# Content-based Light Field Image Compression Method with Gaussian Process Regression

Deyang Liu, Ping An, Ran Ma, Wenfa Zhan, Xinpeng Huang, Ali Abdullah Yahya

***Abstract*—Light field (LF) imaging enables new possibilities for digital imaging, such as digital refocusing, changing of focus plane, changing of viewpoint, scene-depth estimation, and 3D scene reconstruction, by capturing both spatial and angular information of light rays. However, one main problem in dealing with LF data is its sheer volume. In this context, efficient compression methods are needed for such a particular type of content. In this paper, we propose a content-based LF image-compression method with Gaussian process regression to improve the compression efficiency and accelerate the prediction procedure. First, the LF image is fed to the intra-frame codec of HEVC. In the prediction procedure, the prediction units (PUs) are classified as non-homogenous texture units, homogenous texture units, and visually flat units, based on the content property of the LF image. For each category, we design a corresponding Gaussian process regression (GPR)-based prediction method. Moreover, we propose a classification mechanism to exactly decide to which category the current PU belongs, so as to adjust the trade-off between the computational burden and the LF image coding efficiency. Experimental results demonstrate that the proposed LF image compression method is superior to several other state-of-the-art compression methods in terms of different quality metrics. Furthermore, the proposed method can also achieve a good visual quality of views rendered from decoded LF contents.**

***Index Terms*—light field image compression, classification image compression, Gaussian process regression, content property, HEVC.**

## I. Introduction

RECENT light field (LF) imaging, which can capture spatial and angular information of light rays, allows several new applications, including low-level image processing (such as digital refocusing and multi-view image extraction), mid-level vision analysis (such as depth estimation, segmentation, and saliency detection), and high-level user interfaces (such as 3D reconstruction, object detection and recognition, and tracking) [1]. Thanks to such wide use of LF imaging, LF technology has a positive impact on multimedia applications.

In order to represent the full LF, Adelson and Bergen [2] introduced a seven-dimensional plenoptic function $L = L(x, y, z, \theta, \phi, \lambda, t)$, where the $(x, y, z)$ represents the viewing position, the $(\theta, \phi)$ represents the light-ray direction, the $\lambda$ represents the light-ray wavelength, and the $t$ represents the time. Since such high-dimensional data are hard to record and handle, the LF model is simplified to 5D by assuming the measured function is monochromatic and time-invariant. By assuming a 3D region is free of occlusions, the light-field model is further simplified to 4D by Levoy and Hanrahan [3] and Gortler et al. [4]. The simplified 4D function $L = L(u, v, x, y)$ can be described as a two-plane representation to locate the light rays passing through from the aperture (object) plane $uv$ to the image plane $xy$. Moreover, the illumination component can be specified in the simplified 4D plenoptic function, so as to support various lighting configurations such as multiple light sources, light sources with different colors, and arbitrary types of light sources [5]. As a result, we can explore the LFs in a digital perspective with advances of computational photography [6].

Many techniques can be used to derive an LF image, such as multi-sensor capture approach, time-sequential capture approach, and multiplexed imaging approach. A typical method to capture an LF image is to use a plenoptic camera, which is achieved through an array of microlenses placed between the main lens and the camera sensor. Light rays emanating from the 3D scene are refracted through the microlens array and recorded by a 2D image sensor. Each microlens captures one perspective view of the 3D scene at a slightly different angle to its neighbors. The recorded light rays through different microlenses are called micro-images (MIs). Depending on the position of the camera sensor and the microlens array relative to the main lens, plenoptic cameras can be divided into the two categories of standard [7] and focused [8] plenoptic cameras. In the standard plenoptic camera, the microlens array is one focal distance away from the sensor. When the main lens in a standard plenoptic camera creates a focused image of the scene, the microlens is not focused on that image, which leads to a low-resolution rendered image. Compared to the standard plenoptic camera, the focused plenoptic camera puts the focal plane of microlenses away from the image sensor plane and allows each microlens to generate a focused MI. Thus, it provides a tradeoff between spatial and angular information. Thanks to the recorded spatial and angular information, we can render view images at different viewpoints and different focused planes from an LF image.

The rapid development of LF technology has also attracted

This work was supported in part by the National Natural Science Foundation of China under Grant 61801006, 61571285, in part by the Key Project on Anhui Provincial Natural Science Study by Colleges and Universities under Grant KJ2018A0361, in part by the Open Fund of the Key Laboratory of Advanced Display and System Applications under Grant P201801, in part by Program for Innovative Research Team in Anqing Normal University and in part by the Foundation of University Research and Innovation Platform Team for Intelligent Perception and Computing of Anhui Province (Corresponding author: Deyang Liu.)

Deyang Liu, Wenfa Zhan, Ali Abdullah Yahya are with Anqing Normal University, Anqing, China. e-mail: (liudeyang@163.com, zhanwf12@163.com and aselwey1@hotmail.com).

Deyang Liu, Wenfa Zhan, Ali Abdullah Yahya are also with the University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University

Ping An, Ran Ma and Xinpeng Huang are with Key Laboratory of Advanced Display and System Applications, Shanghai University, China

many research groups considering the standardization of LF application. The JPEG started a new study, referred to as JPEG Pleno [9], to standardize the LF image coding methods. The MPEG also includes LF image coding into MPEG-I Visual standardization project [10].

A large amount of data is required to adequately represent an LF scene, which calls for efficient compression schemes for particular types of images for effective transmission and storage. Although many image-compression methods, such as cross-space distortion directed color image compression [11], a closed-form bit-allocation approach for image compression [12], and a content-based image-compression method for arbitrary-resolution display devices [13], have been proposed, they cannot achieve a high coding efficiency for the particular content property of an LF image. In this context, several light-field image-coding frameworks have been proposed, which can be mainly categorized into two groups based on the existing image/video coding standards, such as JPEG [14] and high-efficiency video coding (HEVC) [15]. The first kind of compression method, called pseudo-sequence-based compression, considers creation of a 4D LF representation of an LF image prior to compression. A representative method is to extract multiple view images (VIs) from the LF image and utilize the inter-prediction of HEVC to improve the LF image-coding efficiency [16]. However, the procedure to acquire the 4D LF representation strongly relies on the acquisition device [17] and the geometric information of an LF image, such as the size of the MIs and the focus plane.

The other kind of compression method, called spatial correlation-based compression, tries to compress the LF image directly. Since the MIs in an LF image exhibit repetitive patterns and a lot of redundancy exists between adjacent MIs [25][27][34], such method does not extract VIs from the LF image, and has the potential to achieve better coding efficiency if we can make full use of such high spatial correlation. However, the compression efficiency of this kind method strongly relies on the prediction accuracy of coding blocks. Most LF image-compression methods do not consider the contents of the LF image and cannot derive high prediction accuracy for texture areas. Moreover, high computational complexity restricts the wide use of such a method. How to improve the prediction accuracy according to the content property of an LF image while reducing the computational complexity remains a serious problem.

Our previous work [30] proposed a GPR based prediction method to improve the prediction accuracy. However, the GPR based method strongly relies on the correlations of the current coding block and its adjacent blocks. For some non-homogenous texture areas, such correlations are low, which reduces the prediction accuracy. Moreover, the prediction complexity is high for the GPR based method both in encoder side and decoder side. Therefore, in order to improve the prediction accuracy and reduce the prediction complexity, in this paper, we follow the spatial correlation based compression method and propose a content based LF image compression method with Gaussian process regression. The motivation of this paper lies in two aspects. Firstly, since the non-homogenous texture areas are difficult to predict, a classification method which can segment out the non-homogenous texture areas might be advantageous to improve the prediction accuracy. Secondly, the computational complexity might be reduced by using different prediction methods for different categories.

In the prediction procedure, the PU is used as the basic unit. Firstly, the PUs of LF image is classified into three categories according to the LF image content property. Then, effective prediction methods are proposed for different categories to improve the prediction accuracy and reduce the computational burden in the perspective of Gaussian process regression. The main contributions of this paper are as follows:

1) LF image classification based on the content property. An LF image differs from a natural 2D image, which contains many non-homogenous texture areas (such as the stitched area between neighboring MIs) that are hard to predict. Thus, a region-based classification method which can segment out the non-homogenous texture areas might be advantageous. To improve the prediction accuracy, we divide the PUs of an LF image into three categories: non-homogenous texture units (NHTUs), homogenous texture units (HTUs), and visually flat units (VFUs), based on the content property of the LF image, where the PUs in non-homogenous texture areas constitute an NHTU, the PUs in homogenous texture areas (such as periodic texture areas or regular structure areas) constitute an HTU, and PUs in visually flat regions (such as skies and walls in MIs) constitute a VFU.

2) Adaptive classification mechanism. To exactly decide which category a PU belongs to, we propose a classification mechanism based on the prediction mode correlation and gradient information, so as to further adjust the tradeoff between the computational burden and LF image coding efficiency.

3) GPR-based classification prediction scheme. To achieve high coding efficiency and a low computational burden, we design prediction schemes for different LF image categories. For an NHTU, since the correlation in such an area is low, a hybrid GPR and intra-displacement compensation scheme is proposed to improve the prediction accuracy. For an HTU, a GPR-based prediction method is adopted, since such an area has periodic textures or regular structures. For a VFU, a simplified GPR prediction method is designed to reduce the computational burden.

The rest of the paper is organized as follows. Section II presents a review of related work. Section III gives an overview of the prediction framework and the Gaussian process regression-based LF image-prediction method. The content-based classification and classification mechanism are described in section IV. The GPR-based classification-prediction scheme is introduced in section V. Section VI discusses the simulation results, and the last section is devoted to our conclusions.

## II. Related work for LF image compression

As mentioned above, LF image-compression methods can be mainly divided into two categories: pseudo-sequence-based and spatial-correlation-based. Pseudo-sequence-based coding methods try to extract multiple view images from an LF image, as seen in Fig. 1(b). The extracted VIs are then organized into

sequences before being fed into the image coder and high inter-correlations are used to improve the coding efficiency. A sub-aperture image streaming scheme [18] adopting rotation scan mapping was proposed to compress the lenselet image. A pseudo-sequence-based scheme investigating the coding order of views, rate allocation, and prediction structure was proposed [19] to encode the pseudo-sequence. In order to analyze the relationship between prediction structure and coding performance, a multi-view video-coding prediction structure was put forward [20], where the inter-view prediction is extended into a two-directional parallel structure. In [21], a lossless compression method was presented for rectified LF images, where the high correlation among the VIs was explored. A novel LF image compression method with pseudo-sequence-based 2D hierarchical coding structure was proposed [22]. This method designed a 2D hierarchical reference structure to characterize the inter-correlations among the VIs, then the distance-based reference-frame selection and spatial-coordinate-based motion vector scaling were used to improve the coding efficiency. Moreover, based on the 4D LF representation of an LF image, a depth image-based view synthesis technique [23-24] was proposed for application in LF image compression. The main idea of such methods is to compress a small subset of views and then reconstruct the entire LF image by using the depth-image-based view synthesis technique.

