Light Field Compression with Homography-based Low Rank Approximation

Xiaoran Jiang, Mikaël Le Pendu, Reuben A Farrugia, Christine Guillemot

Light Field Compression with Homography-based Low Rank Approximation

Xiaoran Jiang, Mikaël Le Pendu, Reuben A. Farrugia, Christine Guillemot *Fellow, IEEE*

***Abstract*—This paper describes a light field compression scheme based on a novel homography-based low rank approximation method called HLRA. The HLRA method jointly searches for the set of homographies best aligning the light field views and for the low rank approximation matrices. The light field views are aligned using either one global homography or multiple homographies depending on how much the disparity across views varies from one depth plane to the other. The light field low-rank representation is then compressed using HEVC. The best pair of rank and QP parameters of the coding scheme, for a given target bit-rate, is predicted with a model defined as a function of light field disparity and texture features. The results are compared with those obtained by directly applying HEVC on the light field views re-structured as a pseudo-video sequence. The experiments using different data sets show substantial PSNR-rate gain of our compression algorithm, as well as the accuracy of the proposed parameter prediction model, especially for real light fields. A scalable extension of the coding scheme is finally proposed.**

***Index Terms*—Light fields, Low rank approximation, Homography, Compression.**

I. Introduction

LIGHT field (LF) imaging has emerged as a very promising technology in the field of computational photography. Many acquisition devices have been recently designed to capture light fields, going from arrays of cameras capturing the scene from slightly different viewpoints [1], to single cameras mounted on moving gantries and plenoptic cameras. Plenoptic cameras are becoming commercially available using arrays of micro-lenses placed in front of the photosensor to obtain angular information about the captured scene [2], [3]. Compared to classical 2D imaging, light fields capture the intensity values of light rays interacting with the scene. The recorded flow of rays is in the form of large volumes of data retaining both spatial and angular information of a scene which enables a variety of post-capture processing, such as re-focusing, extended focus, different viewpoint rendering and depth estimation, from a single exposure [1], [2], [4]. For a comprehensive overview of light field image processing techniques, please refer to [5].

Given the very large volume of high-dimensional data, the design of efficient compression schemes of light fields is a key challenge for practical use of this technology. First methods for compressing synthetic light fields appeared in the late 90's essentially based on classical coding tools as vector quantization followed by Lempel-Ziv (LZ) entropy coding [4] or wavelet coding as in [6] and [7], yielding however limited compression performances (compression factors not exceeding 20 for an acceptable quality). Predictive schemes inspired from video compression methods have also been naturally investigated, adding specific prediction modes in schemes inspired from H.264 and MVC, as in [8] and [9]. Motivated by the objective of random access and progressive decoding, which is not enabled by predictive schemes, the authors in [10] describe another approach using a wavelet transform applied in the 4 dimensions of the light field. The scheme naturally inherits the scalable and progressive properties of wavelet-based coding schemes. The authors in [11] instead use a Principal Component Analysis (PCA) to enable random access to pixels and to support scalability.

The compression of real light fields has recently gained attention thanks to the emergence of capturing devices (plenoptic cameras or rigs of cameras). In this paper, we focus on the compression of real light fields captured by plenoptic cameras. Prior work in this area has followed two main directions: either coding the array of sub-aperture images extracted from the lenslet image as in [12]–[14], or directly encoding the lenslet images captured by plenoptic cameras [15]–[20] with an extension of HEVC with dedicated prediction modes.

Instead of directly encoding the light field (the array of sub-aperture images or the lenslet image for light fields captured by plenoptic cameras), the authors in [21] consider the focus stack as an intermediate representation of reduced dimension of the light field and encode the focus stack with a wavelet-based scheme. The light field is then reconstructed from the focus stack using the linear view synthesis approach described in [22].

