Li Y.; Sjöström, M.; Olsson, R; Jennehag, U., "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015

http://dx.doi.org/ 10.1109/TCSVT.2015.2450333

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

Coding of focused plenoptic contents by displacement intra prediction

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

***Abstract*—A light field is commonly described by a two-plane representation with four dimensions. Refocused three-dimensional contents can be rendered from light field images. A method for capturing these images is by using cameras with microlens arrays. A dense sampling of the light field results in large amounts of redundant data. Therefore, an efficient compression is vital for a practical use of these data. In this paper, we propose a displacement intra prediction scheme with a maximum of two hypotheses for the compression of plenoptic contents from focused plenoptic cameras. The proposed scheme is further implemented into HEVC. The work is aiming at coding plenoptic captured contents efficiently without knowing underlying camera geometries. In addition, the theoretical analysis of the displacement intra prediction for plenoptic images is explained; the relationship between the compressed captured images and their rendered quality is also analyzed. Evaluation results show that plenoptic contents can be efficiently compressed by the proposed scheme. Bit rate reduction up to 60 percent over HEVC is obtained for plenoptic images, and more than 30 percent is achieved for the tested video sequences.**

***Index Terms*—Plenoptic images, Plenoptic videos, light field, HEVC, compression.**

I. Introduction

A light field represents the intensity and the direction of the outgoing radiance from a scene and can be sub-sampled by capturing the scene with plenoptic cameras. Arbitrary views and views at different focused planes can be generated by combining the pixels from the captured microlens images, also called Elementary Images (EI). However, a dense sampling of the light field results in highly correlated EIs. Conventional intra prediction in HEVC is inefficient for exploiting this correlation [1] [2]. The question is how much the compression efficiency can be achieved by using spatial displacement intra prediction with more than one hypothesis (reference signal) for the coding of these plenoptic contents.

The plenoptic function $I = P(x, y, z, \theta, \phi, \lambda, t)$ [3] has seven dimensions and captures the intensities I of light rays at any of the viewing positions x, y, z, directions θ, ϕ, wavelengths λ, and time t. For a static scene and to represent the color in RGB, the plenoptic function is reduced to five dimensions without λ and t. The parameter of wavelengths is removed due to the integration over wavelengths for the color intensity $I = [I_R, I_G, I_B]^T$. Further assuming regions are free of occluders, the plenoptic function can be simplified into four dimensions as a light field [4] [5], which is represented by a two-plane representation. The four dimensions, (p, q, r, t), locate the coordinates of a light passing through from one plane (p, q) to another (r, t). A conventional camera averages the intensities of radiances described by the higher dimensional plenoptic function into the two-dimensional image sensor of the camera. A light field capturing is, however, a sampling of the light field in its four dimensions, because acquisition of a full light field is practically infeasible. There are four techniques commonly used for capturing a light field image: with camera arrays [6], with a moving camera [7], with coded apertures [8], and with microlens arrays in a camera. From the capturing with microlens arrays, two different capturing approaches are further derived: standard plenoptic capturing and plenoptic 2.0. Cameras with plenoptic 2.0 techniques are also referred to as focused plenoptic cameras. For clarity, in the context of this paper, we refer to cameras with microlenses as plenoptic cameras.

The first plenoptic camera was introduced by Gabriel Lippmann in 1908 [9]. But the commercially available plenoptic camera was firstly produced by Lytro, Inc., founded by Ng et al. [10] [11] in 2006. The Lytro camera captures the distribution of light rays as described by the light field. This capability is achieved by putting a microlens array in front of the image sensor. Because the focal plane of the microlens is on the image sensor, only angular information is captured in each EI. This camera set-up is the standard plenoptic capturing system, which results in a low spatial resolution of rendered views. To overcome this drawback and trade-off the spatial resolution with the angular resolution, focused plenoptic cameras [12] have been devised. In images captured by focused plenoptic cameras, each EI is essentially a small cropped multi-view image from a specific viewing angle. The focused plenoptic cameras are described in detail in section III.

We define plenoptic images as the images captured by plenoptic cameras. A plenoptic image [13] with rectangular EIs from focused plenoptic cameras is shown in Fig. 1.

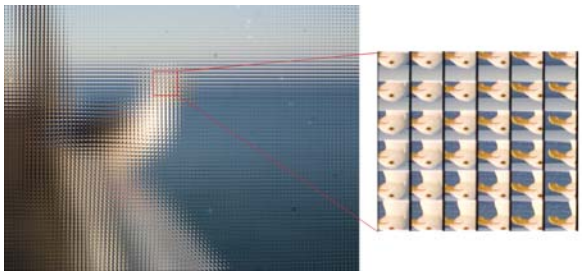


Fig. 1. Images from focused plenoptic cameras [13].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

A. Motivation

Multi-view rendering and refocusing from plenoptic contents only involve re-sampling of the captured dataset. The rendering is also fairly simple compared to traditional photo-realistic graphic rendering techniques that require scene geometry and surface shading models [14]. There are numerous applications benefited from light field rendering techniques, ranging from medical applications to 3D videos and computer games, etc.

The EIs in a plenoptic image resemble multi-view video contents, but, with a small image size. The total amount of pixels in each direction of an EI are usually in the magnitude of tens of pixels to hundreds of pixels [15] [12]. A straightforward approach to encode these images would be to apply standard image or video encoders, e.g., JPEG and H.264, on the captured images. However, the redundancy between EIs is not exploited by these encoders. Alternatively, multi-view encoders, such as MV-HEVC [2], can be applied on EIs for decorrelation. But, it is problematic to separate each EI from the plenoptic images if the geometries of the camera are unknown. The geometries, e.g., modulation images, are related to the camera settings when the image is captured [15]. The modulation images are obtained during camera calibration, depending on the focus of the main lens, the focal length of the main lens, and the opening of the camera aperture. The calibration process leads to different images captured by the same camera can have different geometries.

Even if the geometrical problem is overcome and the EIs can be separated from the plenoptic images, another problem arises: how to arrange them into a two-dimensional grid of image and feed them into multi-view encoders. This is because the arrangement of EI changes with the camera, EI is fairly small, and there are usually hundreds by hundreds of EIs in a plenoptic image. It is possible to rearrange the EIs into other formats, e.g., to rendered views. However, to produce a good rendered view, additional information is required [12] [16], e.g., disparity maps between EIs. In the case of video sequences, disparity maps have to be estimated for every frames.

As mentioned above, the difficulties of coding separated EIs motivate us to devise a general coding scheme that can 1) be applied to plenoptic images and videos from all focused plenoptic cameras without knowing camera geometries and 2) decorrelate EIs in the compression.

B. Proposed method

In this paper, we introduce a three-dimensional displacement intra and inter prediction scheme for the coding of plenoptic images and videos. The scheme is implemented into HEVC. The prediction is able to perform both in the spatial domain (x, y) and the temporal domain (previous and future) with a maximum of two hypotheses.

The novelties of this paper are 1) we develop a multi-hypothesis prediction scheme for the coding of plenoptic contents; 2) the scheme is integrated into HEVC framework; 3) the displacement intra prediction is explained theoretically with the proposed plenoptic signal model; 4) the coding performance is analyzed empirically for plenoptic contents with different input image properties; 5) we also investigate the efficiency of the Test Zone (TZ) search [17] compared to the full search for the proposed scheme; 6) the impacts of coding in the captured image to the rendering quality are theoretically analyzed and visually inspected.

