# EFFICIENT DIRECTIONAL AND L1-OPTIMIZED INTRA-PREDICTION FOR LIGHT FIELD IMAGE COMPRESSION

*Rui Zhong[1], Shizheng Wang[2], Bruno Cornelis[1], Yuanjin Zheng[2], Junsong Yuan[2], Adrian Munteanu[1]*
[1]Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel, Brussels, Belgium
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

## ABSTRACT

Light field images can be conveniently captured by consumer-level plenoptic cameras. However, as the resulting data rates are very high, providing efficient compression for this type of data is of critical importance. This remains an open problem which has recently attracted a lot of attention from the coding community. State-of-the-art compression systems prove to be inefficient when directly applied on this type of data due to the inherent spatial discontinuities in light field images. In this paper, a novel intra-prediction method for disk-shaped pixel clusters is proposed. An L1 minimization of the prediction residuals is performed followed by clustering of the predictors, leading to an optimized set of predictors for the macro-pixels. Furthermore, directional intra-prediction modes based on HEVC are devised for the macro-pixels. Experimental results obtained on the EPFL light field image dataset demonstrate that the proposed coding scheme yields an average of 3.22 dB and 1.45 dB gain in PSNR, and 59.6% and 30.88% average rate savings compared to HEVC and the state-of-the-art in light field image coding respectively.

***Index Terms***— light field images, intra prediction, directional mode, L1 optimization, image compression

## 1. INTRODUCTION

As introduced in [1], the concept of light fields refers to the amount of light traveling in every direction through every point in space. In contrast to conventional cameras, which only capture incoming light rays at a given location, plenoptic cameras record the high-dimensional light field data, accounting for both intensity and directional information.

An example of such a device, capturing 4D light field data, is the Lytro plenoptic camera, combining microlens arrays with high-resolution image sensors [2]. Plenoptic cameras provide sufficient information to enable a broad range of applications, such as re-focusing [3], image-based rendering [4], computer graphics [5], [6], free-viewpoint video [7], and many more.

As illustrated in Fig. 1, plenoptic cameras record the directional light intensity distributions onto the image plane using a microlens array. The Lytro microlenses, having a disk shape in the 'microlens plane', produce disk-shaped pixel clusters, sometimes referred to as 'macro-pixels'–see e.g. [8].

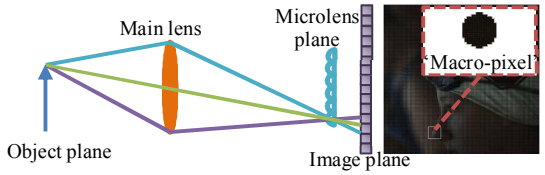


Figure 1. (left) The lens focuses the light from the object onto the microlens plane; (right) a macro-pixel corresponding to a microlens.

The light field image resolution depends on the number of microlenses and the pixel count in each macro-pixel. In Lytro II cameras, there are nearly 200,000 microlenses, with 199 pixels per macro-pixel [9], leading to a vast spatial resolution for light field images. To enable large-scale use of such devices, storing, processing and transmitting of the produced light field images needs to be efficient and user-friendly. These requirements call for the design of an efficient compression system for light field image data.

The HEVC standard is the state-of-the-art in video coding, substantially improving compression performance over all its predecessors [10]. However, the existing HEVC encoder is designed with the assumption of local spatial and temporal continuities in video, which is incompatible with the systematic spatial discontinuities between macro-pixels in light field images.

A novel HEVC-based approach for light field images based on self-similarity compensated prediction was recently presented in [11], in which patch-match based compensated prediction identifies, for a given block, the most similar block from the neighboring reconstructed regions. This compensated prediction method works well for holoscopic images using a regular grid of rectangular macro-pixels. In [12], an HEVC-based coding framework is devised, operating on multiview video that is derived from light field images. This method yields state-of-the-art performance in light field image compression.

In our work [13], we exploit the fact that pixels with the same spatial coordinates within neighboring macro-pixels are spatially correlated. Based on this observation, we proposed an L1-optimized prediction algorithm that linearly predicts the macro-pixels based on the neighboring reconstructed ones. In this paper, novel HEVC based directional intra-modes are designed and a competition between the directional prediction and L1-optimized linear prediction is

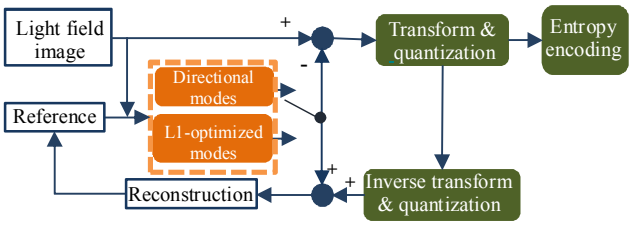


Figure 2. The proposed compression framework whereby directional and L1-optimized prediction modes are competing.