In this paper, we describe a light field compression algorithm based on a low rank approximation exploiting scene and data geometry. In particular, light fields with a dense angular sampling such as those captured by plenoptic cameras are addressed. We consider the coding of the sub-aperture images (*i.e.* views) already extracted from a lenslet image. Thanks to the high correlation between the views in such light fields, the matrix whose columns are formed by vectorizing each view can be well approximated by a low rank matrix. In addition, a prior alignment of the views with homography warpings increases the correlation and thus improves the low rank approximation. In the proposed method, homography projections are searched for each view in order to obtain the best low rank matrix approximation for a given target rank k (where k is less than the number of views). This joint homography alignment and low rank optimization procedure is illustrated in Fig. 1. In the cases where the scene contains several layers of depth, the method has also been extended to

This work has been supported in part by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

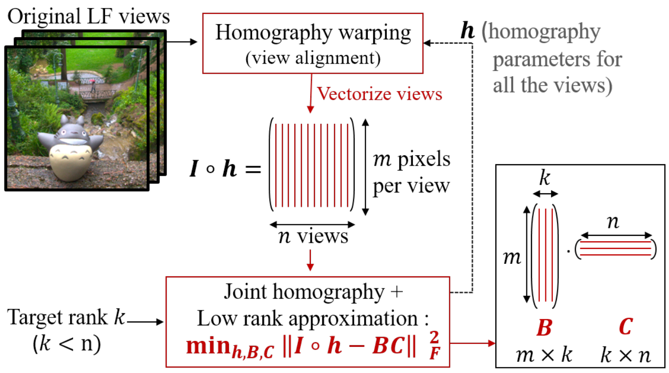


Fig. 1. Overview diagram of the HLRA method.

search for one homography for each depth plane. To cope with artifacts at the frontier of two depth planes when performing the homography warpings, a blending of homographies is performed.

The obtained rank k matrix is expressed as a product of a matrix B, containing k basis vectors, with a matrix C containing weighting coefficients as shown in Fig. 1. This decomposition enables an efficient encoding where the k images formed by reshaping each column of B are compressed with HEVC-Intra. The weights contained in C and the homography parameters are also necessary to reconstruct the light field, hence are transmitted using classical entropy coding. Note that this low rank approximation method is based on our earlier work described in [14].

The PSNR-rate performance of the compression scheme depends on two key parameters: the approximation rank and the quantization parameter (QP) of the HEVC encoder. Using a set of training light fields, we learn a model to predict the pair of rank and QP parameters yielding the best PSNR for a given target bit-rate. The model is learned as a function of several input light field features: disparity indicators defined as a function of the decay rate of the SVD values of the original and registered view matrices, as well as texture indicators defined in terms of the decay rate of SVD values computed on the central view. The parameter prediction problem is cast as a multi-output classification problem solved using a Decision Tree ensemble method, namely the Random Forest method [23].

A scalable extension is finally described in which the homography-based low rank model is used to compress the base layer. The reconstructed base layer is used to predict the light field, and the residue for each light field view is encoded by using the proposed low rank based scheme, with or without view alignment.

Our experiments show the advantage of using the proposed joint alignment and low rank optimization rather than first aligning the views independently of the targeted rank. Furthermore, the compression performances of the proposed scheme are assessed against those obtained with two methods applying HEVC inter-coding on pseudo-sequences of sub-aperture images [12], [13]. In the sequel, the method in [12] will be referred to as HEVC-lozenge, while the method in [13] will be called HEVC-pseudo. Thanks to the robustness of our method to noise and color variations across views, substantial PSNR-rate gains are obtained for two different datasets of real light fields: the INRIA dataset [24] and that of the ICME 2016 Grand Challenge (Light-Field Image Compression) [25].

II. Light Fields: Background and Notations

The light rays emitted by a scene and received by an observer at a particular point (x, y, z) in space, at a given instant t, can be described by the 7D plenoptic function $L(x, y, z, \theta, \phi, t, \lambda)$ where (θ, ϕ) are angles giving the orientation of the light rays and λ their wavelength. For a static light field, the $7D$ plenoptic function can be simplified into a 4D representation called 4D light field in [4] and lumigraph in [26], describing the radiance along rays by a function $L(x, y, u, v)$ of 4 parameters at the intersection of the light rays with 2 parallel planes. This simplification is done assuming constant radiance of a light ray from point to point, and given that an R, G, B sampling of the wavelength is performed by the color filters coupled with the CCD sensors.