The second kind of LF image-compression method is spatial-correlation-based, and this can be further divided into two categories based on template matching prediction (TMP) and intra-displacement compensation (IDC). Both the TMP- and IDC-based compression methods are based on the fact that the adjacent MIs exhibit repetitive patterns and a lot of redundancy exists, which can be seen in Fig. 1(a). TMP-based compression tries to obtain the best candidate prediction block for the current coding unit (CU) through measuring the similarity between the surrounding neighboring pixels of the current CU and the candidate templates that are already reconstructed or decoded within a searching range. A LF image coding method incorporating locally linear embedding (LLE) [25] was proposed by exploiting the inherent non-local spatial correlation of MIs. The coding efficiency is further improved by combining the self-similarity (SS)-based compensated prediction method and the LLE-based method [26]. A two-stage high order block prediction method [27] was proposed for LF image, where a geometric transformation was applied for intra block prediction. This method allows choosing order of prediction model for each block, and achieves a high coding efficiency both for unfocused and focused LF images. To explore the high spatial correlation of MIs, our previous work [28] put forward a disparity compensation-based LF image coding method. This method was further improved by combining the kernel-based minimum mean-square-error estimation [29] and Gaussian process regression [30].

The IDC-based LF compression methods reuse the reconstructed or decoded block templates to drive a best intra-displacement vector, and then the best matching block is used to predict the current CU. In the IDC-based coding methods, intra-block copy (IBC) [31] is a well-known prediction method. Such method tries to search for a reference block in the reconstructed regions to predict the current CU. A self-similarity (SS) mode [32] was introduced to HEVC by Conti et al to compress the LF image, where the decoded regions were also reused to predict the current CU. Both the IBC mode and SS mode are a single hypothesis prediction, which limits their prediction accuracy. To alleviate such problem, a bi-predicted SS estimation and compensation method was put forward [33][45], where two blocks were searched for in the reconstructed regions and a linear combination of such two blocks were utilized to predict the current CU. Moreover, Li et al [34] also proposed a displacement intra-prediction scheme to reduce prediction errors by using more than one hypothesis.

As mentioned above, an LF image has higher texture complexity than a natural 2D image. The TMP-based compression methods cannot derive an accurate prediction for some non-homogenous texture areas, since the current CU is quite different from its neighbor templates. Although the IDC-based compression methods work well in such areas, they must send the obtained displacement vectors to the decoder side, which reduces the coding efficiency. To alleviate such shortcomings and achieve high LF image-coding efficiency, in this paper, we follow the spatial correlation-based compression method and propose a content-based LF image-compression method with Gaussian process regression. The main idea is to improve the compression efficiency of an LF image and reduce the prediction complexity by combining the advantages of TMP- and IDC-based compression methods. The details will be described in the following sections.

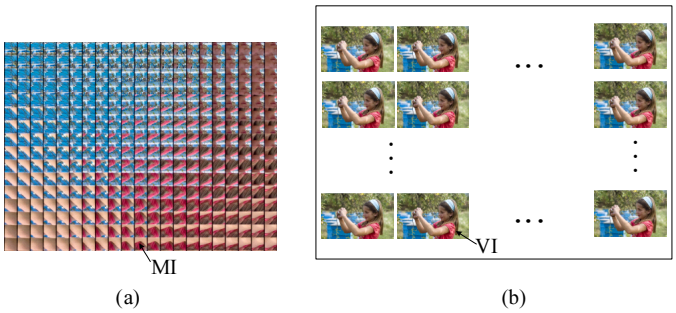



Fig. 1. Acquired focused light field image Laura: (a) lenslet image; (b) 4D LF view images

## III. LF IMAGE-PREDICTION FRAMEWORK AND GPR-BASED PREDICTION METHOD

The proposed LF image-compression method is based on the intra-prediction of the HEVC coding standard. In the prediction procedure, the whole lenslet LF image is fed to the intra-frame codec of the HEVC. The lenslet image is then divided into slices, which are further divided into a sequence of coding tree units (CTUs) with size 64×64. A flexible quad-tree structure is adopted in the prediction process, where the CTUs are recursively split into CUs from size 64×64 to 8×8 [35]. Inside a CU, multiple non-overlapping PUs can be defined. The PUs are used as the basic unit for intra-prediction,which vary from 4×4 to 64×64. Traditional intra-prediction uses the

angular prediction method to predict each PU. HEVC intra-prediction provides 35 directional prediction modes, including the DC and planar modes, which are described in Fig. 2.

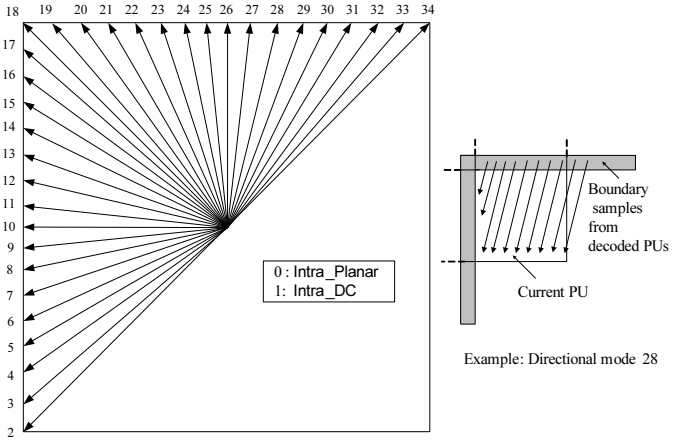



Fig. 2. HEVC angular intra-prediction modes for intra-picture prediction

A Lagrange multiplier is utilized in the mode-decision process to find the optimal prediction mode. The optimal quad-tree depths and the best prediction mode are encoded and transmitted sequentially to the decoder side.

Since the lenslet LF image is different from the natural 2D image, the intra-angular prediction has difficulty achieving an accurate prediction. Our previous work [30] proposed a GPR-based prediction method (also called the GPR mode) to improve the prediction accuracy and acquire high coding efficiency. In order to avoid modifying the bit-stream structure, the GPR mode is implemented in the HEVC architecture through substituting one intra-directional prediction mode. This means that the outputs of the GPR mode will replace the samples produced by the replaced intra-directional prediction mode. The details are briefly summarized as follows.

Suppose that the pixel values in the current PU are compacted in a column vector $\mathbf{x}_0$, and the pixel values in the adjacent templates of current PU with template thickness $T$ are stacked in a column vector $\mathbf{y}_0$; then the vectors $\mathbf{x}_0$ and $\mathbf{y}_0$ can be expressed in a multidimensional vector format $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0)$, as shown in Fig. 3(a). If we assume that the current PU is composed of $n \times n$ pixels, then the vector $\mathbf{x}_0$ has size $n^2 \times 1$ and vector $\mathbf{y}_0$ has size $N = 2nT + T^2$. The goal of the proposed GPR-based coding method is to predict the vector $\mathbf{x}_0$, given a prediction support set $\mathbf{X}$ of size $n^2 \times K$, which is constructed by $K$ nearest neighbor ($K$-NN) template vectors $\{\mathbf{x}_k | k = 1, 2, ..., K\}$ searched within specified searching windows of the current PU. Since the current PU is unknown in the prediction process, we first use the known vector $\mathbf{y}_0$ to search for its $K$-NN template vectors $\mathbf{Y} = \{\mathbf{y}_k | k = 1, 2, ..., K\}$ of size $N \times K$ under the Euclidean distance as the matching criterion within specified searching windows, as shown in Fig. 3(b). After acquiring the template vectors $\mathbf{Y}$, the corresponding vector $\{\mathbf{x}_k | k = 1, 2, ..., K\}$ is used to compose the prediction support vectors $\mathbf{X}$, as shown in Fig. 3(a). To increase the matching accuracy, the template thickness $T$ is set to the size of the current PU in the matching process, as shown in Fig. 3(b).

Given the column vector $\mathbf{x}_0$ for prediction and its prediction

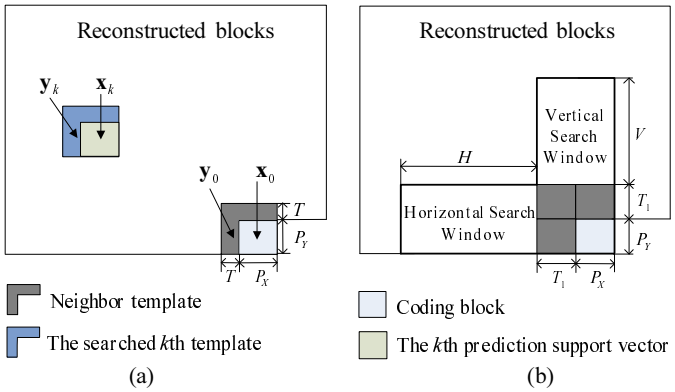



Fig. 3. Multidimensional formalism construction: (a) the neighbor template and the searched $k$th template; (b) searching windows used to derive the prediction support vectors.

support vector $\mathbf{X}$, we suppose that the vector $\mathbf{T} = \{\mathbf{Y}^i, y_0^i\}_{i=1}^N$ is the training set, which is composed of $N$ couples of $K$-tuple training input vectors $\mathbf{Y}^i = [\mathbf{Y}_1^i, \mathbf{Y}_2^i, ..., \mathbf{Y}_K^i]$ and a scalar training output value $y_0^i$ drawn from

$$y_0^i = f(\mathbf{Y}^i) + \epsilon, \tag{1}$$

where $\mathbf{Y}^i$ and $y_0^i$ are the row vectors of vectors $\mathbf{Y}$ and $\mathbf{y}_0$, respectively, and $\epsilon$ is white noise with variance $\sigma_n^2$.