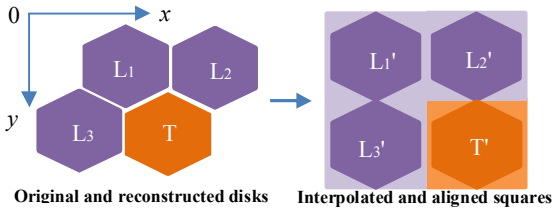


Figure 3. Interpolated and aligned squares: the reference macro-pixels $L_1, L_2, L_3$ are located around the target macro-pixel $T$.

proposed to further reduce spatial redundancies in light field.

In summary, the contributions in this paper are as follows: (i) we establish novel directional intra-modes for the macro-pixels, considered as the coding unit instead of conventional block-based structures used in HEVC; (ii) one proposes an innovative method to fill-in reference samples, accounting for the specific type of circular structures employed in directional intra prediction; (iii) we perform optimized linear prediction by minimizing the L1 residual between the prediction and the original light field image; (iv) one determines the optimized intra-prediction mode among directional and L1-optimized modes; (v) we adapt the HEVC-based coding tools to encode the residuals.

## 2. PROPOSED LIGHT FIELD COMPRESSION SYSTEM COMBINING DIRECTIONAL AND L1-OPTIMIZED LINEAR PREDICTION

In the proposed compression framework, illustrated in Fig. 2, the directional and L1-optimized linear prediction modes are competing in order to minimize the residual.

Directional prediction (Fig. 2) consists of two phases; first, the disk-shaped reference samples are interpolated to form squares; secondly, a directional linear prediction is performed, based on the aligned square-shaped samples. Essentially, the pixels of neighboring reconstructed macro-pixels are placed in the reference buffer to enable weighted pixel-wise prediction associated with different directions.

The L1-optimized linear prediction (Fig. 2) integrates (*i*) an offline phase, whereby the prediction weights expressing the linear relationship between the target macro-pixel and the neighboring macro-pixels are trained, and (*ii*) an online phase, by which the weights are employed for linear prediction. A more detailed description of the different steps is given below.

### 2.1. Directional intra-prediction modes

The main goal of directional intra-prediction is to remove the spatial redundancies by estimating the samples in the target macro-pixel based on a linear prediction from specific reference samples. As mentioned, the inherent spatial discontinuities between neighboring macro-pixels in light field images break the assumption of local spatial continuity which is exploited by the HEVC standard.

Hence, directly applying HEVC's directional intra-prediction is inefficient. To solve this problem, we symmetrically place the reconstructed reference macro-pixels in squares, and interpolate the margin areas, as illustrated in Fig. 3. Moreover, except the DC and Planar modes, which are identical to those in HEVC, new directional modes are proposed to exploit spatial redundancies in various directions. Further details are given below.

*a. Reference samples filling*

In a first step, the reconstructed macro-pixels $L_1, L_2, L_3$ are interpolated to form the $h \times h$ ($h = 17$ for Lytro camera) squares $L_1', L_2', L_3'$, as shown in Fig. 3. To this end, one utilizes vertical and horizontal extrapolation, whereby the pixels in blank areas are copied from the closest boundary pixels in $L_1, L_2, L_3$. For instance, to obtain $L_3'$, a horizontal extrapolation of $L_3$ is performed:

$$L_3'(x,y) = \begin{cases} L_3(x-a(y),y), & x \geq h/2 \\ L_3(x+a(y),y), & x < h/2 \end{cases}, \quad (1)$$

where $a(y)$ is the distance between the pixel $L_3'(x,y)$ and the closest available pixel of macro-pixel $L_3$ at row $y$. A similar extrapolation procedure is performed to determine $L_1'$ and $L_2'$. We note that, in this case, the first step is a vertical extrapolation procedure of $L_1$ and $L_2$ respectively, subsequently followed by a horizontal extrapolation, to cover the pixels which were not determined in the first extrapolation step. The resulting squares $L_1', L_2', L_3'$ offer the necessary reference samples used in directional prediction.

