5%	-35.9%
Fountain & Vincent	-14.0%	0.1%	-36.2%
Friends	-16.8%	-17.6%	-35.0%
Vespa	-24.5%	-25.3%	-45.2%
Lamp&Book	-21.7%	-25.0%	-41.2%
Cards	-31.6%	-19.8%	-43.9%
Dolls	-24.2%	-16.1%	-44.7%
Ferriara Opendoor	-4.8%	6.2%	-32.0%
MagicCubic	-29.0%	-42.9%	-58.7%
Vase	-33.6%	-34.8%	-41.5%
Average	-27.0%	-22.1%	-44.7%

4) Mode selection statistics

The section analyzes the intra mode selection statistics for the proposed method to further demonstrate the effectiveness in improving the compression efficiency. Table X summarizes the

ratio of selected intra mode for luminance component using 4×4 block as a basic unit for image “Fountain_&_Vincent” under different QPs. It is found that compressing the lenslet image directly by *HEVC* results in most PUs selecting DC mode (more than 50% for QPs lower than 44), intra prediction mode 1, and planar prediction mode, intra prediction mode 0. Applying the proposed *IR* algorithm, the proportion of selecting horizontal direction mode and vertical direction mode are becoming much larger relative to that of *HEVC*, which benefits from the spatial correlation improvements introduced. Finally, compressing the lenslet images by the complete solution of the proposed approach, most of PUs select the proposed modes, especially *MWP* and *CSP*. Notably, more blocks will select *MWP* mode at low compression ratio, corresponding to smaller QPs, while a larger proportion of blocks will select *CSP* mode at high compression ratio because of less overhead bits. The statistics illustrate that the proposed scheme provides more precise prediction for intra prediction.

5) Discussion of generalization of the proposed algorithm

Among the existing microlens array arrangements in plenoptic cameras, the hexagonal microlens array is the most advanced arrangement in the commercialized plenoptic cameras to obtain the highest fill-factor. Fill-factor is the maximum coverage of the active area on the sensor. Higher fill-factor corresponds to more efficient acquisition of light field [56][57]. Hence, our method including image reshaping and macroblock-based intra prediction mainly aims at plenoptic images captured by *Lytro Illum* with 15×15 angular resolution recommended by the common test conditions [9] and the testing dataset [58]. Considering that micro-lens array can be of any shape and any arrangement, possible solutions in generalizing the proposed algorithm are discussed in this section, whose further improvement is also under investigation as our future work.

Considering the lenslet image consisting of macropixels with $k \times k$ effective pixels (pixels valid in generating the light field), a possible extension of image reshaping is to rearrange the macropixels by macropixel alignment and adaptive interpolation to generate a regularized image consisting of $n \times n$ blocks with the block centers aligned horizontally and vertically. Different from the original imaging reshaping that guarantees each block is a 16×16 block, in such case, the block can be the smallest block that can cover one macropixel, as the black grids shown in Fig. 13. Then, since the coding unit grids, the grids in red shown in Fig. 13, are misaligned with the block grids, a preliminary extension of the proposed macropixel intra prediction can select the reference blocks according to coordinate relation between the macropixel block and the prediction unit. Like the instance shown in Fig.13 (a), since the current PU (the block in magenta) covers the bottom-right corner of four macropixels (the gray circles), the three reference blocks (the blocks in green), each of which covers the bottom-right corner of the nearest reconstructed macropixels, are selected. The relative position between the current PU and reference block, denoted by D_s in Fig. 13(a), is variant with PU size, which can be calculated by:

$$D_s = \begin{cases} n & N < n \\ \lceil N/n \rceil \times n & N > n \end{cases}, \quad (8)$$

where N represents PU size and $\lceil \cdot \rceil$ rounds the number up to the nearest integer. After selecting the reference blocks, sample prediction for the current PU can be the same with that described above in the proposed MWP, CSP, and BMP modes.