The overall aim of the work is to improve the compression efficiency for light field contents. The work is limited to the compression of plenoptic images and videos from focused plenoptic cameras, and the maximum number of hypotheses for the proposed coding scheme is limited to two. The goal is to investigate the rate-distortion ratio objectively for the proposed method and to relate the PSNR of compressed plenoptic images to the visual quality of rendered views.

C. Outline

The sequence of the paper is organized as follows. Section II describes previous work, and the focused plenoptic camera is presented in Section III. We illustrate the displacement intra prediction with multi-hypothesis for plenoptics in Section IV, the coding block and the intra prediction in HEVC in Section V, and the proposed methods in Section VI. Test arrangements and evaluation criteria are presented in Section VII, and Section VIII shows the results and analysis. Section IX concludes this paper.

II. Previous work

Previous works on light field compression can be categorized into three different groups [18]: vector quantization, progressive coding and predictive coding. The *vector quantization* approach [4] exploits data redundancies by representing an entire vector space with a subset of vectors. In *progressive coding*, Discrete Wavelet Transform (DWT) is commonly used in light field compression for a finer granularity of progressive scalability [14] [18] [19]. A typical example is the disparity-compensated lifting and shape adaptation scheme [14] that employs a Shape Adaptive Discrete Wavelet Transform (SA-DWT) to better preserve object boundaries. 4-D wavelets [20] have also been applied to 4-D light field contents. An early work in *predictive coding* is presented in [21], which arranges light field contents into a grid of images. An image is predicted with different modes from a few intra-coded images within the grid. This work is further improved by using homography [22]. In addition, there are also other approaches that do not distinctively lie in a group mentioned above for the coding of light field contents. Principle Component Analysis (PCA) was utilized in [23] [24]. Paper [25] presents an approach that separates objects in ray space and applies wavelets to code each part separately. The model-based approach that represents objects by voxels and uses geometries for prediction has been investigated in [20]. The performance of distributed video coding for light field contents was analyzed in [26] [27]. An approach based on compressive sensing [28] has recently been proposed to capture and encode light field images.

In order to apply the above approaches for an efficient coding of plenoptic images, EIs must be separated from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

plenoptic image. To avoid the separation process, the Self-Similarity (SS) mode [1] has then been introduced into HEVC for the coding of plenoptic images [29] [1] and videos [30]. The SS mode is to predict the current block from reconstructed signals, and the SS mode uses only one hypothesis (reference) block for the prediction. However, the previous theoretical and empirical studies [31] [32] [33] illustrate that inter-frame prediction using multi-hypotheses is more effective at reducing prediction errors. The displacement intra prediction is similar to the inter-frame prediction. Therefore, for the coding of plenoptic images, we have deployed multi-hypothesis spatial displacement intra prediction into the HEVC framework by using a maximum of two hypotheses. The results from our previous work [34] demonstrates a significant improvement compared to original HEVC intra and the single hypothesis displacement intra for plenoptic images in a fast search mode. With a rapid development, HEVC range extensions have recently introduced the Block Copying (BC) [35] mode with spatial intra prediction for coding of screen contents. However, this mode has a restriction on search areas and performs only one hypothesis prediction. A thorough assessment of the coding with more hypotheses is of interest. In addition, for our previous work in [34], different mode assignments, search methods for displacement vectors, the influences of input signals from various plenoptic cameras, and coding complexity etc. have not been investigated, and it is essential to develop a coding scheme for plenoptic videos with multi-hypothesis and analyze its coding performance as well.

III. Focused plenoptic camera

This section briefly describes the focused plenoptic cameras in terms of capturing and rendering, and the relations between captured images and rendered views. This is to give an insight to the properties of focused plenoptic images and how the rendering is affected by compression.

A. Capturing and EI cross-similarities

A typical focused plenoptic camera set-up [12] is presented in Fig. 2, with the main lens focusing on the main lens image plane, and the microlenses focusing on a plane in front of the image sensor plane. The camera is capable of capturing both spatial and angular information of a scene in each EI, and the angular information is also spread over several EIs for a spatially located point [12]. The trade-off between spatial and angular ratio is adjusted by the parameters (a and b) of the camera and the size of the microlenses, etc. The capturing process results in correlated images in adjacent EIs. The image correlation depends on a and b, which can be shown by a simple ray tracing, e.g., an increase of a can lead to more correlated EIs. But, if a fixed value of a and b is given, the correlation between EIs is determined by the homogeneity and the depth of the scene. More specifically, 1) EIs appear similar if the scene surface is homogeneous; 2) a large number of adjacent EIs capture the same parts of the scene if the objects are farther away from the camera (for the set-up with the main lens image plane in front of the microlenses shown in Fig. 2). In order to exclude the image properties of the scene surface, we introduce the term cross-similarity to measure the repetitiveness of the scene between EIs. High cross-similarity is referred to as a large number of adjacent EIs capturing the same parts of the scene. The repetitive patterns of EIs due to cameras and scene depth are of special interest for our coding scheme. An example of the captured image with respect to the cross-similarity can be seen in Fig. 3, which shows parts of two frames from the $PlaneAndToy$ sequence [36].

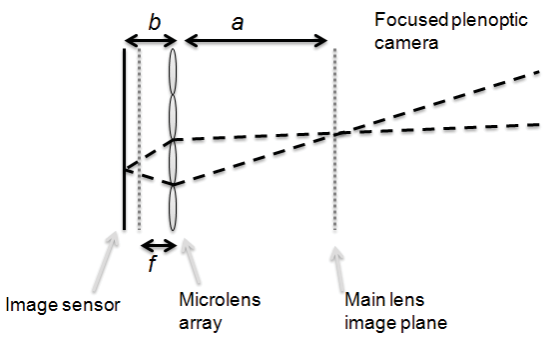


Fig. 2. Focused plenoptic camera [12].

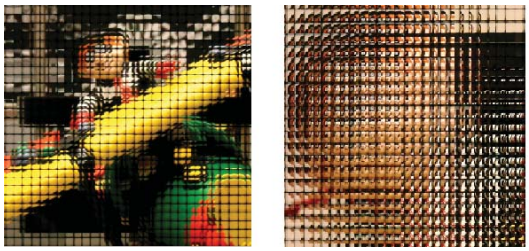


Fig. 3. Focused plenoptic image: low cross-similarity (left), high cross-similarity (right) [36].

B. Rendering

Calibrations are commonly conducted before the actual rendering to reduce rendering artifacts. This is because noise, random shift of microlens image, and vignetting etc. are present in the captured images [15]. The calibration involves intensity compensation with modulation images, microimage center determination, etc.

Multi-view of all-in-focus images is rendered by combining patches from each EI [12]. Fig. 4 shows the rendering approach. Because objects in a scene can be in different depth planes, the cross-similarity between EIs changes and the disparities between EIs vary throughout a plenoptic image. Furthermore, the patch sizes are determined by the disparities, which correspond to the depth planes of the scene. For example, the EIs containing objects farther away from the camera for the set-up in Fig. 2 have a higher cross-similarity. Hence, the disparities and the patches from these EIs are smaller. This implies that the disparities must be estimated for a possible artifact-free rendering. The patches of different size

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

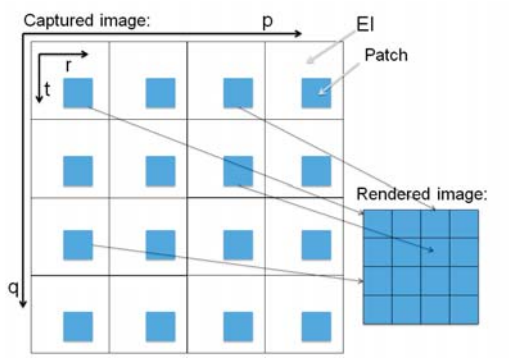


Fig. 4. Captured plenoptic image and rendered view [12].

are then normalized with different magnification factors and combined to form a rendered image [12]. A simpler approach to circumvent the disparity estimation is to use a constant patch size. But, by doing so, artifacts will exhibit on parts of the rendered image [12], if the objects on those parts are not in the depth plane corresponding to the fixed patch size being used.