In a second step, directional prediction is performed. As illustrated in Fig. 4, the reference samples include a left-up pixel $r_0$, an above set $C: \{c_1, c_2, \cdots, c_{2h}\}$ and a left set $R: \{r_1, r_2, \cdots, r_{2h}\}$. The reference sets $C$ and $R$ are filled-in by the interpolated pixels of the above, left-up, and left square.

The pixels of the above reference set $\{c_1, c_2, \cdots, c_{2h}\}$ are taken from the bottom line of $L_1'$ and $L_2'$, whereas the pixels of the left reference set are copied from the rightmost column of $L_3'$. Essentially, the above subsets $\{c_1, c_2, \cdots, c_h\}$ and $\{c_{1+h}, c_{2+h}, \cdots, c_{2h}\}$, obtained from $L_1'$ and $L_2'$ respectively, are determined as:

$$c_x = \begin{cases} L_1'(x,h), & 1 \leq x \leq h \\ L_2'(x-h,h), & 1+h \leq x \leq 2h \end{cases}, \quad (2)$$

The left subsets $\{r_1, r_2, \cdots, r_h\}$ and $\{r_{1+h}, r_{2+h}, \cdots, r_{2h}\}$ are computed from $L_3'$ as:

$$r_y = \begin{cases} L_3'(h,y), & 1 \le y \le h \\ L_3'(h,y-h), & 1+h \le y \le 2h \end{cases}, \quad (3)$$

Finally, the left-up pixel $r_0$ is set as the average of $c_1$ and $r_1$.

*b. Directional prediction*

Directional prediction of a sample in $T'$ at location $(x,y)$ - see Fig. 4, is expressed as a matrix multiplication:

$$T'(x,y) = [\boldsymbol{P}_k \cdot \mathbf{C}' + \boldsymbol{Q}_k \cdot \mathbf{R}' + 16]/32 \ . \quad (4)$$

The vectors $\mathbf{C}'$ and $\mathbf{R}'$ denote column vectors collecting $C$ and $R$, i.e. the above and left set of reference samples respectively. The $k^{th}$ parameter sets $\boldsymbol{P}_k \in \mathbb{R}^{1\times 2h}$ and $\boldsymbol{Q}_k \in \mathbb{R}^{1\times 2h}$ ( $k = h\cdot y + x$ ), which contain two non-zero elements, are functions of the pixel's coordinates $1 \le x, y \le h$ and of the angle labeled by $g_i$ for directional mode $i$.

The row vectors $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ are initialized to zero; these vectors contain two nonzero elements located at index $n = x + \lfloor (y \cdot e_i)/32 \rfloor$ and $n+1$, which are assigned values $d_n$ and $d_{n+1} = 32 - d_n$, with $d_n = f(y \cdot e_i)$, where $(y \cdot e_i)$ is the multiplication of the coordinate $y$ and the angular number $e_i$. The function $f$ is defined as:

$$f(z) = \begin{cases} z - 32 \cdot \lfloor z/32 \rfloor, & z > 0 \\ 32 \cdot \lceil |z|/32 \rceil + z, & z < 0 \end{cases} \quad (5)$$

Here, $e_i$ corresponds to an angle $g_i$; for example, the angles $g_2 = H + 32$ and $g_{19} = V - 32$ correspond to $e_2 = +32$ and $e_{19} = -32$ respectively (see Fig. 4). $d_n$ is employed to determine which vector contains the two nonzero elements:

$$\begin{aligned} &d_n \in \mathbf{Q}_k, \forall g_i \in [H-27,\ H+32],\ with\ \ 2 \le i \le 18, \\ &d_n \in \mathbf{P}_k, \forall g_i \in [V-32,\ V+32],\ with\ \ 19 \le i \le 36. \end{aligned} \quad (6)$$

The prediction of the macro-pixel, extracted from the predicted block, is subtracted from original macro-pixel to yield the residual. Subsequently, the optimized directional mode is selected by minimizing the residual of macro-pixel using the mean absolute difference criterion.