Under the assumption that the vectors $\mathbf{y}_0$ and $\mathbf{Y}$ obey a zero-mean Gaussian process (GP) distribution, the GP prior is given by:

$$\mathcal{P}(\mathbf{y}_0 | \mathbf{Y}) = \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{Y}, \mathbf{Y}) + \sigma^2 \mathbf{I}), \tag{2}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{K}(\mathbf{Y}, \mathbf{Y})$ is a Gram matrix whose element can be determined by: $\mathbf{K}_{i,j} = k(\mathbf{Y}_i, \mathbf{Y}_j)$. The function $k(\mathbf{x}, \mathbf{x}')$ is a kernel function, which is used to measure the similarity between two vectors $\mathbf{x}$ and $\mathbf{x}'$.

Since the GP is a collection of random variables, any finite subset has a joint Gaussian distribution [36]. Therefore, the vectors $\mathbf{x}_0$ and $\mathbf{y}_0$ obey a joint Gaussian distribution, given by

$$\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_Z + \sigma^2 \mathbf{I} & \mathbf{K}_{HZ}^{\mathsf{T}} \\ \mathbf{K}_{HZ} & \mathbf{K}_H \end{bmatrix}\right), \tag{3}$$

where $\mathbf{K}_Z = k(\mathbf{Y}, \mathbf{Y})$, $\mathbf{K}_{HZ} = k(\mathbf{X}, \mathbf{Y})$, and $\mathbf{K}_H = k(\mathbf{X}, \mathbf{X})$.

The marginal distribution of $\mathbf{y}_0$ can be written as [37]

$$\mathcal{P}(\mathbf{x}_0 | \mathbf{y}_0, \mathbf{Y}, \mathbf{X}) = \mathcal{N}(\mu_0, \mathbf{\Sigma}_0) \tag{4}$$

$$\mathbf{u}_0 = \mathbf{K}_{HZ}(\mathbf{K}_Z + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_0 \tag{5}$$

$$\mathbf{\Sigma}_0 = \mathbf{K}_H - \mathbf{K}_{HZ}(\mathbf{K}_Z + \sigma^2 \mathbf{I})^{-1} \mathbf{K}_{HZ}^{\mathsf{T}}. \tag{6}$$

Consequently, $\mathbf{u}_0$ is considered as the best estimator of $\mathbf{x}_0$.

A basic assumption of the GPR model is that points with close inputs are likely to have a similar target output value. In general, a kernel function is used to define the nearness or similarity. In the proposed GPR-based prediction method, the inner product of two input vectors is selected to define the kernel function, which is given by:

$$k(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^{\mathsf{T}} \mathbf{x}'). \tag{7}$$

## IV. Content-based classification and classification mechanism

Different from a natural 2D image, an LF image is composed of numerous MIs. Such a content property makes the LF image difficult to predict, since more non-homogenous texture areas exist. To improve the prediction accuracy of the non-homogenous texture areas and accelerate the prediction procedure, we classify the PUs used as the basic prediction unit into three categories: non-homogenous texture units (NHTUs), homogenous texture units (HTUs), and visually flat units (VFUs). The PUs in non-homogenous texture areas (such as the stitched area between two neighboring MIs) constitute an NHTU, the PUs in homogenous texture areas (such as periodic texture areas or regular structure areas) constitute an HTU, and the PUs in visually flat regions (such as skies and walls in MIs) constitute a VFU. In an NHTU, the texture information of a PU is different from its adjacent templates, which means that it is hard to achieve an accurate prediction of the PU by measuring the similarity between its adjacent templates and corresponding candidate templates. For an HTU, the PUs have similar texture information and structural characteristics, which lead to a high correlation between a PU and its adjacent templates. A VFU is a special case of an HTU. In such areas, the luminance information of a PU and its adjacent templates is simple. By using different prediction methods, the coding efficiency can be improved with low prediction complexity. Fig.4 gives some PU examples in different categories overlaid on the original LF images *Seagull*.

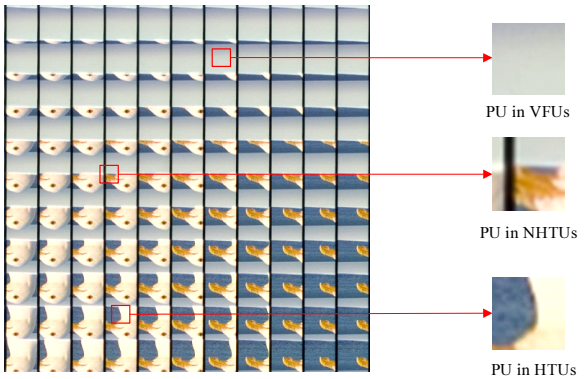



Fig. 4. PUs in different categories overlaid on the original LF images *Seagull*

An effective classification mechanism is also of great importance to the overall compression efficiency. Therefore, an adaptive classification mechanism is proposed, which considers three aspects.

1) To decide if the current PU belongs to an NHTU or HTU, the contents of the current PU should be considered.

2) The visual flatness of LF contents should be considered in the VFU classification method, and the criterion should be simple and fast.

3) In the prediction procedure, the current PU is unknown, and its neighboring blocks with the same size (the left, up, and up-left neighboring blocks) can be used in the classification process.

To account for these three aspects, in this paper, the prediction-mode correlation is used to decide whether the current PU belongs to an NHTU or HTU, and the gradient information is adopted to measure the visual flatness. Below we introduce the classification details.

### A. Prediction-mode correlation-based classification

In an HTU, the PUs have similar texture information and structural characteristics. Since these may lead to a similar prediction mode [39], the PUs in an HTU should have prediction modes similar to their neighboring known blocks. This means that we can apply prediction-mode information to decide whether the current PU belongs to an NHTU or HTU.

Suppose that the optimal prediction modes of current PU and its up, left, and up-left adjacent blocks are denoted by $PM_C, PM_U, PM_L$, and $PM_{UL}$, respectively. According to the above analysis, if $PM_C = PM_L = PM_U = PM_{UL}$, then the current PU can be grouped into an HTU. Otherwise, the current PU is decided to belong to an NHTU. Since the prediction-mode information of the current PU is unknown in the prediction process, we propose to group the current PU into an HTU if $PM_L = PM_U = PM_{UL}$. For simplicity, we define a flag, $flag_{CB}$, to mark whether the current PU belongs to an HTU, given by:

$$flag_{CB} = \begin{cases} 1 & if(PM_L = PM_U = PM_{UL}) \\ 0 & Otherwise \end{cases} . \quad (8)$$

When $flag_{CB}$ equals 1, the current PU is divided into an HTU. Otherwise, the current PU is grouped into an NHTU.

To verify the judgment accuracy, Table I illustrates the accuracy of $PM_C = PM_U = PM_L = PM_{UL}$ at each depth level when $PM_U = PM_L = PM_{UL}$. This judgment accuracy is the probability of $PM_C = PM_U = PM_L = PM_{UL}$ when $PM_U = PM_L = PM_{UL}$ across all the test QPs. The test LF image set covers both the focused LF images (Bike, Fountain, Sergio, and Zhengyun1) and the unfocused LF images (I01, I02, I04, and I06). From Table I, we find that the average judgment accuracy of $PM_C = PM_U = PM_L = PM_{UL}$ at each depth level when $PM_L = PM_U = PM_{UL}$ across all test QPs is from 77.8% to 95.1%, and 81.0% on average. This means that the prediction mode of current PU can be recognized to be the same as its adjacent blocks when $PM_U = PM_L = PM_{UL}$, in most cases. As a result, the prediction-mode information of the adjacent blocks can be utilized to decide whether the current PU is located at an HTU or NHTU.

### B. Gradient information-based classification

Based on the fact that the gradient information can represent some texture characteristics of video content [40], we propose to use the gradient information to decide whether the current PU belongs to a VFU. According to the content feature analysis, the luminance information of the current PU and its adjacent reconstructed blocks is simple in a VFU, so as to the searched *K*-NN template blocks. Considering that the classification criterion should be simple and fast, we use the

TABLE I
JUDGMENT ACCURACY OF $PM_C = PM_U = PM_L = PM_{UL}$ AT EACH DEPTH LEVEL WHEN $PM_U = PM_L = PM_{UL}$

| Images | Depth 0 | Depth 1 | Depth 2 | Depth 3 | Average |
|---|---|---|---|---|---|
| Bike | 67.5% | 88.2% | 86.1% | 74.9% | 79.2% |
| Fountain | 70.0% | 80.7% | 83.4% | 82.6% | 79.2% |
| Sergio | 67.4% | 78.9% | 82.2% | 82.7% | 77.8% |
| Zhengyun1 | 69.2% | 79.8% | 82.3% | 82.4% | 78.4% |
| I01 | 70.4% | 81.1% | 87.0% | 74.2% | 78.2% |
| I02 | 75.6% | 84.0% | 83.7% | 71.9% | 78.8% |
| I04 | 79.4% | 88.7% | 83.8% | 72.4% | 81.1% |
| I06 | 95.4% | 99.0% | 95.8% | 90.0% | 95.1% |

top nearest candidate block of the current PU and corresponding adjacent reconstructed blocks in $K$-NN templates to decide whether the current PU belongs to the VFU or not. As mentioned above, the current PU and its adjacent reconstructed blocks can be expressed by a multidimensional vector format $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0)$. Suppose that the searched top-nearest template set is expressed by a multidimensional vector $\mathbf{z}_1 = (\mathbf{x}_1, \mathbf{y}_1)$, and the gradients of vectors $\mathbf{z}_1, \mathbf{x}_1, \mathbf{y}_1$ are represented by $G_{\mathbf{z}_1}, G_{\mathbf{x}_1}, G_{\mathbf{y}_1}$, respectively. During the gradient calculation, the vectors $\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1$ are all compacted into column vectors, respectively. The dimension of vector $\mathbf{z}_1$ is a size summation of vector $\mathbf{x}_1$ and vector $\mathbf{y}_1$. Suppose the current PU is composed of $n \times n$ pixels, the sizes of vectors $\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1$ are $N_{x1} = n^2 \times 1, N_{y1} = (2nT + T^2) \times 1, N_{z1} = (N_{x1} + N_{y1}) \times 1$, respectively. In order to reduce the computational burden, the mean value of absolute differences at one pixel intervals in each vector is calculated as the corresponding gradient, which means that $G_{\mathbf{x}_1} = (1/N_{x1}) \times \sum_{k=1}^{N_{x1}-1} |\mathbf{x}_1(k+1) - \mathbf{x}_1(k)|, G_{\mathbf{y}_1} = (1/N_{y1}) \times \sum_{k=1}^{N_{y1}-1} |\mathbf{y}_1(k+1) - \mathbf{y}_1(k)|, G_{\mathbf{z}_1} = (1/N_{z1}) \times \sum_{k=1}^{N_{z1}-1} |\mathbf{z}_1(k+1) - \mathbf{z}_1(k)|$. As a result, we derive three scalar gradients. In the classification procedure, the current PU is considered to be located at the VFU if $|G_{\mathbf{x}_1} - G_{\mathbf{y}_1}| < G_{\mathbf{z}_1}$. Otherwise, the current PU is grouped into an HTU. For simplicity, we also define a flag, $flag'_{CB}$, to mark whether the current PU belongs to an VFU or not, given by:

$$flag'_{CB} = \begin{cases} 1 & if(|G_{\mathbf{x}_1} - G_{\mathbf{y}_1}| < G_{\mathbf{z}_1}) \\ 0 & Otherwise \end{cases} . \quad (9)$$

## V. GPR-BASED CLASSIFICATION-PREDICTION SCHEME

The Gaussian process is a powerful, non-parametric tool for learning regression, and the proposed GPR-based LF image-compression scheme considerably outperforms some other state-of-the-art techniques [30]. Therefore, in this paper, we propose a content-based LF image-compression method with Gaussian process regression. The main idea is to divide the PUs into three categories according to the content property of the LF image and design corresponding prediction schemes for different categories based on the GPR to further improve the compression efficiency and reduce the prediction complexity. This section will illustrate the prediction methods for different categories.