Image refocusing is conducted by a blending process, which involves the integration in the angular dimension for a spatial point. In short, the blending is to overlap EIs with a certain disparity, and the summation of the intensity of the overlapped areas must be averaged with the number of overlapping. The objects at the depth plane that corresponds to the current disparity are brought into focus and other areas are blurred.

C. Relations between captured images and rendered views

The quality of the in-focus rendered images are of interest. The out of focus areas, which appear as blurring, are not intended to attract the attention of the viewer. Let the captured plenoptic image be $C(p,q,r,t)$, $(1 \le p < N_p,\ 1 \le q < N_q,\ 1 \le r < N_r,\ 1 \le t < N_t)$, and N_p, N_q, N_r, N_t are the size of each of the dimensions. The impacts of the distortions induced by compression on the captured image to the rendered image are as follows.

Firstly, lowered distortions, e.g., lowered Mean Square Error (MSE), on the captured image implies lower distortions on the all-in-focus rendered views if the distortion is distributed uniformly throughout the decoded captured image. This is because the all-in-focus rendered views are just a subset of the captured image.

Secondly, the distortion is likely more visible on the all-in-focus rendered view than on the refocused image. This is because the intensity of a pixel in a refocused image is a result of averaging multiple corresponding pixels from adjacent EIs for a spatial point. More specifically, assume that the error of a pixel after coding is a random variable $\varepsilon_i = C(p_i,q_i,r_i,t_i) - C'(p_i,q_i,r_i,t_i)$, where p_i, q_i, r_i, t_i are the indexes in the four dimensions, and $C'(\cdot)$ is the decoded captured image. The error ε_i results from quantization. When adjacent EIs are overlapped with a disparity and averaged for the refocusing, the square error of a pixel due to averaging becomes

$$E[(\frac{1}{W}\sum_{i=1}^{W}\varepsilon_i)^2] \approx \quad \frac{1}{W^2}\sum_{i=1}^{W}E[\varepsilon_i^2] \tag{1}$$

where W is the total number of averaged EIs, which depends on the disparity. Eq. 1 holds because the cross term after the the expansion at the left side of the equation $E[\varepsilon_i\varepsilon_j] \approx 0$ under the assumption that ε_i and ε_j are uncorrelated for $i \neq j$ and have a zero mean approximately. It has been pointed out in [19][37] that it is reasonable to assume that the quantization errors ε_i are uncorrelated with each other, which is under the condition with uniform quantization for smooth density distribution and in high rate. Because $\frac{1}{W^2}\sum_{i=1}^{W}E[\varepsilon_i^2] < E[\varepsilon_i^2]$, the square error on the refocused image is smaller than on the all-in-focus rendered view.

The above analysis implies that it is reasonable to perform a visual inspection of the all-in-focus rendered views to evaluate the performance of a compression scheme for plenoptic images.

IV. Displacement Intra Prediction with Multi-Hypothesis for Plenoptics

In this section, we propose a signal model for plenoptic images and incorporate it into the Motion Compensated Prediction (MCP) for multi-hypothesis prediction in [31]. This is to give an insight into how multi-hypothesis can improve the coding efficiency for displacement intra prediction with respect to focused plenoptic images.

Modeling of plenoptic signals: A 3D scene represented by a light field is modeled by a 2D planar surface v_{2D} in [19], and the surface v_{2D} is assumed approximately lambertian. We denote an EI as $C(p_i,q_i,r,t)$ and model the EI image through the mapping shown in Fig. 5 as

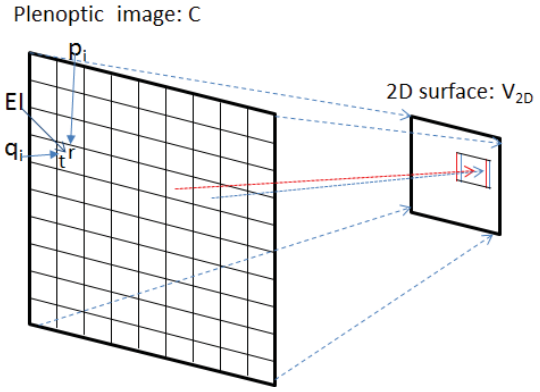


Fig. 5. Mapping between the captured plenoptic image and the 2D surface.

$$\begin{aligned} &C(p_i,q_i,r,t) \\ &= v_{2D}((p_i-1)\ N_r + r + \Delta l_{r_i}, (q_i-1)\ N_t \\ &+ t + \Delta l_{t_i}) + n_{p_i,q_i}(r,t). \end{aligned} \tag{2}$$

The noise term $n_{p_i,q_i}(r,t)$ includes geometrical errors, Non-lambertian effects, lens distortions and disocclusions. Δl_{r_i} and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

Δl_{t_i} are the offsets that map $C(p_i, q_i, r, t)$ to the 2D surface v_{2D}.

We derive a hypothesis texture image from $C(p_i, q_i, r, t)$ as $c_i(r,t) = v_i(r + \Delta l_{r_i}, t + \Delta l_{t_i}) + n_i(r,t)$, where $v_i(o_1, o_2) = v_{2D}((p_i - 1)\ N_r + o_1, (q_i - 1)\ N_t + o_2)$. The current texture image being predicted is $s(r,t) = v_0(r + \Delta l_{r_0}, t + \Delta l_{t_0}) + n_0(r,t)$. The multi-hypothesis signals c_i are assumed to be found from

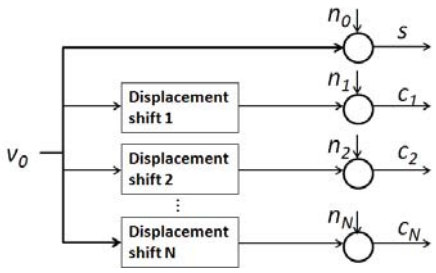


Fig. 6. Signal model for the displacement intra with respect to focused plenoptic images; s: current EI being predicted, and c_i: neighboring EIs.

different neighboring EIs, i.e., c_i is a shifted version of v_0 plus noise n_i (c_i and s are cross-similar). This assumption leads to the same signal model as described in [31] and shown in Fig. 6. The displacements between s and c_i are related by two-dimensional shift displacement vectors.

It is assumed that s and c_i are produced by a jointly wide-sense stationary Gaussian random process as in [31]; s is predicted by a linear prediction from c_i. If W number of hypotheses are used for prediction, in the simple averaging case, s is predicted from $\frac{\sum_{i=1}^{W} c_i}{W}$. We further assume the same signal statistics for n_i, v_0, and shift displacement errors as in [31]. The multi-hypothesis MCP analysis [31] can be applied to the plenoptic signal model between cross-similar EIs. HEVC has also inherited the basic structure of hybrid video encoders [38] with such a MCP scheme.