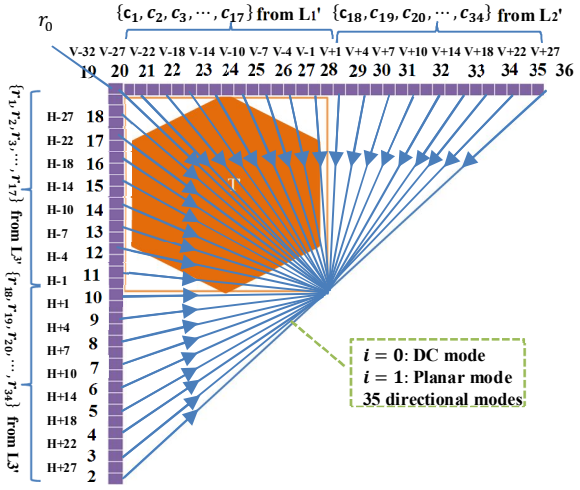


Figure 4. Directional prediction modes for encoding macro-pixel $T$, the directional mode $i \in [2,36]$, the labelled angle $g_i$.

### 2.2. L1-optimized intra-prediction modes

As for L1-optimized modes, a macro-pixel is regarded as the basic prediction unit, lying in contrast to conventional coding paradigms that rely on block-based coding structures. The target macro-pixel $T$ is modelled as a linear combination of its three neighboring reference macro-pixels $L_1, L_2, L_3$, depicted in Fig. 3. The weights $(w_1, w_2, w_3)$, satisfying the obvious constraint that they sum to 1, express the linear relation between the target and reference macro-pixels. By using a standard L1-constrained optimization toolbox [14], the unknown $\mathbf{W} = (w_1, w_2, w_3)$ is computed as solution of the L1 optimization problem:

$$\min_{\mathbf{W}=(w_1,w_2,w_3)} \|\mathbf{T} - \mathbf{W} \cdot \mathbf{L}'\|_1, \quad subject\ to \quad \sum_{p=1}^{3} w_p = 1, \quad (7)$$

where $\boldsymbol{L}' = (L_1, L_2, L_3)'$ denotes the transpose of the matrix consisting of the reference macro-pixels $L_1, L_2, L_3$ surrounding the target $\boldsymbol{T}$ ( $\boldsymbol{T} \in \mathbb{R}^{1\times m}$, $\boldsymbol{W} \in \mathbb{R}^{1\times 3}$, $\boldsymbol{L}' \in \mathbb{R}^{3\times m}$, and $m = 199$ is the number of pixels in a macro-pixel).

The original weights obtained as the solution of (7) are clustered by means of a K-means clustering step [15], which allows for an efficient indexing of the weights according to each of the K cluster centers. During disk-based intra prediction, the $K$ prediction modes are traversed and the best linear prediction mode is determined as follows:

$$\mathbf{W}_{best} = \arg\min_{\mathbf{W} \in \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_K\}} \|\mathbf{T} - \mathbf{W} \cdot \mathbf{L}'\|_1 \ , \quad (8)$$

where $\mathbf{W}_k, 1 \le k \le K$ is the set of trained prediction weights corresponding to intra-mode $k$. One uses $\log_2^K$ bits to index the best linear prediction mode $\mathbf{W}_{best}$ for each target macro-pixel $\boldsymbol{T}$; in our experiments K was set to 32.

Finally, the proposed coding method selects the mode yielding the lowest residual between the directional and L1-optimized modes.

### 2.3. Entropy coding of the intra-prediction modes

The compressed stream makes use of HEVC's syntax elements of intra prediction and residual information [10]; we adapt them to the proposed codec and design the necessary syntax elements. The original syntax elements of HEVC intra prediction consist of the 'CU_skip_flag', the Most Probable Modes, and the block residual coefficients for intra prediction [16], [17]. Here, the skip flag, the directional mode index, and the residual coefficients are encoded using CABAC. Furthermore, the syntax element of L1-optimized modes is combined with that of directional modes named 'mpm_idx'. The range of 'mpm_idx' is modified from $[0,36]$ to $[0,68]$ ( $[37,68]$ belongs to L1-optimized modes). Following the closed-loop coding paradigm, entropy decoding, inverse quantization and transformation, all parts of the encoder, are performed to generate the reconstructed macro-pixels. One notes that entropy coding and decoding are also parts of the loop, that is, to guarantee that, even if the processing unit is changed from a block to a macro-pixel, correct encoding is performed by matching the coder with the decoder.