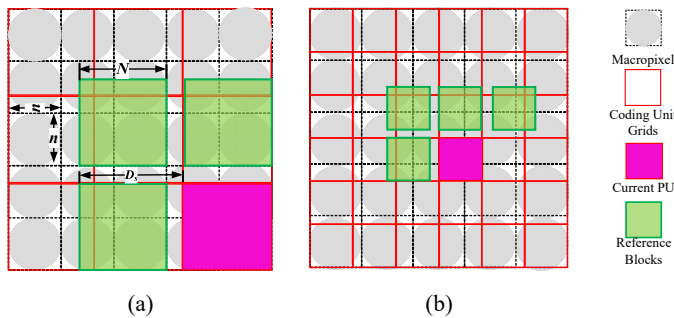


Fig. 13 The relationship between the current prediction unit and the reference blocks for different coding unit size: (a) Coding unit size is larger than macropixel block size; (b) Coding unit size is smaller than macropixel block size.

The performance of the generalized algorithm is further tested on another ten plenoptic images shown in Fig. 14, within which five images captured by Lytro 1.0 are downloaded from light field image dataset [68] with angular resolution 11×11 , and five images are downloaded from Stanford light field archive [69] with angular resolution 17×17 . The test conditions and evaluation methods of compression efficiency are the same with those mentioned above. Since the light field images provided by the Stanford light field dataset are captured by a conventional camera hanging on a lego gantry, the views are arranged regularly on the 17×17 grid, which results in that the synthesized plenoptic image consists of square macropixels in size of 17×17 pixels. Thus, the image reshaping method is skipped for the Stanford data. Table IX summarizes the compression efficiency results by comparing with *HEVC* and *LLE* (the best except the proposed approach). As shown in the table, the generalized image reshaping method, denoted by *GIR*, which is applied to lenslet images captured by Lytro 1.0 can outperform *HEVC* by an average of 16.2% bitrate reduction. Compared with *HEVC* and *LLE*, the generalization of the proposed compression method (denoted by *GProp.*), including *GIR* and the extended macropixel intra prediction, can achieve bitrate reduction by an average of 49.0% and 30.3%, respectively. The results demonstrate that the proposed method can be generalized to benefit compressing plenoptic images with different or novel macropixel shapes and arrangements. While, for the situation that the regularized macropixel block grid is misaligned with the coding unit grid, how to optimize the compression efficiency needs to be further investigated, which has been put as one of our future works as well.

V. CONCLUSIONS

This paper proposes a novel plenoptic image compression scheme, which can efficiently exploit the inherent correlation among macropixels. After applying the previously proposed invertible image reshaping method to the lenslet image, the reshaped image is compressed by adding the three proposed macropixel-based intra prediction modes as additional candidate modes. The proposed modes predict the current PU

by the co-located blocks or their combinations in the spatially adjacent macropixels, which can bring significant compression performance improvement. A maximum of 80.5% bitrate reduction with an average of 47.0% bitrate reduction can be achieved relative to *HEVC* under the same reconstructed light field quality. Also, significant compression performance is demonstrated by outperforming state-of-the-art methods *IBC/LLE/PVTA* by an average of 27.7%/22.7%/45.0% bitrate reduction. Moreover, a better tradeoff between the compression efficiency and computational complexity can be achieved by the combination of MWP and CSP if the computational resources are limited at the encoder. The performance of the proposed scheme can be further improved by designing a specific entropy coding engine and a fast mode selection method, which are under investigating as future works.

Lytro 1.0 Dataset [68], Angular resolution 11×11 ,
lenslet image resolution 3936×3786



Stanford Light Field Data [69], Angular resolution 17×17

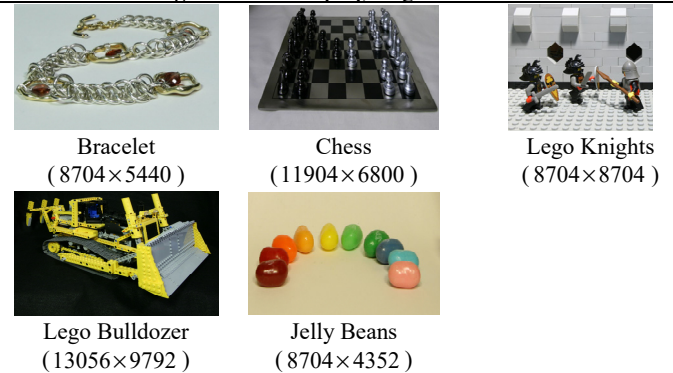


Fig. 14 Tested images with different macropixel sizes.