The important theoretical analysis results [31] of the multi-hypothesis MCP over an optimal intra coding for Gaussian wide sense stationary sources are the following: 1) increasing the number of hypotheses reduces the coded bit rate, especially when the motion prediction is accurate, (e.g., a quarter pixel prediction); 2) the spectral noise on the signal influences the prediction, i.e., the bit rate reduction with multi-hypothesis becomes smaller with an increasing level of noise; 3) when the power of residue noise increases, doubling the number of hypotheses is more effective than doubling the accuracy of the prediction; 4) averaging several good quality hypotheses always reduces the coded bit rate. Additionally, in the simple averaging case when more than one hypothesis is used, the analysis also shows the largest coding gain is obtained when the number of hypothesis increases from one to two, which assumed a realistic level of noise, n_i, on the images. Following the multi-hypothesis MCP analysis, a later study [32] concludes that using more than two hypotheses is less effective in reducing prediction error variance in theory and provides insignificant coding gains in experimental results.

The above analyses suggest it is appropriate to increase the number of hypothesis to achieve a more efficient compression for focused plenoptic images. Within the scope of this paper, we only consider the compression efficiency related to the number of hypothesis increased to the maximum of two in our experiment.

V. The Coding Block and the Intra Prediction in HEVC

A Coding Unit (CU) in HEVC consists of one luma Coding Block (CB), two chroma CBs and its associated syntax [38][2][39]. The Coding Tree Unit (CTU) is a basic processing unit and the largest block allowed in HEVC. A CU can be split recursively until it reaches a maximum predefined depth. In each splitting level, the CU is further split into Prediction Units (PUs) of different sizes, which are the smallest blocks upon which a prediction is conducted. An exception is when the Transform Unit (TU) is utilized for intra prediction. The TU is also split recursively from the CU, and transform coding is performed on each TU. For an efficient compression, the Rate-Distortion Optimization (RDO) criterion, $\min(D + \lambda R)$, is employed to determine the best CU splitting depth, PU fragmentation, TU size, and PU prediction modes, etc. D is the distortion introduced by coding, λ the Lagrange multiplier, and R the coded bit rate. Based on the RDO, a set of prediction modes, both intra and inter, are tested. The RDO minimizes the distortion D given a bit rate constraint R. In relation to our proposed scheme, we briefly illustrate the intra prediction for HEVC here. A detailed description is found in [2].

The intra prediction of a PU in HEVC is based on its top and left boundary samples from neighboring reconstructed PUs. When a CU is split into multiple TUs, the intra prediction is conducted on each TU based on neighboring reconstructed TU samples. Fig. 7 illustrates the prediction that a PU is predicted diagonally from its top reconstructed boundary samples. The PU can be predicted with 33 directional modes plus a planar and a DC mode, and these modes are also applied when TU is used for prediction. The prediction modes are encoded predictively by using the most probable modes, which are derived from the previously coded neighboring PUs.

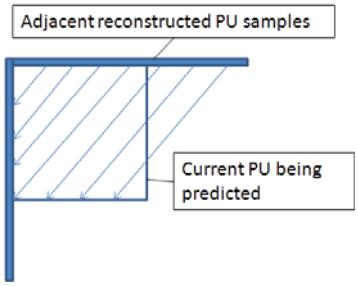


Fig. 7. Intra-prediction: the current PU is predicted from the boundary of its neighboring reconstructed PUs.

VI. Proposed Methods

The intra prediction scheme in HEVC is unable to explore the repetitive patterns of plenoptic images, because only

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

boundary samples are used for the prediction. We apply the displacement intra prediction into HEVC as an intra prediction scheme to solve this problem. However, the prediction is with a maximum of two hypotheses, which are also limited by the HEVC framework. In addition, the encoding time complexity will increase substantially if more hypotheses are used. Furthermore, the analyses of displacement intra prediction for plenoptic images illustrated in Section IV suggest that increasing the number of hypotheses from one to two is effective in reducing coding bit rates.

A. Coding of plenoptic images (intra mode)

In video coding, the bit rates for intra-coded frames are much higher than for inter-coded frames. Therefore, the bit rate produced by intra prediction can account for a large portion of the total bit rates if video frames are intra-coded frequently, either by configuration or because of drastic scene changes between frames. Hence, the intra image compression efficiency is of paramount importance.

Fig. 8 presents the concept of the proposed displacement intra prediction coding scheme. It can be seen as a derivation from the inter-prediction scheme of HEVC. The proposed scheme takes the adjacent reconstructed region in the same plenoptic image as reference pictures in HEVC for the prediction. The region is limited by a predefined search range parameter within the reconstructed CU blocks. An example is shown in Fig. 9, in which the reconstructed region is marked with color light gray and dark gray and is neighboring the current PU, block B. In order to perform the bi-directional prediction with two references, the reconstructed region is separated into two parts colored light gray and dark gray, respectively. These two parts are assumed to be two reference pictures in each of the two picture reference lists L_0 and L_1. The picture reference lists are utilized to accommodate the reference pictures for the prediction. For the uni-directional prediction with a single reference, the entire reconstructed region (light gray and dark gray) is assumed to be a reference picture in the list L_0. We call the intra mode with bi-prediction as *image B-coder* and the one with uni-prediction as *image P-coder*.

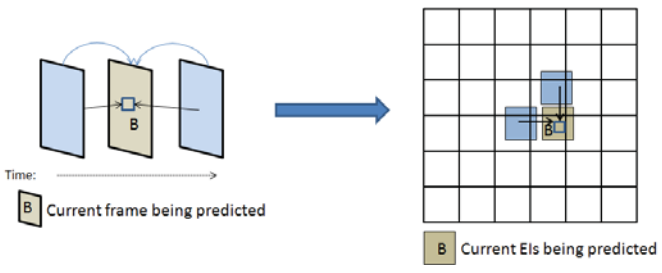


Fig. 8. Displacement intra-prediction derived from inter-prediction for plenoptic image compression.

Image B-coder (intra mode): the prediction for the *B-coder* is similar to the prediction for the B frame coding in HEVC. There are three candidates for the prediction of a current PU, block B in Fig. 9. As to the first two candidates, they are the best matching blocks from each of the reference lists L_0 and L_1. These two blocks are, however, not shown in Fig. 9. The third candidate is obtained by $\frac{P_0+P_1}{2}$, where P_0 and P_1 are two blocks from the lists L_0 and L_1, respectively. More specifically, block P_0 is found by a refinement search in the neighborhood of the best matching block from L_0, and block P_1 is obtained from L_1 in the same way. This refinement search is designed to find the best matching signal $\frac{P_0+P_1}{2}$ to the current PU within the refinement search range. As a result, the best of the three is chosen to predict the current PU block.

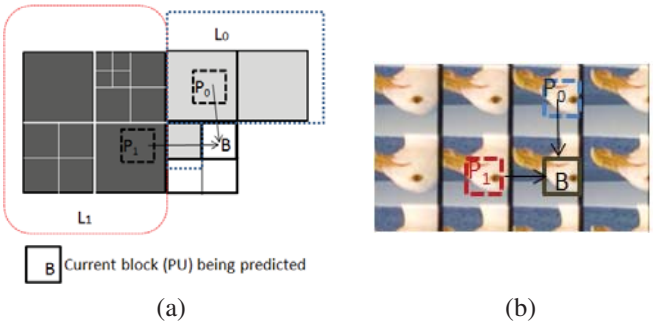


Fig. 9. Bi-prediction within an image. (a) Two parts in color light gray and dark gray are assumed as two reference pictures and available in the reference list L_0 and L_1; (b) an illustration of the prediction on a plenoptic image.

Image P-coder (intra mode): the proposed *P-coder* is essentially performing a single reference prediction. The best matching block is obtained from a search process from the entire reconstructed region (i.e. with light gray and dark gray regions in Fig. 9). Consequently, the best block is selected from the region to predict the current PU block.