TABLE I. BD-PSNR OF THE PROPOSED METHOD

| | Proposed Vs HEVC | | Proposed Vs [12] | | Proposed Vs [13] | |
|---|---|---|---|---|---|---|
| | *PSNR gain (dB)* | *Bitrate saving (%)* | *PSNR gain (dB)* | *Bitrate saving (%)* | *PSNR gain (dB)* | *Bitrate saving (%)* |
| I01 | 3.25 | -54.48 | 0.32 | 6.37 | 0.5 | -8.53 |
| I02 | 1.83 | -37.19 | 0.52 | -16.9 | 0.36 | 0.15 |
| I03 | 1.34 | -28.48 | 0.52 | -8.41 | 0.61 | -2.75 |
| I04 | 1.62 | -32.61 | -0.01 | 26.95 | 0.38 | 1.16 |
| I05 | 3.12 | -64.62 | 1.34 | -40.78 | 0.62 | -8.61 |
| I06 | 4.77 | -88.76 | 0.53 | -24.49 | 0.84 | -9.27 |
| I07 | 2.66 | -51.75 | 3.16 | -61.39 | 0.39 | -2.08 |
| I08 | 4.17 | -91.36 | 2.86 | -76.79 | 0.82 | -29.51 |
| I09 | 4.85 | -69.16 | 1.02 | -21.87 | 0.66 | -7.1 |
| I10 | 1.82 | -49.06 | 2.1 | -52.97 | 1.61 | -26.16 |
| I11 | 5.61 | -82.11 | 4.25 | -72.65 | 0.54 | -12.4 |
| I12 | 3.62 | -65.64 | 0.78 | -27.7 | 0.22 | -0.94 |
| **Avg.** | **3.22** | **-59.60** | **1.45** | **-30.88** | **0.63** | **-10.5** |

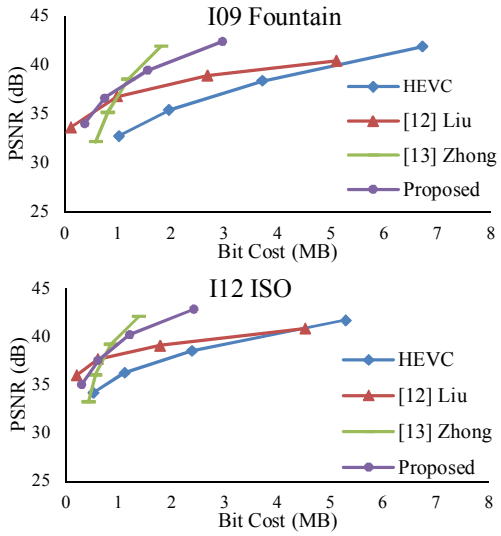


Figure 5. Rate-distortion performance comparison.

## 3. EXPERIMENTAL EVALUATION

### 3.1. Experimental Setup

In the experimental evaluation of the proposed coding system, the conventional evaluation procedure of [18] is followed, whereby the EPFL test set [18] including 12 light field images, having a resolution of 7728x5368 pixels (requiring 51854880 bytes) is employed. We compare the PSNR and cost in bytes of the encoded light field images against the state-of-the-art methods, namely, HEVC operating in intra-mode [16], the pseudo-sequence-based compression of [12], and our previous method in [13]. The experiments are performed using 4 QPs, namely 22, 27, 32, and 37.

### 3.2. Experimental results and analysis

Table I reports the BD-PSNR and BD-BR computed using Bjontegaard's evaluation tools [19]. In Fig. 5, the rate-distortion curves are shown for light field images 'I09_ Fountain' and 'I12_ ISO_Chart_12' from the EPFL dataset. We notice from these results that the PSNR obtained with the proposed method is higher than that of HEVC. Moreover, at low and medium rates, the proposed method reaches much better compression performance compared to our work [13].

The results demonstrate that the proposed compression method achieves high rate savings compared to the state-of-the-art. Overall, the average PSNR gain is 0.63dB, 1.45dB and 3.22dB against [13], [12], and HEVC respectively, corresponding to 10.5%, 30.88%, and 59.6% rate savings respectively. Maximum gains in rate relative to [13] go as high as 29.5%. These large rate savings prove that the proposed macro-pixel-based directional and linear prediction approaches are particularly effective on this type of data.

The lower PSNR at high rates relative to [13] is caused by the additional bit cost incurred by encoding the directional prediction modes. We also have to observe that we perform distortion optimization (and not rate-distortion optimization) to select the best mode which has lower distortion between L1 prediction and directional prediction. At high rates, distortion optimization turns out to not be efficient enough to select the best prediction mode for each macro-pixel.