TABLE IX. COMPRESSION PERFORMANCE FOR LENSLET IMAGES WITH DIFFERENT MACROPIXELS SIZES

Data	Image Name	GIR vs. HEVC	GProp. vs. HEVC	GProp. vs. LLE
Lytro 1.0	BSNMom	-23.1%	-51.8%	-43.2%
	Cocktails	-19.8%	-60.1%	-57.4%
	Dessert	-13.6%	-23.7%	-23.6%
	Edelweiss	-4.1%	-18.4%	-11.7%
	Flat_Toes	-20.6%	-69.0%	-53.4%
Stanford Light Field Database	Bracelet	-	-40.5%	-21.7%
	Chess	-	-41.0%	8.0%
	Lego Knights	-	-71.4%	-48.3%
	Lego Bulldozer	-	-64.3%	-36.5%
	Jelly Beans	-	-50.1%	-15.3%
Average		-16.2%	-49.0%	-30.3%

VI. REFERENCES

- [1] Lytro, <https://www.lytro.com/>.
- [2] Raytrix, <https://www.raytrix.de/>

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) < 12

TABLE X. PROPORTION OF DIFFERENT INTRA PREDICTION MODES

QP	Methods	Planar:0	DC:1	Horizontal: 10	Vertical: 26	CSP				MWP:39	BMP:40	Others
						35	36	37	38			
26	HEVC	14.9%	64.3%	11.9%	1.6%	-	-	-	-	-	-	7.3%
	IR	16.7%	32.4%	14.1%	16.5%	-	-	-	-	-	-	20.2%
	Proposed	4.9%	7.0%	2.6%	1.4%	9.7%	8.4%	0.5%	1.3%	47.4%	11.2%	5.4%
32	HEVC	16.8%	56.8%	16.7%	3.7%	-	-	-	-	-	-	6.0%
	IR	15.7%	23.7%	20.6%	28.5%	-	-	-	-	-	-	11.6%
	Proposed	5.2%	6.2%	4.0%	2.2%	13.5%	16.4%	0.8%	3.0%	32.6%	11.8%	4.2%
38	HEVC	14.8%	51.4%	20.3%	6.3%	-	-	-	-	-	-	7.2%
	IR	12.2%	13.4%	24.7%	40.2%	-	-	-	-	-	-	9.4%
	Proposed	6.3%	6.1%	7.0%	3.6%	13.6%	21.0%	1.1%	4.5%	23.4%	9.6%	3.9%
44	HEVC	13.8%	39.7%	19.6%	16.6%	-	-	-	-	-	-	10.3%
	IR	12.5%	9.6%	18.8%	51.4%	-	-	-	-	-	-	7.7%
	Proposed	8.4%	6.5%	6.7%	5.3%	11.2%	28.4%	1.3%	4.4%	15.2%	8.3%	4.2%