The Advanced Motion Vector Prediction (AMVP) [2] technique in HEVC is also applied in the proposed scheme. The displacement vector obtained from the search is not encoded directly. Instead, the AMVP encodes the difference between the current displacement vector and its predictor. The predictors are found from the spatially neighboring and temporally collocated reconstructed PUs of the current PU. In the merge mode of AMVP, the displacement vectors of the current PU can also be derived from its spatially neighboring and temporally collocated reconstructed PUs, where the latter is only available in the case of video coding. We describe the plenoptic video coding in the next section.

The proposed displacement intra prediction is tested using the RDO criterion along with the traditional intra during encoding. The best of them is selected as the final intra prediction mode for the current PU. The *B-coder* is expected to improve the compression efficiency over the *P-coder*. The reason for such an improvement is demonstrated in Section IV. The proposed intra mode *P-coder* and the *B-coder* use the same syntax as the P slice and the B slice [2] in HEVC.

B. Coding of plenoptic videos

The coding of videos aims to extend the prediction within an image into the temporal domain. Fig. 10 illustrates the process of plenoptic video coding. Up to 16 reference pictures can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

loaded into each of the lists, L_0 and L_1. We call the proposed video encoder with two hypotheses as *video B-coder*.

Video B-coder: In addition to the reference pictures from the adjacent reconstructed part of the same image (video frame), neighboring reconstructed frames in the temporal domain are also loaded as reference pictures into the reference lists, L_0 and L_1 as shown in Fig. 10. The exact frames loaded into the lists depend on the HEVC configurations. For the prediction, the predictor candidates are the best matching blocks from each of the lists and the $\frac{P_0+P_1}{2}$. The P_0 and P_1 are from the reference pictures in the lists L_0 and L_1, respectively, in a similar way as for the proposed intra frame coding by a refinement search. More specifically, the refinement search here is performed iteratively for all the combinations of the reference pictures in the lists L_0 and L_1, and is around the proximity of the best matching block from each reference picture. A detailed description of the iteration process is referred to [2] and [40]. Furthermore, the search for the best matching block from the temporal domain is limited to the reference pictures within a predefined search range. As a result, the best of the three candidates is used to predict the current PU.

The prediction for the video coder can be seen as a hybrid of displacement intra and inter frame prediction. To encode the current PU, the encoder will find the best prediction modes with the RDO among the conventional intra, and the hybrid of displacement intra and inter frame prediction.

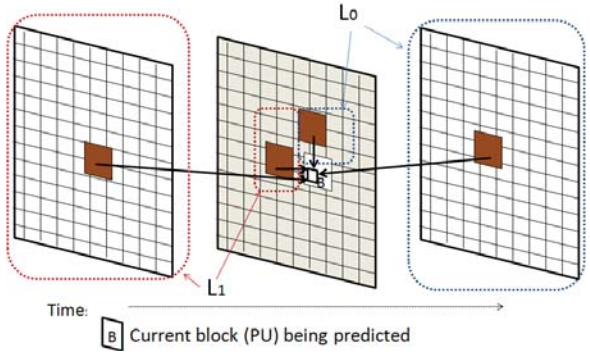


Fig. 10. Plenoptic video coding: prediction candidates can be selected from spatial domains as well as from temporal domains.

C. Encoding complexity of the proposed schemes

The time complexity of HEVC has been analyzed empirically in [41]. It is shown that the motion prediction in HEVC takes up a significant portion of time in encoding. For the proposed *image P-coder*, the coding complexity is equivalent to HEVC P frame coding with one reference picture, and for the proposed *image B-coder* equivalent to HEVC B frame coding with one reference picture in each of the picture reference lists. Therefore, the proposed image coders have higher encoding time complexity than HEVC original intra. In addition, the *video B-coder* has a higher encoding time complexity than HEVC B frame coding given the same prediction structure, and the complexity increased is equivalent to adding one more reference picture for the HEVC B frame coding in each of the picture reference lists. However, the complexity analysis for the decoding in [41] illustrates that coding with temporal P and B prediction is faster in a real time implementation than with all intra frame coding, because less time is spent on the entropy decoding. This suggests that it can be advantageous for the proposed image coders in decoding over the original HEVC intra.

VII. Test arrangement and evaluation criteria

Due to the importance of the intra prediction in reducing bit rates for a coded intra frame, it is of interest to test the intra frame coding efficiency. In addition, in order to draw a more general conclusion, images from different plenoptic cameras with various image characteristics were considered for the test. The characteristics include microlens structures, EI cross-similarity, etc.

Light field images, e.g., *Seagulls*, *Books* [13], *PlaneAndToy* [36], were used in the test. We also captured an image, *OpticalTable*[1], by using Raytrix camera [42]. This image is of size 6576 by 4384 with hexagonal EIs with a width of approximately 36 pixels.

The video sequence of *PlaneAndToy*, see Fig. 11, contains 250 frames of plenoptic images with a stationary background. The cross-similarity between EIs changes over the sequence. In addition, another video sequence, DemichelisSpark [36], was also used for the test. The cross-similarity between EIs does not change throughout the sequence, therefore, only the first 50 frames were tested. Table I summarizes the tested plenoptic input data.

TABLE I
Plenoptic images/videos

Image/Video	Resolution	EI shape and size in pixels	Cross-similarity (EIs)
Seagull	7240×5236	Square(75)	High
Books	3913×3913	Circular(50)	Low
OpticalTable	6576×4384	Hexagonal(36)	Low
PlaneAndToy (high)	1920×1088	Square(27)	High
PlaneAndToy (low)	1920×1088	Square(27)	Low
PlaneAndToy (video)	1920×1088	Square(27)	Change
DemichelisSpark	2880×1620	Circular(38)	Low

The HEVC Test Model (HM) reference software version 11 [40] was modified for the proposed scheme. The test was conducted in TZ search unless otherwise mentioned. The search range was set to 192 for *Seagull*, *Books*, and *OpticalTable*, and to 128 for *PlaneAndToy* and *DemichelisSpark*. The search range is different because the EI size is larger for *Seagull*, *Books*, and *OpticalTable*. The search range refinement for bi-prediction was set to 4 as default.

A. Intra coding

The images were transformed into YUV 4:2:0 format. The tested Quantization Parameters (QPs) for images were selected to be 20, 30, 40, and 50 for *Seagull*, *Books*, and *OpticalTable*.

[1]http://plenoptics.droppages.com/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

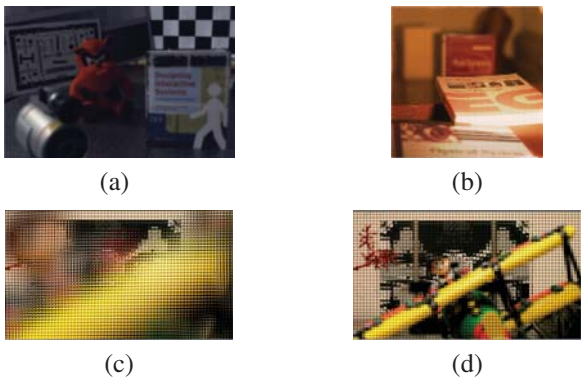


Fig. 11. Light field images: (a) *OpticalTable*; (b) *Books*; (c) *PlaneAndToy* (186^{st} frame) with high cross-similarity; (d) *PlaneAndToy* (1^{st} frame) with low cross-similarity.

Two images from the *PlaneAndToy* sequences were extracted for the intra coding test. They are the first frame of *PlaneAndToy* with low EI cross-similarity denoted as *PlaneAndToy(low)* and the 186^{th} frame with high cross similarity denoted as *PlaneAndToy(high)*. The QPs for *PlaneAndToy* are set to 22, 27, 32, and 37. The bit rate, bits per pixel (bpp), was obtained from the coded bit stream for all YUV components, whereas the quality was measured on the luminance component by using PSNR. The rate-distortion curves of PSNR versus bpp are presented for the *image P-coder*, the *image B-coder*, and the original HEVC intra coder. Bit rate changes for the *P-coder* and the *B-coder* over the original HEVC have also been expressed in the BD-PSNR [43], which computes the average difference between two rate-distortion curves.