The complexity of the proposed method is evaluated by the average encoding time. Specifically, the average encoding time for the 4 QPs (22, 27, 32, and 37) is 204 seconds, while the encoding time for HEVC is 6927 seconds, which is approximately 28 times higher compared to that of the proposed method. The reason is that in HEVC, the best coding unit is selected between $H$ block sizes ($64\times64, 32\times32, \cdots, 4\times4$). For each coding unit, distortion optimization is used to obtain the best mode from intra directional predictions, which spends time on the calculation of block-wise distortion $O(m)$ and bit cost $O(n)$. The time cost is $H\times(O(m)+O(n))$. In the proposed method, the coding unit is fixed to be a macro-pixel, and the block-wise distortion is the only cost in time complexity, which is of the order $O(m)$.

## 4. CONCLUSIONS

A novel compression system for light field image data has been proposed in this paper. Our approach exploits the spatial correlation amongst neighboring disk-shaped pixel clusters corresponding to each microlens in the light field image. We capture these correlations by assuming a linear dependency model between a target macro-pixel and its neighboring reference macro-pixels in the lenslet image. To further improve encoding efficiency, we propose new directional modes for this type of data. The competition between the directional and the L1-optimized intra prediction increases the efficiency of the subsequent residual encoding step performed using an adapted HEVC codec. The experimental results demonstrate that the proposed coding method achieves significantly higher PSNR and particularly higher rate savings compared to the state-of-the-art.

## 5. REFERENCES


[1] A. Gershun, "The light field," Journal of Mathematics and Physics XVIII, pp.51–151, 1936.

[2] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Computer Science Technical Report CSTR, 2(11), pp.1-11, 2011.

[3] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," Procceedings of SPIE, vol. 8667, 2013.

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

[5] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays," ACM Transactions on Graphics, vol. 30, no. 4, July 2011.

[6] K. Changil, K. Subr, K. Mitchell, A. Sorkine-Hornung, and M. Gross, "Online view sampling for estimating depth from light fields," IEEE International Conference on Image Processing, ICIP 2015, pp. 1155-1159, 2015.

[7] J. Carranza, C. Theobalt, M.A. Magnor, H.-P. Seidel, "Free-viewpoint video of human actors," ACM SIGGRAPH 2003, pp. 569-577, 2003.

[8] G. Wetzstein, I. Ihrke, W. Heidrich, "On plenoptic multiplexing and reconstruction," International Journal on Computer Vision, vol. 101, no. 2, pp. 384-400, 2013.

[9] G. Wetzstein, "I2. 1: Invited Paper: On the duality of compressive light field imaging and display," SID Symposium Digest of Technical Papers, vol. 46. no. 1. 2015.

[10] G.J., Sullivan, J.-R. Ohm, W. 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.

[11] C. Conti, L.-D. Soares, and P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction," Signal Processing: Image Communication, vol. 42, pp. 59-78, 2016.

[12] D. Liu, L.Z. Wang, L. Li, Z.W. Xiong, F. Wu, W.J. Zeng, "Pseudo-sequence-based light field image compression," Multimedia & Expo Workshops (ICMEW), IEEE International Conference on. IEEE, pp. 1-4, 2016.

[13] R. Zhong, S.Z. Wang, B. Cornelis, Y.J. Zeng, J.S. Yuan, A. Munteanu, "L1-Optimized Linear Prediction for Light Field Image Compression," Picture Coding Symposium, 2016.

[14] J. Duchi, "L1-norm: Methods for Convex-Cardinality Problems", Stanford Technical Report, 2008.

[15] J. A. Hartigan, M. A. Wong, "Algorithm AS 136: A K-means Clustering Algorithm", Journal of the Royal Statistical Society, Series C 28, no. 1, pp.100-108, 1979.

[16] "High Efficiency Video Coding", Recommendation ITU-T H.265/International Standard ISO/IEC 23008-2, 2015.

[17] Joint Collaborative Team on Video Coding (JCT-VC), "HEVC reference software, HM version 16.8," https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.

[18] M. Rerabek, L. Yuan, L. Authier, and T. Ebramini, "EPFL light-field image dataset," ISO/IEC JTC1/SC 29/WG1 69th Meeting, http://mmspg.epfl.ch/EPFL-light-field-image-dataset, 2015.

[19] G. Bjontegaard, Calculation of average PSNR differences between RD-curves, VCEG Contribution VCEG-M33, Austin, April 2001.