### A. Prediction for NHTU

An LF image is composed of numerous MIs. For this reason, a non-homogenous texture area exists in an LF image, such as the stitched area between two neighboring MIs. In such areas, the traditional TMP-based compression methods do not necessarily lead to a good prediction, since the current PU is quite different from its adjacent templates in such areas. We propose a hybrid GPR and IDC scheme to improve the prediction accuracy. The IDC scheme (also called IDC mode) is similar to inter-picture prediction, which uses the original pixels of the current PU to search an optimal matching block as its best candidate block within a searching window. The displacement vector (DV) of the current PU and the searched best candidate block is thus sent to the decoder as overhead, so that the decoder side can find the best candidate block. In the proposed hybrid prediction method, the searching window of IDC is set to the whole reconstructed or decoded blocks, as shown in Fig. 5.

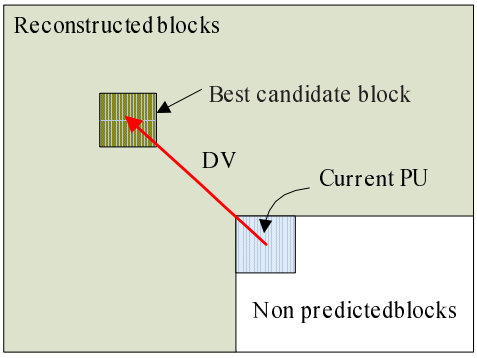



Fig. 5. Intra-displacement Compensation (IDC)

The IDC-based prediction method is implemented in the intra-frame coding framework of the HEVC coding standard through adding a new competing prediction mode other than the 35 directional prediction modes. Therefore, in the hybrid prediction method, the GPR prediction method (GPR mode), IDC prediction method (IDC mode), and intra-directional prediction mode are all used as the competing prediction modes.

The proposed hybrid prediction method explores the idea of using the IDC prediction or intra-directional prediction to find the best prediction $\widehat{\mathbf{x}}_0^{NHTU}$ of the current PU when the GPR prediction method fails based on the rate-distortion optimization (RDO) procedure, given by:

$$\begin{cases} J* = \underset{d \subset D}{\operatorname{argmin}} (\underset{p \subset P}{\operatorname{argmin}} J_m) \\ J_m = (SSE_{luma} + \omega_{chroma} \times SSE_{chroma}) \\ \quad + \lambda_{mode} \times R_{mode} \end{cases} , \quad (10)$$

where $D \in \{0, 1, 2, 3\}$ represents the set of candidate PU depths; $P$ is the set of all of the candidate-prediction modes; $SSE_{luma}$ and $SSE_{chroma}$ denote the distortion between the

current PU and its reconstructed block when the current PU is encoded with depth $d$ and mode $p$ for luma and chroma components, respectively; $\omega_{chroma}$ represents the weight parameter for the chroma component; $\lambda_{mode}$ denotes the Lagrange multiplier; and $R_{mode}$ denotes the total bits used to encode the current PU.

In other words, the proposed hybrid prediction method uses the "try all then select best" intra-mode decision method to find the best prediction mode and optimal depth for each PU.

### *B. Prediction for HTU*

In homogenous texture areas, the current PU and its adjacent reconstructed blocks have a similar texture structure, which means that they have a high correlation. For these areas, the GPR-based prediction method can achieve high compression efficiency. Given this high correlation, the current PU is likely to have prediction errors similar to its adjacent reconstructed blocks. As a result, high prediction accuracy can be achieved by using the GPR-based method for the PUs in an HTU. The prediction of the current PU in an HTU can be expressed as:

$$\widehat{\mathbf{x}}_0^{HTU} = \mathbf{K}_{HZ}(\mathbf{K}_Z + \sigma^2\mathbf{I})^{-1}\mathbf{y}_0. \tag{11}$$

We have mentioned that for the PUs in an NHTU, the "try all then select best" intra-mode decision method is used to find the optimal prediction mode and quad-tree depths. According to the content property of the HTU, the GPR-based prediction method can acquire good enough prediction accuracy. Therefore, we skip the IDC mode and only derive the optimal prediction mode and quad-tree depths among the rest 34 intra-directional prediction modes and the GPR mode to further reduce the prediction complexity for PUs in an HTU. Moreover, the LF image consists of numerous MIs, which results in that the texture-homogeneous areas hardly prevail. Thus, the coding unit size $64 \times 64$ is seldom selected as the best block size. Therefore, for an HTU, only four coding block sizes, ranging from $32 \times 32$ to $4 \times 4$, are chosen in the mode-decision process to further reduce the computational burden.

### *C. Prediction for VFU*

A VFU is a special case of an HTU. The luminance information of the PUs in visually flat regions and their adjacent reconstructed blocks is simple. Applying a GPR-based prediction method would be overkill since similar reconstruction quality could be achieved by a simpler (and faster) prediction method for such relatively simple structures. Therefore, we propose a simplified GPR prediction method (also referred to as SGPR mode) for the PUs in a VFU.

A Gaussian process is a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of these random variables has a multivariate Gaussian distribution. This means that the distribution of a finite vector subset can be simplified by a multivariate Gaussian distribution. Based on this distribution's properties, the noise-free prediction of the current PU can be rewritten as:

$$\widehat{\mathbf{x}}_0^{VFU} = \widetilde{\mathbf{x}}_0 + \mathbf{K}_{HZ}\mathbf{K}_Z^{-1}(\mathbf{y}_0 - \widetilde{\mathbf{y}}_0), \tag{12}$$

where $\widetilde{\mathbf{x}}_0$ and $\widetilde{\mathbf{y}}_0$ express the linear predictions of $\mathbf{x}_0$ and $\mathbf{y}_0$ from the prediction supports set $\mathbf{X}$ and $\mathbf{Y}$, drawn from:

$$\widetilde{\mathbf{x}}_0 = \sum_{k=1}^{K} \mathbf{w}_k(\mathbf{y}_0)\mathbf{x}_k \tag{13}$$

$$\widetilde{\mathbf{y}}_0 = \sum_{k=1}^{K} \mathbf{w}_k(\mathbf{y}_0)\mathbf{y}_k, \tag{14}$$

where $\mathbf{w}_k(\mathbf{y}_0)$ is the weight vector.

From Eq. (12), we see that the prediction of the current PU in a VFU has two parts. The first is a linear prediction of vector $\mathbf{x}_0$, and the second can be seen as a correction vector representing that the unpredictable part (prediction errors) of $\mathbf{y}_0$ is transformed into subspace $\mathbf{x}_0$ [37]. As mentioned above, the flat regions have a simple luminance information, where the unpredictable part of $\mathbf{y}_0$ has a negligible effect on the prediction accuracy. This means that a linear prediction can be directly used to predict the PUs in a VFU and the correction vector can be neglected. The prediction of the current PU in a VFU can be derived as:

$$\widehat{\mathbf{x}}_0^{VFU} = \sum_{k=1}^{K} \mathbf{w}_k(\mathbf{y}_0)\mathbf{x}_k. \tag{15}$$

To predict the current PU with Eq. (15), we must acquire the weight vector $\mathbf{w}_k(\mathbf{y}_0)$. An acquisition method [25] [38] has been proposed to calculate such weight vectors, in which the adjacent template $\mathbf{y}_0$ and its $K$-NN template vector set $\mathbf{Y} = \{\mathbf{y}_k | k = 1, 2, ..., K\}$ are used by solving an optimization problem given by:

$$\min_{\mathbf{w}} \|\mathbf{y}_0 - \sum_{k=1}^{K} \mathbf{w}_k(\mathbf{y}_0)\mathbf{y}_k\| \ subject\ to \ \sum_k \mathbf{w}_k(\mathbf{y}_0) = 1. \tag{16}$$

This method aims to derive the weight vector by calculating the best-fitting of the $K$-NN template vectors $\{\mathbf{y}_k | k = 1, 2, ..., K\}$ to the vector $\mathbf{y}_0$, where the relevance between the vector $\mathbf{y}_0$ and its $K$-NN template vector set is important in deriving an accurate weight vector. Nevertheless, this method does not consider such relevance, which may reduce the reliability when the correlation between vector $\mathbf{y}_0$ and its $K$-NN template vectors is rather low.

To this end, a locally correction-weight-based method is proposed to derive the weight coefficient $\mathbf{w}_k(\mathbf{y}_0)$, where we assign a correction weight to each pixel in vector set $\{\mathbf{y}_k | k = 1, 2, ..., K\}$. The correction weight is greater for such pixels that are closer to the corresponding pixels in vector $\mathbf{y}_0$. As a consequence, calculating $\mathbf{w}_k(\mathbf{y}_0)$ can be rewritten as a problem of constrained minimization given by:

$$\min_{\mathbf{w}} \sum_{k=1}^{N} \mathcal{R}[h_{\mathbf{w}}(\mathbf{y}_k^n) - \mathbf{y}_0^n]^2 \ subject\ to \ \sum_k \mathbf{w}_k(\mathbf{y}_0) = 1, \tag{17}$$

where $\mathcal{R}$ denotes the correction weight vector, $N$ represents the total number of pixels in vector $\mathbf{y}_0$, and $h_{\mathbf{w}}(\mathbf{y}_k^n)$ is given by $h_{\mathbf{w}}(\mathbf{y}_k^n) = \sum_{k=1}^{K} \mathbf{w}_k(\mathbf{y}_0)\mathbf{y}_k^n$.