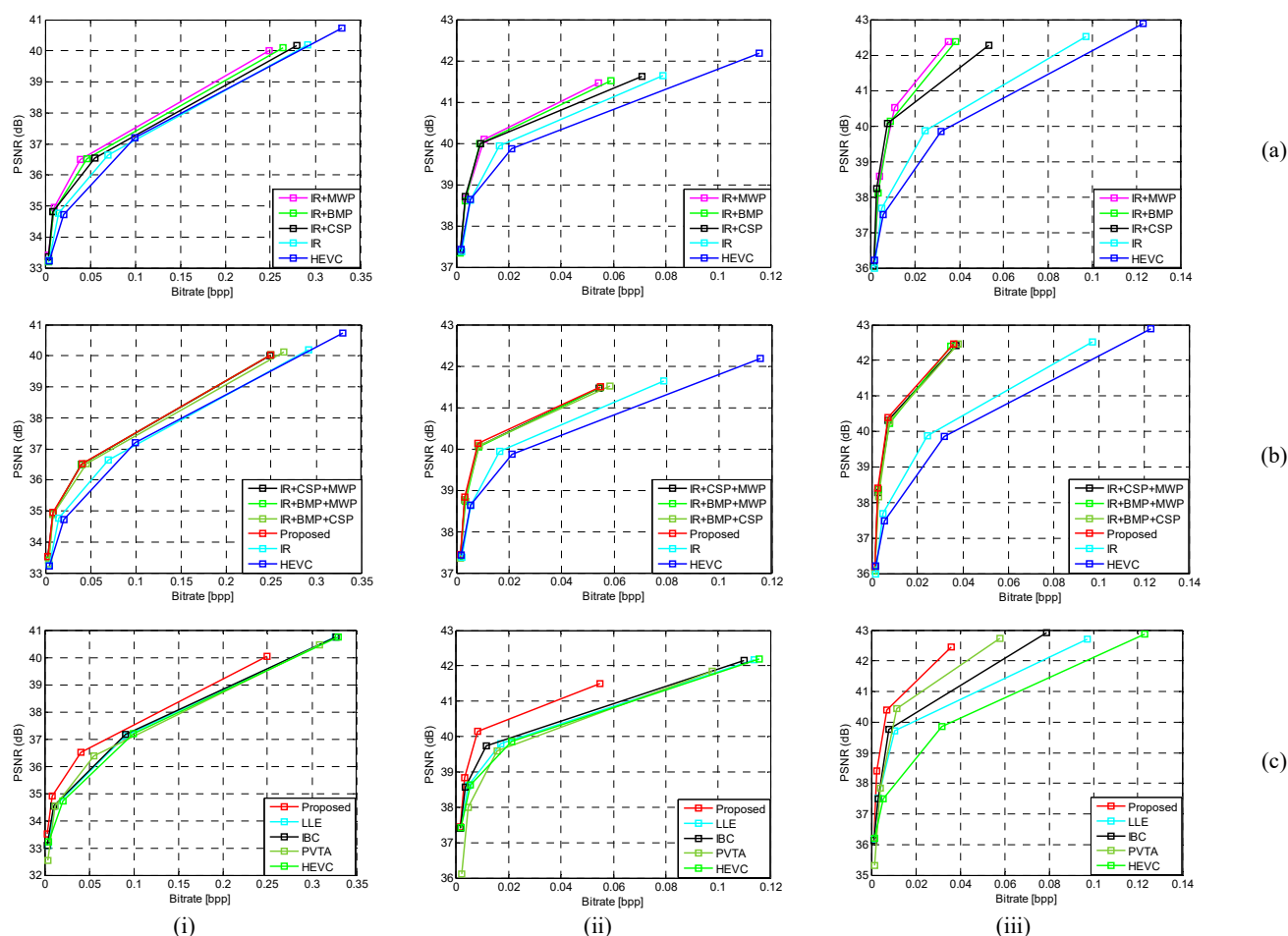


Fig. 15. RD performance of test images. The images in each column: (i) Vase; (ii) Magic Cubic; (iii) Ankylosaurus_& Diplodocus. (a) Comparison among IR, dual-coding-tool and HEVC; (b) Comparison among IR, tri-coding-tool, Proposed and HEVC; (c) Comparison among the coding methods.

[3] M. Martínez-Corral, A. Dorado, H. Navarro, A. Llavador, G. Saavedra, B. Javidi, "From the plenoptic camera to the flat integral-imaging display." Proc. of SPIE - The International Society for Optical Engineering 9117(2014): 91170H-91170H-6.

[4] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: sensing, display, and applications," Applied optics, 2013, 52(4): 546-560.

[5] J. Arai, "Integral three-dimensional television," 2015 14th Workshop on Information Optics (WIO), Kyoto, 2015, pp. 1-3.

[6] N. Li, J. Ye, Y. Ji, H. Ling, J. Yu, "Saliency Detection on Light Field," Computer Vision and Pattern Recognition IEEE (CVPR), 2014:2806-2813.

[7] A. Ghasemi, M. Vetterli, "Scale-invariant representation of light field images for object recognition and tracking," IS&T/SPIE Electronic Imaging, 2014:159-164.