The configuration parameters for the proposed image coders were defined as the "Low delay-Main" setting in JCTVC-L1100 [44], and for the original HEVC as the "All intra-Main" setting in JCTVC-L1100 [44]. Additionally, the coding efficiency of JPEG2000 [45] was compared to the HEVC original intra. We used the OpenJPEG software [46] and the default setting with one quality layer and one tile. The coding quality is set to the same as the HEVC original intra coding for the four bit rate points.

The following aspects are of interest for the analysis: 1) cross-similarity between EIs; 2) search strategies (the full search versus the TZ search); 3) intra mode *image B-coder* versus *image P-coder*; 4) the percentage of different modes used.

The decompressed images were also rendered for a visual quality investigation. The rendering approach with a constant patch size was employed to render all-in-focus images. The calibration step mentioned in Section III-B was skipped because the camera geometry is unknown. This will not affect the evaluation, because all-in-focus images are rendered under the same condition from different compressed images. The part of the image that is in the depth plane corresponding to a given patch size is of special interest.

B. Video coding

For the video sequences *planeAndToy* and *DemichelisSpark*, the "Random access-Main" setting (hierarchical B frame coding) in JCTVC-L1100 [44] with a GOP of size 8 was used for HEVC with temporal prediction and the proposed *video B-coder*. The QPs used were 22, 27, 32, and 37. The CTU size was set to 16 to reduce the coding time complexity. A bigger CTU, e.g., 32, than the size of an EI will likely cause more CU splitting and increase side information (e.g., splitting flags, displacement and motion vectors). The rate-distortion curves and the BD-PSNR were also acquired for the video sequence. However, the bit rate is measured in kilo-bits per second (kbps).

VIII. Results and analysis

A. Intra coding

(1) Cross-similarity between EIs: The results in Table II show that the compression efficiency is related to the cross-similarity between EIs. A higher cross-similarity between EIs implies that more hypotheses signals are available for the prediction. Therefore, larger bit rate saving is achieved. The bit rate savings for the *image B-coder* are 56.71 and 26.64 percent for *Seagull* and *PlaneAndToy(high)* compared to the HEVC intra, respectively, in the TZ fast search case shown in Table II. For the *image P-coder*, the improvement compared to the *image B-coder* is less. But, substantial bit rate reduction can still be seen in Table II. The bit rate reductions are 42.89 and 20.18 percent for *Seagull* and *PlaneAndToy(high)*, respectively.

Although the image *Books* has low cross-similarity, the EIs are still highly correlated as a consequence of a more homogeneous scene surface. By using the proposed *image B-coder*, a bit rate reduction of 64.47 percent is gained, as shown in Table II. For the low cross-similar EIs in image *PlaneAndToy(low)*, a smaller bit rate saving of 14.52 percent is obtained, see Table II. The rate-distortion curve is plotted in Fig. 15. Although the EIs in *OpticalTable* are less cross-similar, the bit rate saving is still convincing, which is 45.16 and 32.59 percent for the *image B-coder* and the *image P-coder* shown in Table II.

At a similar bit rate, the proposed scheme outperforms conventional HEVC intra with a significant improved visual quality, which is illustrated in Fig. 13 and Fig. 14 for image *Seagull*. It is further shown that the blockiness artifact is less visible for the proposed scheme at the lower bit rate, see Fig. 14. This is not only because the proposed scheme improves the compression efficiency compared to HEVC intra, but also due to an EI being well predicted from its neighbors by the displacement intra prediction. Therefore, with less energy on the prediction residues, the blockiness distortions from the quantization of the transform coefficients in HEVC are reduced.

Table III additionally illustrates that JPEG2000 performs worse than the original HEVC intra for all the tested images. The rate-distortion curve for *Books* is plotted in Fig. 12. In the figure, the highest bit rate point for JPEG2000 was removed due to a significant bit rate increase to 4.5 bpp at PSNR of 46.73.

2) Full search vs. TZ search: Given a search area of a fixed size, it can be observed in Table II that full search outperforms TZ search in terms of rate-distortion by a large margin,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

TABLE II
BD-PSNR/RATE FOR THE PROPOSED INTRA MODES

Images	P-coder		B-coder	
	BD-PSNR (dB)	BD-rate (%)	BD-PSNR (dB)	BD-rate (%)
Seagull (TZ)	+2.67	-42.89	+3.80	-56.71
Books (TZ)	+4.46	-56.79	+5.26	-64.47
Optic. (TZ)	+2.06	-32.59	+3.01	-45.16
Plane. (high) (TZ search)	+1.83	-20.18	+2.44	-26.64
Plane.(low) (TZ search)	+0.73	-10.85	+0.98	-14.52
Plane.(high) (Full search)	+2.59	-27.59	+2.90	-31.09
Plane.(low) (Full search)	+0.97	-14.11	+1.09	-15.93

TABLE III
BD-PSNR/RATE FOR JPEG2000 COMPARED TO HEVC ORIGINAL INTRA

Images	BD-PSNR (dB)	BD-rate (%)
Seagull (TZ)	-2.62	+59.33
Books (TZ)	-3.55	+83.10
Optic. (TZ)	-2.10	+64.92
Plane. (high)	-2.60	+49.14
Plane. (low)	-2.61	+35.09

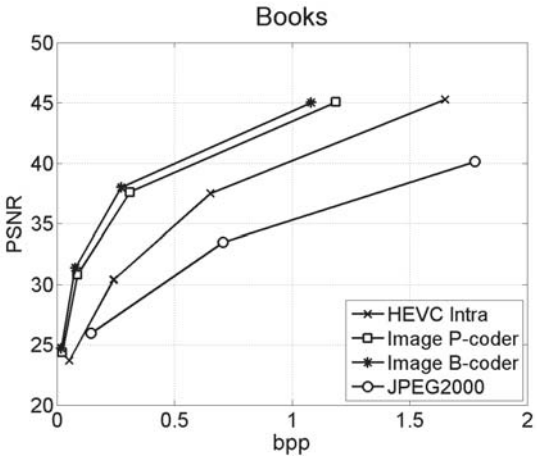
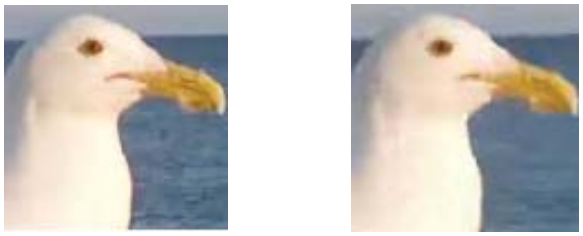


Fig. 12. Rate distortion curve for *Books*.

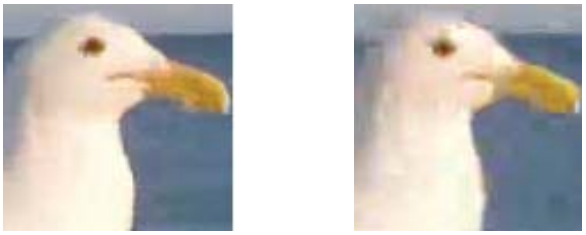
especially in the case of *image P-coder*. The inefficiency of TZ search for the plenoptic contents compared to the natural images [17] is partly because TZ search does not consider and exploit the repetitive patterns.