In Eq.(17), the vector $\mathcal{R}$ is utilized to represent the relevance between vector $\mathbf{y}_0$ and corresponding vector in $\{\mathbf{y}_k | k =$

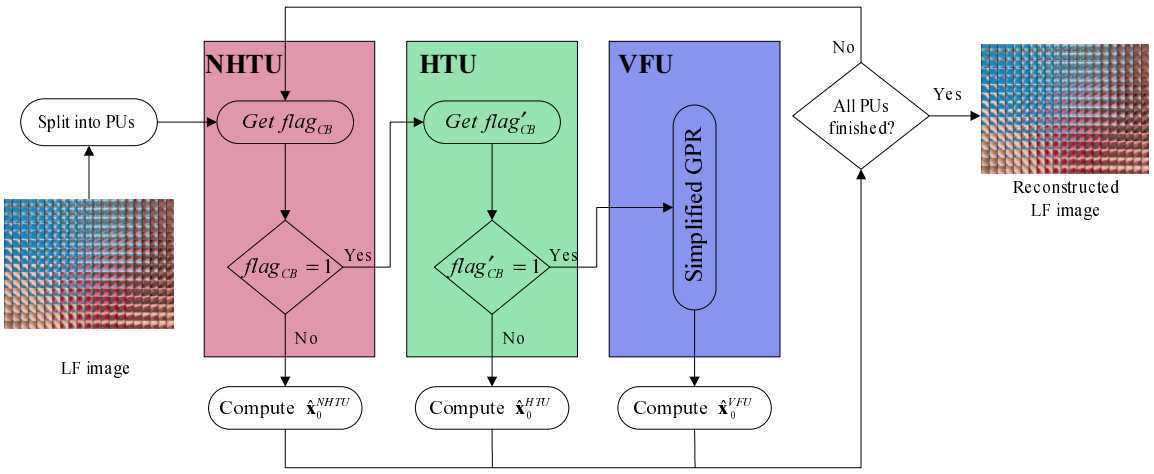



Fig. 6. Flow graph explaining layered GPR based prediction algorithm

$1, 2, ..., K\}$. To derive the vector $\mathcal{R}$, a kernel function is used to weight nearby pixels more heavily than other pixels in $\{\mathbf{y}_k|k = 1, 2, ..., K\}$. Since the Gaussian kernel is a commonly used kernel function, therefore, we use Gaussian kernel to calculate the correlation of vector $\mathbf{y}_0$ and corresponding vector in $\{\mathbf{y}_k|k = 1, 2, ..., K\}$, which is given by:

$$\mathcal{R}^{(n,k)} = \exp(-\frac{(\mathbf{y}_k^n - \mathbf{y}_0^n)^2}{2\tau^2}), \text{where } \tau = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{y}_k^n - \mathbf{y}_0^n)^2. \tag{18}$$

For simplicity, we rewrite the locally correction-weighted-based scheme in vector form as:

$$\min_{\mathbf{w}}(\mathbf{y}\mathbf{w} - \mathbf{y}_0)^{\mathrm{T}}\mathcal{R}(\mathbf{y}\mathbf{w} - \mathbf{y}_0) \ subject\ to \ \sum_k \mathbf{w}_k = 1. \tag{19}$$

After deriving the $\mathbf{w}_k(\mathbf{y}_0)$, the current PU can be predict by using Eq. (15).

### D. Flow graph of the proposed method

The proposed compression method is described by the flow graph in Fig. 6. In the proposed LF image compression procedure, the whole LF image is grouped into three categories according to the content property of LF image. For PUs located at non-homogenous texture areas, a hybrid method based on GPR and IDC is adopted to improve the prediction accuracy. GPR based prediction method is applied to predict the PUs in homogenous texture areas. Regarding the PUs in visually flat areas, we put forward a simplified GPR-based prediction method to reduce the computational complexity, where a locally correction-weight-based method is used. Note that the entire framework also applies to the chrominance block prediction.

## VI. Experimental results

### A. Simulation setup

To validate the effectiveness of the proposed method, ten LF test images are used in the test set, including eight unfocused LF test images used for the ICME 2016 grand challenge in light field image compression[41] and two focused LF test images provided by Dr. T. Georgiev [42]. The resolution of the unfocused LF test images is $7728 \times 5368$. The size of each MI within an unfocused LF test image is $15 \times 15$. The resolution of the focused LF test images is $7240 \times 5432$, and the size of each MI is $75 \times 75$. The LF images in the test set are all transformed into the YUV 4:2:0 format.

The HEVC Test Model (HM) reference software version 15 is modified for the proposed compression method. The coding configurations are set as "All Intra" [43]. Four quantization parameters, 22, 27, 32, and 37, are applied. The proposed GPR-based classification prediction method (referred to as Proposed [1]) is compared with five prediction schemes: the original HEVC (referred to as HEVC), the self-similarity compensated prediction method [32] (referred to as HEVC-SS), the variants of the high-order intra-block prediction method (HEVC-HOP) [27] (referred to as HEVC-HOP-A (6 degrees of freedom (DoF)), HEVC-HOP-P (8 DoF), and HEVC-HOP-B (8 DoF)). The Bjontegaard-Delta rate (BD-rate) is used to measure the performance, and we choose Y-PSNR between the original LF image and decompressed LF image [41] as the objective quality metric.

Since GPR-based prediction is time-consuming, scaling cubically with the size of the training data [44], the template thickness $T$ in Fig. 3(a) is set to 4 (equal to the smallest PU size) to construct the $K$-NN template vectors $\mathbf{Y}$ within the GPR prediction procedure to reduce the computational complexity and ensure prediction accuracy. Since the causal searching window size is set to 128 in both HEVC-SS and HEVC-HOP [27], in the proposed method, the dimensions of the searching windows is also set to 128, which means $V = 128, H = 128$, as shown in Fig. 3(b). The number of nearest-neighbor template vectors $K$ is set to 5 in consideration of computation complexity. As mentioned above, the GPR-based prediction method is integrated into the HEVC coding standard by replacing one of the 35 intra-directional prediction modes. In the proposed method, the least statistically used mode of the

[1]The executable files are available online at https://github.com/yangyangshu

35 intra-prediction modes is replaced by the GPR mode, and a pre-coding procedure of the light field image test set using HEVC is needed to find the least statistically used mode.

### B. Experimental results

The main goal of the proposed LF image-compression scheme is to further improve the coding efficiency according to the content feature as well as to reduce the prediction complexity. In this subsection, we verify the effectiveness of the proposed method from five aspects.

*1) Comparison of RD Performance between the proposed method and state-of-the-art methods*

The RD performance comparison of the proposed compression method with state-of-the-art methods under objective quality metric Y-PSNR is shown in Table II, from which we see that the proposed method is superior to the state-of-the-art methods in most cases both for the focused LF test image set and unfocused LF test image set. An average gain of up to 1.64 dB is achieved by the proposed method over the HEVC intra-standard for all of the LF test image sets. Compared to HEVC-SS, around 0.45 dB average PSNR gain can be achieved for all the LF test image set. When compared to the variants of HEVC-HOP, we find from Table II that around 0.30 dB, 0.26 dB, and 0.24 dB average PSNR gains can be achieved by the proposed method over HEVC-HOP-A (6 DoF), HEVC-HOP-P (8 DoF), and HEVC-HOP-B (8 DoF), respectively. The proposed method can achieve a higher coding efficiency than HEVC-SS because it integrates the IDC and GPR modes in HEVC, which can lead to a high prediction accuracy for texture areas. For HEVC-HOP, the best case is when eight DoFs are used, which allows translations, rotations, scaling, shearing, and perspective changes. Even though HEVC-HOP can obtain high prediction accuracy, extra information that must be encoded reduces the overall coding efficiency. Specifically, the RD gains of HEVC-HOP-P (8 DoF) and HEVC-HOP-B (8 DoF) are respectively 0.03 dB and 0.01dB higher than the proposed method for unfocused LF test image I05, which can be seen from Table II. The reason is closely connected to the contents of unfocused LF test image I05. The scene of image I05 is a Vespa scooter. The foreground texture information is monotone and the background areas belong to the flat regions in most cases. Therefore, for image I05, the PUs are almost divided into an HTU or a VFU, which reduces the prediction accuracy. As a result, the RD gains of HEVC-HOP-P (8 DoF) and HEVC-HOP-B (8 DoF) are a little higher than those of the proposed method. However, the coding efficiency of the proposed method is still higher than that of HEVC-HOP-A (6 DoF) and HEVC-SS.

*2) Comparison of RD performance between the integrated prediction methods*

The proposed LF image-compression method tries to integrate the IDC prediction method (IDC mode), GPR prediction method (GPR mode), and SGPR prediction method (SGPR mode) into the HEVC coding standard to improve the compression efficiency and reduce the complexity of PU prediction. To further verify the effectiveness of the proposed method, we provide an RD performance comparison of the proposed method with IDC mode, SGPR mode, and GPR mode under objective quality metric Y-PSNR, which is shown in Table III. From Table III, we find that the proposed method can achieve a better average compression efficiency than the other prediction modes. Around 0.68 dB, 0.29 dB, and 0.06 dB average PSNR gains can be obtained by the proposed method over IDC mode, SGPR mode, and GPR mode, respectively. This is mainly because the proposed method uses the "try all then select best" intra-mode decision method to find the best prediction mode and optimal depth for each PU in an NHTU and uses the GPR mode to predict the PUs in an HTU, which leads to higher prediction accuracy than the other prediction modes. From Table III, we also observe that the Y-PSNR gains of some LF images are higher for GPR mode compared to the proposed method. For example, the RD gains of GPR mode are 0.02 dB, 0.04 dB, and 0.03 dB higher than the proposed method for unfocused LF test images I03, I08, and I10, respectively. This is mainly because the classification is not accurate. Some PUs that belong to the HTU are divided into the VFU. Since the performance of GPR is superior to that of SGPR, the compression efficiency is reduced. Nevertheless, the effect on the prediction accuracy is negligible. Regarding the other prediction modes, the proposed method can always achieve a higher compression efficiency for all the tested LF images.

The results in Fig. 7 and Fig. 8 further confirm that the proposed scheme can achieve a better RD performance than HEVC intra, IDC mode. It also performs better than the SGPR mode and GPR mode for all tested QPs. The results are consistent for the tested images.