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) < 13

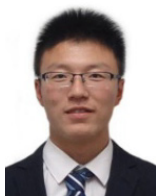
- [8] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, 2016, 23(4): 14-20.
- [9] ISO/IEC JTC 1/SC29/WG1 JPEG "JPEG Pleno Call for Proposals on Light Field Coding," N74014, Geneva, Switzerland, Jan. 2017.
- [10] "Technical report of the joint ad hoc group for digital representations of light/sound fields for immersive media applications", ISO/IEC JTC1/SC29/WG11 MPEG2016/M38503, May 2016, Geneva, Switzerland.
- [11] "Working Draft 0.1 of TR: Technical Report on Immersive Media", ISO/IEC JTC1/SC29/WG11/N16718, Geneva, Jan. 2017.
- [12] Grand Challenge on Light-Field Image Compression, <http://www.icme2016.org/>.
- [13] Grand challenges: Light field image coding, <http://2017.ieeeicp.org/GrandChallenges.asp>
- [14] Y. Li, M. Sjostrom, R. Olsson, U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, 26(7):1308-1319.
- [15] Y. Li, M. Sjostrom, and R. Olsson, "Coding of plenoptic images by using a sparse set and disparities." *IEEE International Conference on Multimedia and Expo IEEE*, 2015:1-6.
- [16] Y. Li, R. Olsson, and M. Sjostrom, "Compression of unfocused plenoptic images using a displacement intra prediction," *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1-4.
- [17] C. Conti, L. D. Soares, P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction". *Signal Processing: Image Communication*, 2016, 42: 59-78.
- [18] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," *IEEE International Conference on Multimedia Expo Workshop(ICMEW)*, July 2016, pp. 1-4.
- [19] C. Conti, P. Nunes, and L. D. Soares, "Light field image coding with jointly estimated self-similarity Bi-prediction", *Signal Processing: Image Communication*, 60(2018), pp: 144-159.
- [20] G. J. Sullivan, J. Ohm, W. J. Han, et al. "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits & Systems for Video Technology*, 2012, 22(12):1649-1668.
- [21] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the Emerging HEVC Screen Content Coding Extension", *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 26(1):50-62.
- [22] L. F. R. Lucas, C. Conti, P. Nunes, et al. "Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC", *Signal Processing Conference (EUSIPCO)*, 2014 Proceedings of the 22nd European. *IEEE*, 2014:11-15.
- [23] R. Monteiro et al., "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," 2016 *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, WA, 2016, pp. 1-4.
- [24] Perra C. "Lossless plenoptic image compression using adaptive block differential prediction," *IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE*, 2015:1231-1234.
- [25] C. Perra and D. Giusto, "Light field compression on sliced lenslet array", *International Journal of Internet Technology and Secured Transactions (IJITST)*, 2017.
- [26] A. Vieira, H. Duarte, C. Perra, L. Tavora and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," 2015 *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Orleans, 2015, pp. 494-497.
- [27] D. Liu, P. An, R. Ma, et al. "Disparity compensation based 3D holoscopic image coding using HEVC," 2015 *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, 2015: 201-205.
- [28] C. Conti, P. T. Kovács, T. Balogh, et al. "Light-field video coding using geometry-based disparity compensation," 2014: *The True Vision-Capture, Transmission and Display of 3D Video. IEEE*, 2014: 1-4.
- [29] A. Aggoun, "A 3D DCT compression algorithm for omnidirectional integral images," 2006 *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006*, pp. II-II.
- [30] A. Aggoun, "Compression of 3D integral images using 3D wavelet transform," *Journal of Display Technology*, 2011, 7(11): 586-592.
- [31] R. J. S. Monteiro, P. J. L. Nunes, N. M. M. Rodrigues, S. M. M. Faria, "Light Field Image Coding Using High-Order Intra-block Prediction," *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(7): 1120-1131.
- [32] T. Sakamoto, K. Kodama and T. Hamamoto, "A study on efficient compression of multi-focus images for dense Light-Field reconstruction," *Visual Communications and Image Processing (VCIP)*, 2012 *IEEE*, San Diego, CA, 2012, pp. 1-6.
- [33] D. Taubman and M. W. Marcellin, *JPEG 2000: Image compression fundamentals, standards and practice*. Boston, MA:Kluwer, 2002.
- [34] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits & Systems for Video Technology*, 2003, 13(7):560-576.
- [35] S. Zhao, Z. Chen, K. Yang and H. Huang, "Light field image coding with hybrid scan order," 2016 *Visual Communications and Image Processing (VCIP)*, Chengdu, 2016, pp. 1-4.
- [36] F. Dai, J. Zhang, Y. Ma, Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," *IEEE International Conference on Image Processing IEEE*, 2015: 4733 - 4737.
- [37] S. Zhao, Z. Chen. "Light field image coding via linear approximation prior", 2017 *IEEE International Conference on Image Process (ICIP)*, Sept 2017, pp. 1-4.
- [38] I. Tabus, P. Helin, P. Astola. "Lossy compression of lenslet images from plenoptic cameras combing sparse predictive coding and JPEG2000," 2017 *IEEE International Conference on Image Process (ICIP)*, Sept 2017, pp. 1-4.
- [39] S. Kundu, "Light field compression using homography and 2D warping," 2012 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1349-1352, 2012.
- [40] X. Jiang, M. Le Pendu, R. A. Farrugia, S. S. Hemami and C. Guillemot, "Homography-based low rank approximation of light fields for compression," 2017 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 1313-1317.
- [41] X. Jiang, M. Le Pendu, R. A. Farrugia and C. Guillemot, "Light Field Compression With Homography-Based Low-Rank Approximation," *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(7): 1132-1145.
- [42] C. Perra, P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2016:1-4.
- [43] C. Perra and D. Giusto, "JPEG 2000 compression of unfocused light field images based on lenslet array slicing," 2017 *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2017, pp. 27-28.
- [44] C. Perra and D. Giusto, "Raw light field image compression of sliced lenslet array," 2017 *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, 2017, pp. 1-5.