The light field can be seen as capturing an array of viewpoints (called sub-aperture images in particular in the case of micro-lens based capturing devices) of the imaged scene with varying angular coordinates u and v. The different views will be denoted here $I_{u,v} \in \mathbb{R}^{X \times Y}$, where X and Y represent the vertical and horizontal dimensions of each sub-aperture image. Each sub-aperture image corresponds to a fixed pair of (u, v) coordinates. In the following, the notation $I_{u,v}$ for the different views (or sub-aperture images) is simplified as I_i with a bijection between (u, v) and i. The complete light field can hence be represented by the matrix $I \in \mathbb{R}^{m \times n}$:

$$I = [\text{vec}(I_1) \mid \text{vec}(I_2) \mid \ldots \mid \text{vec}(I_n)]\,, \qquad (1)$$

with $\text{vec}(I_i)$ being the vectorized version of the sub-aperture image I_i, and where m is the number of pixels in each view $(m = X \times Y)$ and n is the number of views in the light field.

III. Related work

A. Low-rank approximation

Many algorithms taking advantage of the low rank property of a matrix have been developed in the recent literature. For instance, the methods in [27]–[34] tackle the problem of completing a low rank matrix from a subset of its entries. A closely related problem referred to as robust principal component analysis (RPCA) consists in decomposing a matrix as the sum of a low rank and a sparse matrix. The authors in [35] solved this problem using an augmented Lagrangian multiplier method. Note that although those problems are NP-hard in general and involve non-convex optimization (the rank of a matrix being non-convex), recent theoretical papers [36]–[38] have shown that under surprisingly broad conditions, a convex relaxation replacing the rank by the nuclear norm leads exactly to the same unique solution.

Inspired by the RPCA, a low rank approximation model has been considered in the RASL method [39] for aligning correlated images. In their model, each input image is warped by a homography projection. The homography parameters are determined in order to optimize the low rank and sparse decomposition of the matrix formed by concatenating the vectorized warped images. In the context of light filed compression, however, this sparse and low rank decomposition approach does not necessarily yield a compact representation of the original light field. Although useful for ignoring inconsistent pixels between views in the RASL alignment, the sparse term still contains important visual features of the light field (*e.g.* specular light, disocclusion) which essentially consist of high frequencies. Therefore, encoding this term is likely to degrade the overall coding performance. The authors in [40] have nevertheless shown the efficiency of such a compression scheme for encoding videos captured with a fixed camera. In that case the sparse term only captures the moving objects which can be efficiently compressed with a classical video encoder. The static background is essentially contained in the low rank term with a typical rank of 1 or close to 1, and can be encoded at a very low cost.

Note also that for encoding efficiently a low rank matrix A of size $m \times n$ and of rank $k \ll \min(m,n)$, a preliminary factorization step of the form $A = BC$ (where $B \in \mathbb{R}^{m \times k}$ and $C \in \mathbb{R}^{k \times n}$) must be performed. Such a factorization can be obtained with a singular value decomposition. Alternatively, in the SLRMA compression method [41], a similar factorization is found by solving an optimization problem which additionally constrains the sparsity of the matrix B in a given dictionary.

B. Light fields compression

In this section, we focus on prior work dealing with the compression of light fields captured by plenoptic cameras. The methods proposed in the literature can be classified into two categories: those which aim at directly compressing the raw lenslet data after de-vignetting and demosaicing (*e.g.*, [15]–[20]) and those which compress the sub-aperture images extracted from the lenslet data (*e.g.*, [12], [13]).

Most solutions proposed for directly encoding the lenslet data aim at exploiting spatial redundancy or self-similarity between the micro-images. The micro-image is the set of pixels behind each micro-lens, and is also sometimes called elemental image. Spatial prediction modes have thus been proposed for unfocused cameras in [15] and [42] based on a concept of self-similarity compensated prediction or using locally linear embedding techniques in [43]. These self-similarity prediction modes have been further extended to bi-directional prediction in [18] and [19]. The authors in [16] introduce a bi-directional spatial prediction mode in HEVC for encoding elemental images captured by a focused 2.0 camera [44] which has been further extended in [20]) for unfocused 1.0 cameras. While the elemental images (EI) in the focused 2.0 plenoptic cameras can be seen as a cropped multi-view image from one viewing angle, the EI produced by the unfocused 1.0 cameras capture angular information of one point in space. A scalable extension of HEVC-based scheme is also proposed in [45] where a sparse set of micro-lens images (also called elemental images) is encoded in a base layer. The other elemental images are reconstructed at the decoder using disparity-based interpolation and inpainting. The reconstructed images are then used to predict the entire lenslet image and a prediction residue is transmitted yielding a multi-layer scheme. The authors in [17] instead partition the raw light field data into tiles which are then encoded as a pseudo-video sequence using HEVC.