*3) Objective performance under JPEG Pleno LF coding test conditions*

As mentioned above, the JPEG Pleno aims to provide a standard framework for LF coding. By contributing towards the JPEG Pleno effort, this paper further provides the objective performance results, including $PSNR_Y$, $PSNR_{Cb}$, $PSNR_{Cr}$ and $SSIM_Y$ [48], by using the proposed method. The $PSNR_Y$, $PSNR_{Cb}$ and $PSNR_{Cr}$ are computed by averaging the corresponding PSNRs of all the extracted view images from decoded LF image. The $SSIM_Y$ is acquired by averaging the luminance component (Y) SSIM values of all the extracted view images from decoded LF image. In this procedure, the extracted view images are transformed from RGB to YUV colorspace, and the corresponding code is provided in [48]. Four lenslet LF images included in the JPEG Pleno test set are used, namely I01, I02, I04 and I09. For original lenslet LF image, $15 \times 15$ view images can be extracted. In our experiment, only the central $13 \times 13$ views are used by removing the dark views. In order to illustrate the performance of the proposed method and facilitate comparative analysis, four QPs, notably $B_1 = 37, B_2 = 32, B_3 = 27, B_4 = 22$, are used. Table IV gives the objective performance results by using the proposed method under JPEG Pleno LF coding test conditions across all the used QPs. The correspondences between bits per pixel (bpp) defined in [48] and the four QPs are given in Table V. From Table IV, we see that the proposed method can achieve a good objective qualities in terms of PSNR and SSIM. Specifically, the average $SSIM_Y$ is more

TABLE II
Y-RATE-DISTORTION GAINS OF PROPOSED METHOD AND STATE-OF-THE-ART METHODS OVER HEVC

| Images | HEVC-SS | | HEVC-HOP-A(6 DoF) | | HEVC-HOP-P(8 DoF) | | HEVC-HOP-B(8 DoF) | | Proposed | |
|---|---|---|---|---|---|---|---|---|---|---|
| | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) |
| Laura | 2.26 | -30.35 | 2.41 | -33.11 | 2.53 | -35.13 | 2.58 | -35.97 | 2.92 | -35.63 |
| Seagull | 2.81 | -42.78 | 3.03 | -47.67 | 3.12 | -49.60 | 3.24 | -51.99 | 3.88 | -51.02 |
| I03 | 0.17 | -4.10 | 0.20 | -4.85 | 0.20 | -4.87 | 0.20 | -4.77 | 0.25 | -5.79 |
| I05 | 0.89 | -28.73 | 0.99 | -32.63 | 1.03 | -34.03 | 1.01 | -33.62 | 1.00 | -28.91 |
| I07 | 0.48 | -13.73 | 0.74 | -21.78 | 0.73 | -21.51 | 0.77 | -22.66 | 0.83 | -21.83 |
| I08 | 0.66 | -22.95 | 0.72 | -26.47 | 0.72 | -28.14 | 0.71 | -25.57 | 0.97 | -33.68 |
| I09 | 1.49 | -30.72 | 1.66 | -35.11 | 1.69 | -36.03 | 1.69 | -36.00 | 2.09 | -37.90 |
| I10 | 0.22 | -8.10 | 0.27 | -10.11 | 0.28 | -10.31 | 0.28 | -10.46 | 0.32 | -10.99 |
| I11 | 1.49 | -42.84 | 1.76 | -54.03 | 1.80 | -55.14 | 1.81 | -55.46 | 2.44 | -57.19 |
| I12 | 1.42 | -41.35 | 1.66 | -50.63 | 1.69 | -51.77 | 1.69 | -51.82 | 1.70 | -43.71 |
| Average | 1.19 | -26.57 | 1.34 | -31.64 | 1.38 | -32.65 | 1.40 | -32.83 | 1.64 | -32.67 |

TABLE III
Y-RATE-DISTORTION GAINS OF THE PROPOSED METHOD, IDC MODE, SGPR MODE, AND GPR MODE OVER HEVC

| Images | IDC mode | | SGPR mode | | GPR mode | | Proposed | |
|---|---|---|---|---|---|---|---|---|
| | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) | BD-PSNR (dB) | BD-Rate (%) |
| Laura | 1.61 | -20.92 | 2.71 | -32.79 | 2.86 | -33.91 | 2.92 | -35.63 |
| Seagull | 2.32 | -34.17 | 3.54 | -46.63 | 3.52 | -45.55 | 3.88 | -51.02 |
| I03 | 0.14 | -3.28 | 0.18 | -4.18 | 0.27 | -6.17 | 0.25 | -5.79 |
| I05 | 0.56 | -16.81 | 0.86 | -25.18 | 0.99 | -28.48 | 1.00 | -28.91 |
| I07 | 0.66 | -17.35 | 0.64 | -16.98 | 0.71 | -19.05 | 0.83 | -21.83 |
| I08 | 0.40 | -13.49 | 0.98 | -36.21 | 1.01 | -37.48 | 0.97 | -33.68 |
| I09 | 1.26 | -24.51 | 1.95 | -35.58 | 2.08 | -37.37 | 2.09 | -37.90 |
| I10 | 0.18 | -6.28 | 0.28 | -9.76 | 0.35 | -12.08 | 0.32 | -10.99 |
| I11 | 1.48 | -37.41 | 2.39 | -56.07 | 2.44 | -56.75 | 2.44 | -57.19 |
| I12 | 1.00 | -27.55 | 1.51 | -39.04 | 1.56 | -40.07 | 1.70 | -43.71 |
| Average | 0.96 | -20.18 | 1.35 | -30.24 | 1.58 | -31.69 | 1.64 | -32.67 |

than 0.94 across all the QPs.

*4) Comparison of computational complexity*

Table VI gives the execution time ratios of the proposed method and the state-of-the-art methods to HEVC for LF image *Vesper* (I05) when $QP = 32$. From this, we see that the proposed method needs less execution time than HEVC-SS and the variants of HEVC-HOP. For example, the proposed method can save 83.0%, 88.8%, 97.5%, and 96.7% of the coding time compared to HEVC-SS, HEVC-HOP-A (6 DoF), HEVC-HOP-P (8 DoF), and HEVC-HOP-B (8 DoF), respectively. There are two main reasons. One is that the PUs in the proposed method are divided into three categories. By using a simpler and faster prediction method for PUs in a VFU, the computational complexity is reduced on the encoder side. Also, in the GPR method, we only use the specified vertical and horizontal searching window, shown in Fig. 3(b), to search for the prediction support vector $\mathbf{X}$. Regarding the decoder side, the proposed method needs slightly more decoding time than HEVC-SS and the variants of HEVC-HOP, as can be seen from Table VI. This is mainly because the decoder side must perform the same prediction procedure as the encoder side in the proposed method and the matrix operations of GPR are time-consuming.

Table VII gives the average prediction mode usage across the four QPs for all of the test LF images. From this, we can observe that the SGPR mode is used more frequently than the IDC and GPR modes, other than the angular prediction mode. This means that more PUs are divided into VFUs, and then a simpler SGPR mode is utilized to predict these PUs. This also demonstrates that the computational complexity of the proposed method is lower than that of the compared state-of-the-art methods. However, even though the SGPR mode is used more frequently than the IDC and GPR modes, a considerable part of the PUs are predicted using the GPR mode. Since the computational complexity of GPR is high, it still needs more coding time than HEVC, especially on the decoder side.

TABLE IV
OBJECTIVE PERFORMANCE RESULTS BY USING PROPOSED METHOD UNDER JPEG PLENO LF CODING TEST CONDITIONS ACROSS ALL USED QPS

| Images | PSNR-Y (dB) | | | | PSNR-Cb (dB) | | | | PSNR-Cr (dB) | | | | PSNR-YCbCr (dB) | | | | SSIM-Y | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ |
| I01 | 36.08 | 39.23 | 42.63 | 46.27 | 43.11 | 44.0 | 45.51 | 47.64 | 44.02 | 44.87 | 46.14 | 48.0 | 37.95 | 40.53 | 43.43 | 46.66 | 0.948 | 0.968 | 0.982 | 0.99 |
| I02 | 37.62 | 40.74 | 43.91 | 47.06 | 44.21 | 44.95 | 46.45 | 48.54 | 45.24 | 45.86 | 47.04 | 48.75 | 39.39 | 41.9 | 44.62 | 47.46 | 0.952 | 0.972 | 0.984 | 0.99 |
| I04 | 36.72 | 39.81 | 42.99 | 46.26 | 44.67 | 45.13 | 46.34 | 48.25 | 45.06 | 45.59 | 46.41 | 48.22 | 38.76 | 41.20 | 43.84 | 46.76 | 0.96 | 0.976 | 0.985 | 0.99 |
| I09 | 35.5 | 38.71 | 42.31 | 46.38 | 42.87 | 43.64 | 44.87 | 47.07 | 44.0 | 44.69 | 45.83 | 47.39 | 37.49 | 40.07 | 43.07 | 46.59 | 0.933 | 0.956 | 0.976 | 0.99 |
| Avg. | 36.48 | 39.62 | 42.96 | 46.49 | 43.72 | 44.43 | 45.79 | 47.88 | 44.58 | 45.25 | 46.36 | 48.09 | 38.38 | 40.93 | 43.74 | 46.87 | 0.948 | 0.968 | 0.982 | 0.99 |

TABLE VIII
ENCODING AND DECODING TIME RATIO TO HEVC OF IDC MODE, GPR MODE, SGPR MODE, AND THE PROPOSED METHOD

| Images | IDC mode | | SGPR mode | | GPR mode | | Proposed | |
|---|---|---|---|---|---|---|---|---|
| | Encoder side | Decoder side | Encoder side | Decoder side | Encoder side | Decoder side | Encoder side | Decoder side |
| Laura | 41.45 | 0.81 | 4.07 | 14.02 | 21.13 | 98.35 | 55.05 | 38.78 |
| Seagull | 32.35 | 0.86 | 3.94 | 17.52 | 30.09 | 134.17 | 38.43 | 37.32 |
| I03 | 5.26 | 0.95 | 2.66 | 5.95 | 20.12 | 58.15 | 9.60 | 10.68 |
| I05 | 3.79 | 0.80 | 3.62 | 14.37 | 27.72 | 115.96 | 10.49 | 22.53 |
| I07 | 4.32 | 1.14 | 2.95 | 8.78 | 20.68 | 65.99 | 10.87 | 14.49 |
| I08 | 3.40 | 1.13 | 4.11 | 17.55 | 31.96 | 169.13 | 7.99 | 22.93 |
| I09 | 21.35 | 0.99 | 3.73 | 12.93 | 26.90 | 133.02 | 26.33 | 24.44 |
| I10 | 2.80 | 1.20 | 2.63 | 3.13 | 16.85 | 35.39 | 6.34 | 7.86 |
| I11 | 14.29 | 0.97 | 4.20 | 13.96 | 29.64 | 154.68 | 20.15 | 33.35 |
| I12 | 9.69 | 0.88 | 4.06 | 18.71 | 27.22 | 165.58 | 19.29 | 37.51 |
| Average | 13.87 | 0.97 | 3.60 | 12.69 | 25.23 | 113.04 | 20.45 | 24.89 |