The time complexity for encoding is presented in Table IV for images *PlaneAndToy(high)*. The results show that the time spent on the *image B-coder* with full search is 12 times more than with TZ search. Approximately the same encoding time ratio is obtained for the *image P-coder*. The order of the encoding complexity can be generalized as: *image B-coder* with full search $>$ *image P-coder* with full search $>>$ *image B-coder* with TZ search $>$ *image P-coder* with TZ search, and the coding efficiency as: *image B-coder* with full search $>$ *image B-coder* with TZ search $\approx$ *image P-coder* with full search $>$ *image P-coder* with TZ search. Therefore, in our opinion, the *image B-coder* with TZ search is a better trade-off between quality and time complexity.



(a) (b)

Fig. 13. Parts of the rendered views from the decoded plenoptic images at bit rate around 0.13bpp: (a) the proposed *image B-coder* (b) original HEVC intra.



(a) (b)

Fig. 14. Parts of the rendered views from the decoded plenoptic images at bit rate around 0.048bpp: (a) the proposed *image B-coder* (b) original HEVC intra.

TABLE IV
ENCODING TIME

QP	P-coder (TZ) (s)	B-coder (TZ) (s)	P-coder (Full) (s)	B-coder (Full) (s)
22	420	579	5727	7035
27	395	539	5757	7201
32	353	496	5763	7363
37	321	460	5783	7256
Average	*372*	*519*	*5758*	*7214*

3) Image B-coder vs. image P-coder (intra mode): Results from Table II portray a similar pattern; the *B-coder* is superior to the *P-coder*. Regardless of the effects of EI borders, the rate-distortion pattern agrees well with the multi-hypothesis prediction analysis described in Section IV that two hypotheses reduce the coded bit rate more than one. However this improvement diminishes with a decreasing cross-similarity between EIs, as is shown for *PlaneAndToy(low)* in Table II. Only 3.7 and 1.8 percent bit rate savings for *B-coder* over *P-coder* are achieved for the TZ search and the full search, respectively. Additionally, the difference in bit rate reduction between *B-coder* and *P-coder* becomes smaller when comparing TZ search to full search for the *PlaneAndToy* images. For the *PlaneAndToy(high)* in Table II, percentage of bit rate differences between *B-coder* and *P-coder* reduces from 6.5 percent with TZ search to 3.5 percent with full search. This reduction is because the TZ search does not explore all the possible searching points while the full search provides the encoder with a higher probability to find a single matched reference block that achieves the best in terms of RDO.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

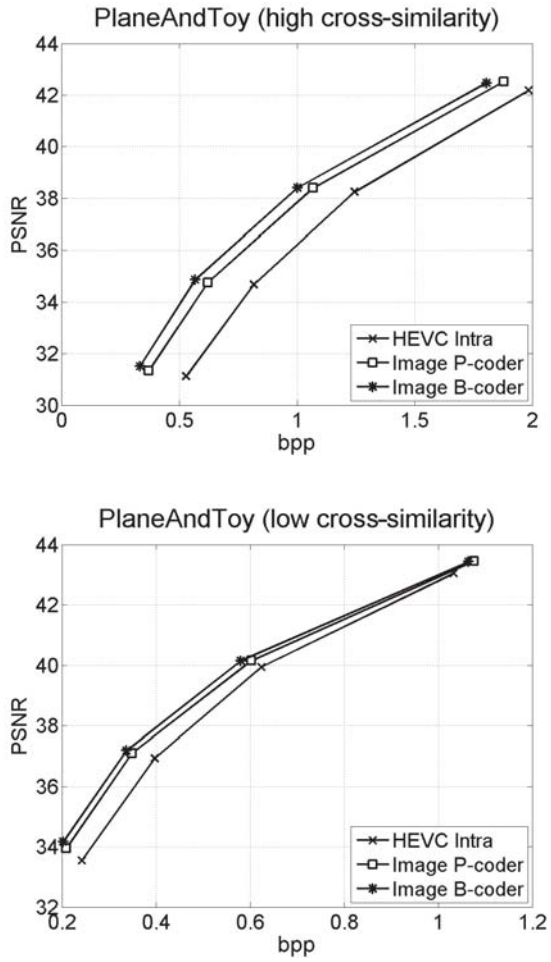


Fig. 15. *PlaneAndToy* with TZ fast search.

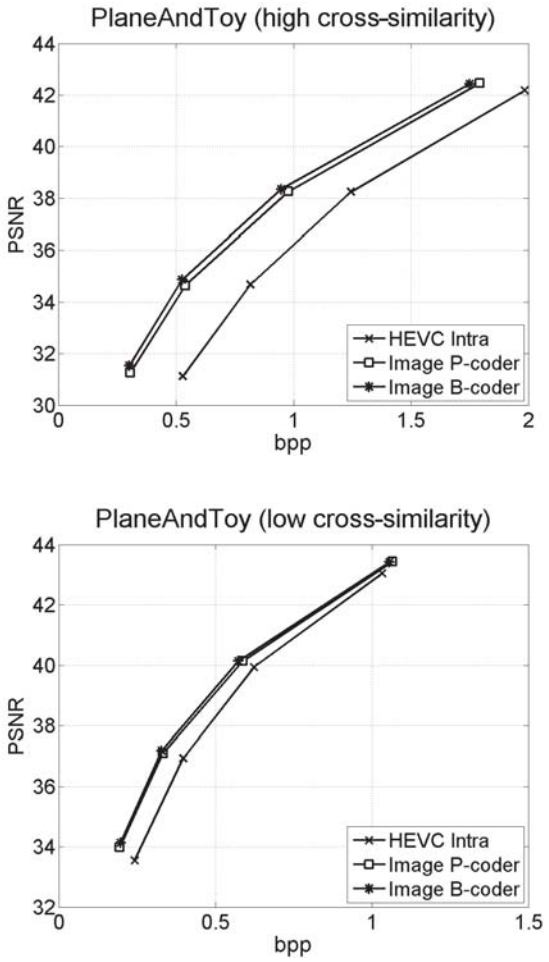


Fig. 16. *PlaneAndToy* with full search.

(4) Employment of different modes: Table V shows that the proposed intra modes are used over 50 percent of the image area for the high cross-similar *PlaneAndToy*. The results were obtained by using *image B-coder*, which includes both the proposed spatial uni-directional prediction and bi-directional prediction into the RDO evaluation in HEVC. As expected, the employment of the proposed modes drops when the cross-similarity between EIs is low. It is less than 20 percent for the low cross-similarity *PlaneAndToy*.

TABLE V
DIFFERENT INTRA MODES USED IN THE *image B-coder*

Contents-PlaneAndToy	**Uni-predicted**	**Bi-predicted**	**HEVC intra**
High cross-similarity (full)	22%	33%	45%
High cross-similarity (TZ)	28%	19%	53%
Low cross-similarity (full)	11%	7%	82%
Low cross-similarity (TZ)	8%	6%	86%

B. Video coding

Even for a scene with a stationary background, 13.7 percent bit rate reduction is obtained with the proposed encoder for the entire *PlaneAndToy* video sequence as shown in Table VI. It is further illustrated in Fig. 17 that the bit rate reduction is more significant at the lower bit rate points. In addition, a larger bit rate saving of 32.01 percent is achieved for the *DemichelisSpark* sequence, see Table VI. Through the examination of coded bit rates frame by frame, we further observe that the main contribution to the coding efficiency is from the intra coded frames.