A second category of method consists in encoding the set of sub-aperture images (or views) which can be extracted from the lenslet images after de-vignetting, demosaicing and alignment of the micro-lens array on the sensor, following the raw data decoding pipeline described in [46]. The author of [47] exploits inter-view correlation by using homography and 2D warping to predict views. Homographies are computed via Random Sample Consensus (RANSAC) [48]. The authors in [12] form a pseudo-sequence by using a lozenge scanning order and encode this pseudo-sequence using HEVC inter-coding. In [13], a coding order and a prediction structure inspired from those used in the multi-view coding (MVC) coding standard is proposed, showing significant performance gains compared with HEVC-Intra.

Here, we consider instead a very different approach which aims at reducing the dimensionality of the data using low rank approximations prior encoding.

IV. Homography-based Low Rank Approximation

The error introduced by the low rank approximation model depends on how well the sub-aperture images are aligned. We hence propose to search for the homographies minimizing the low rank approximation error for a targeted rank.

Let I_i and I_j be two sub-aperture images for which we assume there exists an invertible homography transformation h_i, such that

$$I_j(x,y) = (I_i \circ h_i)(x,y) = I_i(h_i(x,y)). \tag{2}$$

A homography transformation h_i can be characterized by a 3×3 matrix H_i which transforms each coordinates (x,y) in I_i into the coordinates $(\frac{x_H}{w_H}, \frac{y_H}{w_H})$, where

$$[x_H, y_H, w_H]^\top = H_i \cdot [x, y, 1]^\top . \tag{3}$$

However, without loss of generality, the last element $H_{i(3,3)}$ can be fixed to 1. The eight remaining elements are then sufficient to parametrize the homography.

Let h be the set of homographies associated to each view of the light field. In what follows, we will consider h as the matrix $[h_1 \mid ... \mid h_n]$ where $h_1, ..., h_n$ are vectors of size 8×1 whose elements are the homography parameters. The low rank optimization problem is then formulated as

$$\text{argmin}_{h,B,C} \left\| I \circ h - BC \right\|_F^2 , \tag{4}$$

where $\|.\|_F$ is the Frobenius norm, $B \in \mathbb{R}^{m \times k}$, $C \in \mathbb{R}^{k \times n}$ $(k < n)$, and $I \circ h$ stands for the matrix containing all views aligned using homographies $h_1, ...h_n$ and can be written as

$$I \circ h = [\text{vec}(I_1 \circ h_1) \mid ... \mid \text{vec}(I_n \circ h_n)] . \tag{5}$$

Note that for aligning all the views, only $n-1$ homographies would be sufficient assuming that one view (*e.g.* the central view) is not warped and is used as reference for aligning the other views. However, this would require constraining the minimization problem (4) to ensure that the homography of the central view is equal to the identity. For simplicity, we did not consider such a constraint in our formulation.

A. Linear Approximation

Minimizing Eq. (4) is non trivial due to the non linearity of the term $I \circ h$. Nevertheless, when the change in h is small, we can approximate it by local linearity as follows:

$$I \circ (h + \Delta h) \approx I \circ h + \sum_{i=1}^{n} J_i \Delta h_i \epsilon_i^\top, \tag{6}$$

where $\Delta h = [\Delta h_1 \mid ... \mid \Delta h_n]$, J_i is the Jacobian matrix of the warped and vectorized sub-aperture image, $\text{vec}(I_i \circ h_i)$, with respect to the parameters of h_i (*i.e.* $J_i = \frac{\partial}{\partial \zeta}\text{vec}(I_i \circ \zeta)|_{\zeta = h_i}$). And ϵ_i is a $n \times 1$ vector with element i equal to 1 and all the other elements equal to 0.