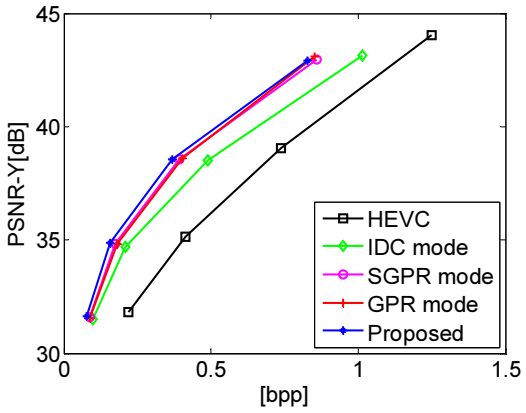



Fig. 7. Rate-distortion curves for Seagull

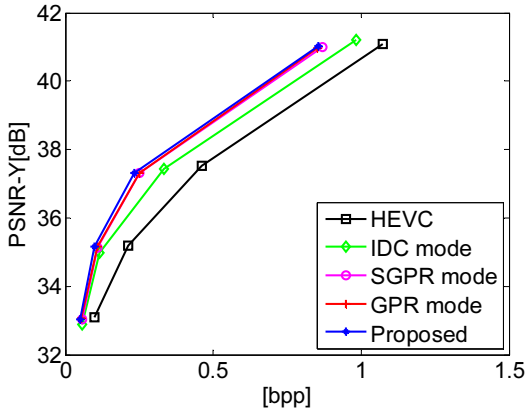



Fig. 8. Rate-distortion curves for LF image I12

In order to compare the computation complexity of the IDC prediction method, GPR prediction method, SGPR prediction method and the proposed method, Table VIII shows the execution time ratios of the four methods to HEVC intra-standard, on both the encoder and decoder sides. As mentioned above, the IDC prediction method integrates the IDC mode into HEVC intra standard as a new competing prediction mode. The GPR prediction method and the SGPR prediction method

TABLE V
THE CORRESPONDENCE BETWEEN BPP AND THE FOUR QPS

| Image | QP=37 | QP=32 | QP=27 | QP=22 |
|---|---|---|---|---|
| I01 | 0.10 | 0.22 | 0.44 | 0.90 |
| I02 | 0.09 | 0.17 | 0.35 | 0.70 |
| I04 | 0.10 | 0.20 | 0.40 | 0.80 |
| I09 | 0.10 | 0.21 | 0.47 | 1.06 |

TABLE VI
ENCODING AND DECODING TIME RATIO TO HEVC FOR VESPER (QP=32)

| Encoder side | HEVC -SS | HEVC-HOP-A | HEVC-HOP-P | HEVC-HOP-B | Proposed |
|---|---|---|---|---|---|
| vs HEVC | 64.67 | 98 | 442.33 | 338.33 | 11.01 |
| Decoder side | HEVC -SS | HEVC-HOP-A | HEVC-HOP-P | HEVC-HOP-B | Proposed |
| vs HEVC | 19.74 | 17.96 | 17.64 | 16.8 | 22.92 |

integrate the GPR mode and SGPR mode into intra standard by replacing the least statistically used mode of the 35 intra prediction modes, respectively. From Table VIII, we observe that the GPR-based method requires the most execution time on both sides. This is mainly because the calculation of matrix $\mathbf{K}_{HZ}$ and $\mathbf{K}_Z^{-1}$ is time-consuming for all of the PUs. Since the decoder side must execute the same prediction procedure as the encoder side, the GPR-based method needs 113.04 times the average execution time of the HEVC intra-standard on the decoder side. To reduce the computational complexity, we propose a content-based LF image-classification algorithm with GPR. Table VIII gives the time ratios of the proposed method to HEVC on both the encoder and decoder sides. From Table VIII, we see that the execution time ratios of the proposed method to HEVC are 20.45 and 24.89 on the encoder and decoder sides, respectively. The proposed method achieves 19.0% and 78.0% of the average coding-time saving compared to the GPR-based method on the encoder side and decoder side, respectively. There are two main reasons. One is that more PUs are divided into VFUs. By using a simpler and faster prediction method for PUs in VFUs, the computational complexity is reduced on both the encoder and decoder sides. The other is that the IDC mode is selected as the optimal prediction mode by many PUs. The IDC mode costs much less than the GPR mode, especially on the decoder side. Therefore, the execution time ratios of the proposed method to GPR on the decoder side are less than that on the encoder side.

*5) Comparison of visual quality of rendered view images*

Since an LF image contains both spatial and angular information of 3D scene, VIs could be rendered from LF contents. In the compression process, the prediction accuracy is higher, the visual quality of rendered VIs from the decoded LF image is better. Therefore, we can use the visual quality of the rendered VI to verify the effectiveness of the proposed method, especially for the areas located in an NHTU.

In the proposed method, we use the basic rendering scheme [46] to extract the viewpoint image of focused LF test images,

TABLE VII
AVERAGE PREDICTION MODE USAGE ACROSS THE FOUR QPS

| Images | DC, Planar directional | Proposed prediction modes | | |
|---|---|---|---|---|
| | | IDC mode | SGPR mode | GPR mode |
| Laura | 33.88% | 9.04% | 31.37% | 25.71% |
| Seagull | 28.86% | 10.54% | 38.45% | 22.15% |
| I03 | 63.89% | 14.36% | 11.34% | 8.16% |
| I05 | 45.21% | 19.72% | 17.89% | 17.19% |
| I07 | 54.76% | 24.22% | 10.95% | 10.07% |
| I08 | 44.56% | 16.62% | 23.13% | 15.69% |
| I09 | 42.41% | 22.13% | 17.80% | 17.67% |
| I10 | 66.94% | 12.20% | 13.28% | 7.71% |
| I11 | 26.04% | 16.52% | 33.53% | 23.94% |
| I12 | 34.36% | 18.40% | 25.17% | 22.09% |
| Average | 44.09% | 16.38% | 22.29% | 17.04% |

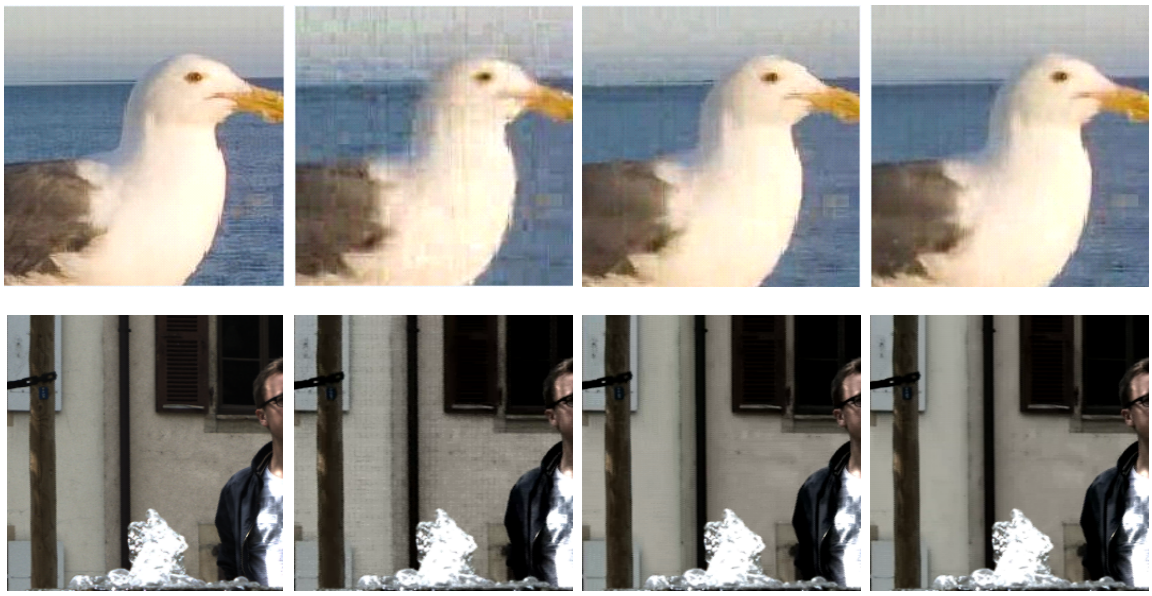
where a block is extracted from each MI first, and then all of the extracted blocks are stitched together to construct the viewpoint image. In this paper, an $8 \times 8$ block is extracted from each MI to construct the viewpoint image. As for the unfocused LF test image, the Light Field Toolbox v0.4 software (LFT software) [47] is used to obtain a specific viewpoint image.

As shown in Fig. 9, the visual quality of the view image rendered from the decoded LF image derived by using the proposed method is superior to the HEVC coding standard and the IDC prediction method, especially in some texture regions. There are two main reasons. One is that by using a classification-prediction method, the proposed scheme can achieve better prediction accuracy in the non-homogenous texture areas than the directional-prediction and IDC modes. The other is that the proposed method can effectively keep the detail information of MIs in the prediction process by dividing the LF image into three categories and using corresponding prediction methods for PUs in different categories.

## VII. CONCLUSION

In this paper, we propose a content based LF image compression method with Gaussian process regression to effectively compress the LF image. The LF image is divided into three categories adaptively according to the content analysis. For PUs in an NHTU, a hybrid GPR and IDC method is proposed in the prediction procedure to improve the prediction accuracy, which explores the idea of using the IDC scheme or intra directional prediction to find the best prediction of the current PU when GPR prediction method fails based on the RDO procedure. For PUs in an HTU, we propose to use the GPR based method to ensure the coding efficiency for homogenous texture area. For PUs in a VFU, we propose to directly use a SGPR prediction method to accelerate the prediction procedure.