TABLE VI
BD-PSNR/RATE FOR THE PROPOSED *video B-coder*

Video	**BD-PSNR (dB)**	**BD-rate (%)**
PlaneAndToy	+0.74	-13.70
DemichelisSpark	+0.92	-32.01

IX. CONCLUSION

We have proposed a displacement intra prediction scheme with a maximum of two hypotheses for the compression of plenoptic contents from focused plenoptic cameras. The scheme has been implemented into HEVC and is capable of exploring the inter-microlens redundancy efficiently. Furthermore, we have formulated a signal model for plenoptic images

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

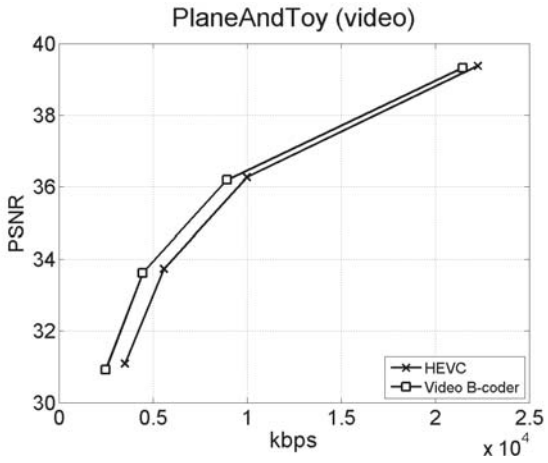


Fig. 17. Rate-distortion curve for *video B-coder* over *PlaneAndToy*.

and explained the theoretical aspects of the displacement intra prediction with multiple hypotheses. In order to assess the rendering artifacts, the impacts of distortions on the compressed captured image to the rendered view were also analyzed.

The results showed that the plenoptic images and videos were compressed efficiently by using the proposed scheme. The compression efficiency is related to several parameters. More cross-similar EIs facilitated the compression for the proposed schemes over HEVC. The TZ search was less effective in searching for prediction candidates for plenoptic contents than for natural images. For the intra modes, although full search can improve the coding efficiency, the *image B-coder* with TZ search is a trade-off between quality and time complexity. In addition, the *image B-coder* has a more extensive bit rate reduction and agrees with the multi-hypothesis analysis. For the tested images, up to 60 percent bit rate reduction was achieved for the proposed scheme compared to HEVC intra, and more than 30 percent was obtained compared to HEVC in temporal mode for the tested video sequences. The visual quality inspection on the rendered views also showed that the proposed schemes outperformed HEVC intra with a better visual quality.

X. Future work

The focus of our future research includes utilizing approximate camera geometries and rendering techniques for prediction to reduce the energy on prediction residues.

Acknowledgment

This work has been supported by grant 20120328 of the Knowledge Foundation, Sweden, by grant 00156702 of the EU European Regional Development Fund, Mellersta Norrland, Sweden, and by grant 00155148 of Länsstyrelsen Västernorrland, Sweden. We also want to acknowledge Todor Georgiev for providing the light field images online.

References

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

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

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

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

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

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

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

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

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

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

[11] R. Ng, M. Levoy, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," *Stanford Tech Report CTSR*, 2005.

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

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

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

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

[16] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "Plenoptic depth map in the case of occlusions," *in Proc. of the International Society for Optics and photonics (SPIE)*, vol. 8667, pp. 86671S–86671S–9, 2013.

[17] X.-l. Tang, S.-k. Dai, and C.-h. Cai, "An analysis of tzsearch algorithm in jmvc," in *Green Circuits and Systems (ICGCS), 2010 International Conference on*, June 2010, pp. 516–520.

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

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

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

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

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

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

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

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2450333, IEEE Transactions on Circuits and Systems for Video Technology

[26] X. Zhu, A. Aaron, and B. Girod, "Distributed Compression of Light Fields," *Stanford University, report, online in Citeseer*.
[27] N. Gehrig and P. Dragotti, "Distributed compression of multi-view images using a geometrical coding approach," *Electronic Engineering*, pp. 421–424, 2007.
[28] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 46:1–46:12, July 2013.
[29] C. Conti, P. Nunes, and L.D. Soares, "New HEVC prediction modes for 3D holoscopic video coding," *2012 19th IEEE International Conference on Image Processing*, pp. 1325–1328, Sept. 2012.
[30] C. Conti, P. Nunes, and L. D. Soares, "3D Holoscopic Video Coding Based on HEVC with Improved Spatial and Temporal Prediction," *Telecommunications-conftele, Castelo Branco, Portugal*, pp. 1–4, May 2013.
[31] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding.," *IEEE Transactions on Image Processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 2, pp. 173–83, Jan. 2000.
[32] M. Flierl and T. Wiegand, "A video codec incorporating block-based multi-hypothesis motion-compensated prediction," *in Proceeding of the SPIE conference on visual communications and image processing, Perth, Australia*, 2000.
[33] M. Flierl, T. Wiegand, and B. Girod, "Rate-constrained multihypothesis prediction for motion-compensated video compression," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 11, pp. 957–969, Nov 2002.
[34] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient Intra Prediction Scheme For Light Field Image Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florance, Italy*, May 2014.
[35] D. Flynn, J. Sole, and T. Suzuki, "High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 4," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-N1005*, April 2013.
[36] A. Aggoun, O. A. Fatah, J. C. Fernandez, C. Conti, P. Nunes, and L.D. Soares, "Acquisition, Processing and Coding of 3D Holoscopic Content for Immersive Video Systems," *3DTV-Conference: Vision Beyond Depth (3DTV-CON), Aberdeen, Scotland*, 2013.
[37] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transaction-Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 2012.
[38] G.J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.
[39] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J.H. Park, "Block partitioning structure in the hevc standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1697–1706, Dec 2012.
[40] "HM reference software version 11 ," https://hevc.hhi.fraunhofer.de/svn/, retrieved: 08, 2013.
[41] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "Hevc complexity and implementation analysis," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1685–1696, Dec 2012.
[42] "Raytrix website," http://raytrix.de/, retrieved: 05, 2014.
[43] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33*, 2001.
[44] F. Bossen, "Common test conditions and software reference configurations," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-L1100*, 2013.
[45] D.S. Taubman and M.W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Norwell, MA, USA, 2001.
[46] "OpenJPEG project," http://www.openjpeg.org/, retrieved: 04, 2015.

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

Mårten Sjöström received the M.Sc. degree in electrical engineering and applied physics from Linkping University, Sweden, in 1992, the Licentiate of Technology degree in signal processing from KTH, Stockholm, Sweden, in 1998, and the Ph.D. degree in modeling of nonlinear systems from EPFL, Lausanne, Switzerland, in 2001. He worked as an Electrical Engineer at ABB, Sweden, from 19931994, was a fellow at CERN from 19941996, and a Ph.D.-student at EPFL, Lausanne, Switzerland during 19972001. In 2001, he joined Mid Sweden University and was appointed Associate Professor and Full Professor in Signal Processing in 2008 and 2013, respectively. He is the head of the subject Computer and System Sciences at Mid Sweden University since 2013. He founded the Realistic 3D research group in 2007. His current research interests are within multidimensional signal processing and imaging, as well as system modelling and identification.

Roger Olsson received the M.Sc. degree in Electrical Engineering and the Ph.D. degree in Telecommunication from Mid Sweden University, Sweden, in 1998 and 2010 respectively. He worked in the video compression and distribution industry 1997-2000. He again joined Mid Sweden University as a junior lecturer 2000-2004 where he taught courses in telecommunication, signals- and systems, and signal- and image processing. Since 2010 he is employed as a researcher at Mid Sweden University where his research interest includes plenoptic image capture, processing, and compression; plenoptic system modelling; and depth map capture and processing.

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