- [45] C. Perra, "Light field coding based on flexible view ordering for unfocused plenoptic camera images" International Journal of Applied Engineering Research, 2017, pp. 10563-10569.
- [46] C. Perra, "Assessing the quality of experience in viewing rendered decompressed light fields", Multimedia Tools & Applications, 2018(4):1-20.
- [47] D. Liu, L. Wang, L. Li, et al, "Pseudo-sequence-based light field image compression," IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2016:1-4.
- [48] S. Shi, P. Gioia, G. Madec, "Efficient compression method for integral images using multi-view video coding," 18th IEEE International Conference on Image Processing, 2011: 137-140.
- [49] W. Ahmad, R. Olsson, M. Sjostrom. "Interpreting plenoptic images as multi-view sequences for improved compression," 2017 IEEE International Conference on Image Process (ICIP), Sept 2017, pp. 1-4.
- [50] L. Li, Z. Li, B. Li, D. Liu, H. Li, "Pseudo-Sequence-Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," IEEE Journal of Selected Topics in Signal Processing, 11(7): 1107 - 1119, 2017.
- [51] X. Jin, H. Han and Q. Dai, "Image Reshaping for Efficient Compression of Plenoptic Content," IEEE Journal of Selected Topics in Signal Processing, 11(7): 1173-1186, 2017.
- [52] H. Han, X. Jin and Q. Dai, "Lenslet image compression based on image reshaping and macro-pixel Intra prediction," 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, 2017, pp. 1177-1182.
- [53] H. Han, X. Jin and Q. Dai, "Lenslet image compression using adaptive micropixel prediction," IEEE Int'l. Conf. on Image Proc. (ICIP), Beijing, China, Sept. 17-20, 2017, pp. 4008-4012.
- [54] E. Y. Lam, "Computational photography with plenoptic camera and light field capture: tutorial," Journal of the Optical Society of America A, 2015, 32(11):2021-2032.
- [55] D. G. Dansereau, O. Pizarro and S. B. Williams, "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras," Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, Portland, OR, 2013, pp. 1027-1034.
- [56] T. H. Lin, M. C. Lin, C. K. Chao, "A novel and rapid fabrication method for a high fill factor hexagonal microlens array using thermal reflow and repeating spin coating", International Journal of Advanced Manufacturing Technology, 2017, 92(9-12):1-8.
- [57] K. H. Liu, M. F. Chen, C. T. Pan, et al. "Fabrication of various dimensions of high fill-factor micro-lens arrays for OLED package", Sensors & Actuators A Physical, 2010, 159(1):126-134.
- [58] JPEG Pleno Dataset: EPFL Light-field data set, <https://jpeg.org/plenodb/lf/epfl/>.
- [59] D. G. Dansereau, "Light Field Toolbox for Matlab," <http://cn.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4?requestedDomain=true>.
- [60] I. K. Kim, J. Min, T. Lee, "Block Partitioning Structure in the HEVC Standard", IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(12):1697-1706.
- [61] S. Boyd and L. Vandenberghe. "Convex optimization." Cambridge university press, 2004.
- [62] J. Lainema, F. Bossen, W. J. Han, et al, "Intra Coding of the HEVC Standard", IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(12):1792-1801.
- [63] HEVC Format Range Extension, <http://www.hevc.info/rext>.
- [64] Downloaded from: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9+SCM-8.0/
- [65] O. C. Au, X. Zhang, C. Pang, and X. Wen, "Suggested Common test conditions and software reference configurations for Screen Content Coding," Joint Collaborative Team on Video Coding (JCT-VC), Torino, JCTVC-F696, July, 2011.
- [66] G. Bjontegaard, "Calculation of average PSNR difference between RD-curves," ITU-T VCEG-M33, 2001.
- [67] I. Viola, M. Řeřábek, T. Ebrahimi, "Comparison and Evaluation of Light Field Image Coding Approaches," IEEE Journal of Selected Topics in Signal Processing, 2017, 11(7): 1092 – 1106.
- [68] Lytro first generation dataset: <https://www.irisa.fr/temics/demos/lightField/index.html>.
- [69] Stanford light field archive: <http://lightfield.stanford.edu/>.