The experimental results demonstrate that the proposed method can compress the light field image efficiency. It outperforms the HEVC intra standard with 1.64 dB average quality improvements for all the LF test images with Y-PSNR



(a) (b) (c) (d)

Fig. 9. Visual rendering views from the decoded LF images at a similar bit-rate: (a) original image; (b) HEVC intra-standard; (c) IDC prediction method; (d) proposed prediction method. The bit-rate for Seagull is 0.063 bpp, and the bit-rate for LF I09 is 0.102 bpp.

as the objective quality metric. Regarding to the computation complexity, the proposed method can save around 96.7% coding time compared to HEVC-HOP-B(8 DoF). And 19.0% and 78.0% average coding time can be saved compared with the GPR based prediction method in encoder side and decoder side, respectively.

Future work will include more effective classification mechanism and prediction method to further improve the coding efficiency, especially for some texture and edge regions.

## REFERENCES


[1] G. Wu, B. Masia, A. Jarabo, Y. Zhang et al, "Light Field Image Processing: An Overview", *IEEE Journal of Selected Topics in Signal Processing*, vol.11, No. 7, 926-954, 2017.

[2] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*. Cambridge, MA, USA: MIT Press, pp. 3-20, 1991.

[3] M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH*, pp. 31-42, 1996.

[4] C. Michael, G. Steven, J., S. Richard, G. Radek, and S. Rick, "The lumigraph," *SIGGRAPH*, 1996.

[5] T. T. Wong,C. W. Fu,P. A. Heng, "The plenoptic illumination function", *IEEE Transactions on Multimedia*, vol. 4, no.3, pp. 361-371, 2002.

[6] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, pp. 46-55, Aug. 2006.

[7] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, and J. C. J. Fernndez, "Light field geometry of a standard plenoptic camera," *Optics Express*, vol. 22, no. 22, pp. 26659-26673, Nov. 2014.

[8] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," *Proc. IEEE Inter. Conf. Computational Photography*, 2009, pp. 1-8.

[9] T. Ebrahimi, JPEG PLENO Abstract and Executive Summary, ISO/IEC JTC 1/SC 29/WG1 N6922, Sydney, Australia, 2015.

[10] "Working Draft 0.1 of TR: Technical Report on Immersive Media", ISO/IEC JTC1/SC29/WG11/N16718, Geneva, Jan. 2017.

[11] S. Zhu, M. Li, C. Chen, "Cross-Space Distortion Directed Color Image Compression," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp.525-538, 2018.

[12] S. Li, M. Xu, Y. Ren, Z. L. Wang, "Closed-Form Optimization on Saliency-Guided Image Compression for HEVC-MSP," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 155-170, 2018.

[13] C. Deng, W. Lin, J. Cai, "Content-Based Image Compression for Arbitrary-Resolution Display Devices," *IEEE Transactions on Multimedia*, vol.14, no. 4, 1127-1139, 2012.

[14] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii-xxxiv,1992.

[15] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, 2012.

[16] I. Viola, Martin Rerabek et al, "Comparison and Evaluation of Light Field Image Coding Approaches," *IEEE Journal of Selected Topics in Signal Processing*, vol.11, no. 7, 1092-1106, 2017.

[17] 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*, vol.11, no. 7, 1107-1119, 2017.

[18] F. Dai, J. Zhang, Y. Ma and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4733-4737, 2015.

[19] D. Liu, L. Wang, L. Li, ZhiweiXiong, Feng Wu and Wenjun Zeng, "Pseudo-sequence-based light field image compression," *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1-4, 2016.

[20] G. Wang, W. Xiang, M. Pickering and C. W. Chen, "Light Field Multi-View Video Coding With Two-Directional Parallel Inter-View Prediction," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5104-5117, Nov. 2016.

[21] P. Helin, P. Astola, B. Rao and I. Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," *2016 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1-4, 2016.

[22] L. Li, Z. Li, B. Li, D. Liu and H. Li, "Pseudo Sequence Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," *2017 Data Compression Conference (DCC)*, pp. 131-140, 2017.

[23] X. Jiang, M. Le Pendu, C. Guillemot, "Light field compression using depth image based view synthesis," *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 19-24, 2017.

[24] X. Huang, P. An, L. Shen, K.Li, "Light Field Image Compression Scheme Based on MVD Coding Standard," *Pacific-Rim Conference on Multimedia 2017*, pp. 1-5, 2017.

[25] L. F. R. Lucas, C. Conti, P. Nunes, L. D. Soares, N. M. M. Rodrigues, C. L. Pagliari, E. A. B. da Silva, S. M. M. de Faria, "Locally linear embedding-based prediction for 3D holoscopic image coding using

HEVC," *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*., pp. 11,15,1-5, 2014.
[26] 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)*, pp. 1-4, 2016.
[27] R. J. S. Monteiro, P. J. L. Nunes, N. M. M. Rodrigues et al, "Light Field Image Coding using High Order Intra Block Prediction," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1120-1131, 2017.
[28] D. Liu, P. An, R. Ma, L. Shen, "Disparity compensa-tion based 3D holoscopic image coding using HEVC," *2015 IEEE China Summit & Int. Conf. Signal and Information Processing (ChinaSIP)*, pp. 201-205, 2015.
[29] D. Liu, P. An, R. Ma, C. Yang, L. Shen, K. Li, "Three-dimensional holoscopic image coding scheme using high-efficiency video coding with kernel-based minimum mean-square-error estimation," *J. Electron. Imaging*, vol. 25, no. 4, pp. 043015-1-043015-9, 2016.
[30] D. Liu, P. An, R. Ma, C. Yang, L. Shen, "3D holoscopic image coding scheme using HEVC with Gaussian process regression," *Signal Processing: Image Communication*, vol. 47, pp. 438-451, 2016.
[31] 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*, vol. 26, no. 1, pp. 50-62, 2016.
[32] C. Conti, L. D. Soares, and P. Nunes, "HEVC-based 3D holoscopicvideocoding using self-similarity compensated prediction," *Signal Processing: Image Communication*, vol. 42, pp. 59-78, 2016.
[33] C. Conti, P. Nunes and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1-4, 2016.
[34] Y. Li, M. Sjostrom, R. Olsson and U. Jennehag, "Coding of Focused Plenoptic Contents by Displacement Intra Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1308-1319, 2016.
[35] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 675-684, Sep. 2016.
[36] C. E. Rassmussen and C. K. I.Williams, Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press, 2006.
[37] J. Koloda, A.M. Peinado, V. Sanchez, "Kernel-Based MMSE Multimedia Signal Reconstruction and Its Application to Spatial Error Concealment," *IEEE Transactions on Multimedia*, vol.16, no.6, pp.1729-1738, Oct. 2014.
[38] D. Liu, P. An, C. Yang, R. Ma,L. Shen, "Coding of 3D holoscopic image by using spatial correlation of rendered view images," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2002-2006, 2017.
[39] L. Shen, Z. Zhang and Z. Liu, "Adaptive Inter-Mode Decision for HEVC Jointly Utilizing Inter-Level and Spatiotemporal Correlations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.24, no. 10, pp. 1709-1722, 2014.
[40] S. Wang, F. Luo, S. Ma, X. Zhang, S. Wang, D. Zhao, W. Gao, "Low complexity encoder optimization for HEVC," *Journal of Visual Communication and Image Representation*, vol. 35, pp. 120-131, 2016.
[41] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, "Call for Proposals and Evaluation Procedure," ICME 2016 Grand Challenge: Light Field ImageCompression, Seattle, USA, pp. 1-8, 2016.
[42] T. Georgiev, January 2013 (Online), Available: http://www.tgeorgiev.net, Website (Online).
[43] F. Bossen, Common HM Test Conditions and Soft-ware Reference Configurations, Document JCTVC-L1100, 2013.
[44] Y. Cao, M.A. Brubaker, D.J. Fleet, A. Hertzmann, "Efficient optimization for sparse Gaussian process regression," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 37, no. 12, pp. 2415-2427, 2015.
[45] C. Conti, P. Nunes and L. D. Soares, "Light field image coding with jointly estimated self-similarity bi-prediction", *Signal Processing: Image Communication*, vol. 60, pp. 144-159, 2018.
[46] T. Georgiev, A. Lumsdaine, Focused plenoptic camera and rendering, J. Electron. Imag. 19 (2010) 021106-1C021106-11.
[47] D. Dansereau, "Light Field Toolbox v0.4" (Online), Available: http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4 (Online).
[48] ISO /IEC JTC 1/SC 29 /WG 1, JPEG, "JPEG PLENO LIGHT FIELD CODING COMMON TEST CONDITIONS", Doc. N80027, Berlin, Germany, July 2018.

**Deyang Liu** received his BS degree in communication engineering from Anqing Normal University, Anqing, China, in 2011, and his MS degree and Ph.D. degree in signal and information processing from Shanghai University, Shanghai, China, in 2014 and 2017. He is currently a lecturer in School of Computer and Information, Anqing Normal University. He is in charge of several projects including the National Natural Science Funds project. His research interests include 3-D video processing, video coding.

**Ping An** received the B.E. and M.E. degrees from the Hefei University of Technology, Hefei, China, in 1990 and 1993, respectively, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2002. In 1993, she joined Shanghai University. From 2011 to 2012, she was a Visiting Professor with the Communication Systems Group, Technical University of Berlin, Berlin, Germany. She is currently a Professor with the Video Processing Group, School of Communication and Information Engineering, Shanghai University. Her research interests include image and video processing, with a focus on 3-D video processing and computer vision. She has completed more than 15 projects supported by the National Natural Science Foundation of China, the National Science and Technology Ministry, and the Science and Technology Commission of Shanghai Municipality. She was a recipient of the Second Prize of the Shanghai Municipal Science and Technology Progress Award in 2011 and the Second Prize in Natural Sciences of the Ministry of Education in 2016.

**Ran Ma** received the M.S. degree and the Ph.D. degree in school of communication and information engineering from Shanghai University, Shanghai, China, in 2000 and 2008, respectively. In 2000, she joined Shanghai University, where she became an Associate Professor in 2009. From 2012 to 2013, she was a visiting professor with the New York University, New York, NY, USA. Her current research interests include image and video compression. She was a recipient of the Second Prize of the Shanghai Municipal Science and Technology Progress Award in 2011 and the Second Prize in Natural Sciences of the Ministry of Education in 2016.

**Wenfa Zhan** received his Ph.D degree in Hefei University of Technology, Anhui Province in 2009. He is currently a professor in the Department of Educational Technology, Anqing Normal University, Anhui Province. His research interests include test data compression, ATPG algorithms, etc. He has published over 50 papers and hold ten Chinese patents.