Xin Jin (S'03–M'09–SM'11) received the M.S. degree in communication and information system and the Ph.D. degree in information and communication engineering, both from Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2005, respectively.

From 2004 to 2005, she was an Intern with the Internet Multimedia Group, Microsoft Research Asia, Beijing, China. From 2006 to 2008, she was a Postdoctoral Fellow with The Chinese University of Hong Kong. From 2008 to 2012, she was a Visiting Lecturer with the Information Technology Research Organization, Waseda University, Fukouoka, Japan. Since Mar. 2012, she has been with Graduate School at Shenzhen, Tsinghua University, China, where she is currently a professor. Her current research interests include computational imaging and power-constrained video processing. She has published over 120 conference and journal papers. Dr. Jin is an IEEE Senior Member, and a member of SPIE and ACM. She is the chair of 3D video compression standard ad-hoc group of the Audio Video Standard Workgroup of China (AVS). She received AVS Outstanding Contributor Award of year 2004 and ISOC Best Paper Award in 2010. She has served on many conference committees, e.g., PCM, VCIP, etc. and served as a reviewer for many transactions, e.g., IEEE TIP, TCSVT, etc. and international conferences, e.g., ICIP, ISCAS, ICME, etc.



Haixu Han received the B.S. degree from the Department of Automation, Dalian University of Technology, Liaoning, China in 2015. He is currently pursuing the M.E. degree in Department of Automation at Tsinghua University, Beijing, China.

His current research interests include light-field image compression and image/video coding. He has publish several papers, including journal paper and conference papers, e.g., the J-STSP, the ICME, the ICIP, the APSIPA, etc.



Qionghai Dai (SM'00) received the B.S. degree in mathematics from Shanxi Normal University, China, in 1987, and the M.E. and Ph.D. degrees in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since being a Postdoctoral Researcher in the Automation Department, he has been with the

Media Lab, Tsinghua University, China, where he is currently an Associate Professor and Head of the Lab. His research interests are in signal processing, broad-band networks, video processing, and communication.