B. Iterative minimization

The minimization problem in Eq. (4) is iteratively solved by updating alternatively the matrices B and C and the homographies $h_1, ..., h_n$. Each homography h_i is first initialized so that the corresponding 3×3 matrix H_i is equal to the identity.

- Given h fixed, B and C are found by computing the singular value decomposition $I \circ h = U \Sigma V^\top$. Then B is set as the k first columns of $U\Sigma$ and C is set as the k first rows of $V^\top$, so that BC is the closest k-rank approximation of $I \circ h$.
- h is updated by solving Eq. (4) for B and C fixed. By noting the updated homography parameter matrix $h' = h + \Delta h$, the minimization Eq. (4) becomes:

  $$h' = h + \underset{\Delta h}{\text{argmin}} \|I \circ (h + \Delta h) - BC\|_F^2 \tag{7}$$

  Given the approximation in Eq. (6), the problem is then to find:

  $$\Delta h = \underset{\Delta h}{\text{argmin}} \|I \circ h - BC + \sum_{i=1}^{n} J_i \Delta h_i \epsilon_i^\top \|_F^2 \tag{8}$$

  This problem is independently solved for each column Δh_i of Δh and can be equivalently stated as:

  $$\forall i,\ \Delta h_i = \underset{\Delta h_i}{\text{argmin}} \|(I \circ h - BC)\epsilon_i + J_i \Delta h_i\|_F^2 \tag{9}$$

  This is a linear least squares problem with solution:

  $$\forall i,\ \Delta h_i = J_i^\dagger (BC - I \circ h)\epsilon_i \tag{10}$$

where $J_i^\dagger$ denotes the Moore-Penrose pseudoinverse of J_i.

C. Recalculating C to account for quantization errors in B

Since the matrix B will need to be compressed to be transmitted to the receiver side, the receiver will obtain a matrix B' with compression artifacts. To reduce the impact of the compression (*i.e.* quantization) errors on the light field reconstruction, the matrix C is recalculated to account for these quantization errors, as follows:

$$C' = \underset{C}{\text{argmin}} \|I \circ h - B'C\|_F^2 = (B')^\dagger (I \circ h). \tag{11}$$

In practice, this adaptation of C to the compression artifacts of the matrix B can increase the PSNR by about 1 dB when strong compression is applied (*e.g.* QP = 38). Fig. 2 shows the PSNR gain of this adaptation for the light field “TotoroWaterfall” (*cf.* Fig. 6). The PSNR gains are shown for the case where one homography per view is applied.

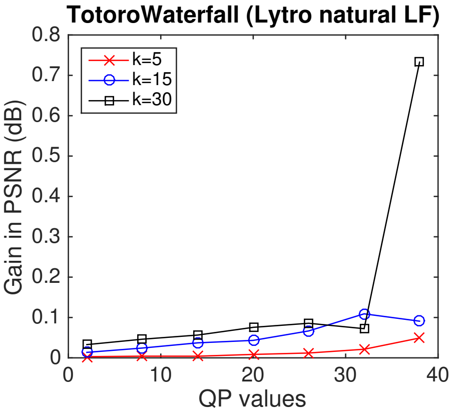


Fig. 2. PSNR gain when the matrix C is re-calculated to account for the compression errors of the matrix B. Alignment has been performed using one homography per view.

V. Low Rank Approximation using multiple homographies

When the disparity varies from one depth plane to another, one global homography per view is not sufficient to well align the whole views. For such light fields, the above homography-based low rank approximation method can be extended to the case where different homographies are computed for different depth planes segmented thanks to a scene depth map.

In order to avoid the high transmission cost of one depth map per view, we instead consider a single depth map of the central view which can be directly estimated using methods as proposed in [49]–[51]. The same depth map is then used for all the views. Note that for the type of light fields addressed, with a dense angular sampling but a limited baseline, using only the depth map of the central view results in a limited loss of accuracy for extracting the depth planes and for computing the corresponding homographies.

Each value in the depth map D is normalized between 0 and 1, indicating if the corresponding pixel is close to the camera (close to 0) or not (close to 1). q depth planes are then obtained by uniformly quantizing D with quantization thresholds $\{\sigma_p\}_{p \in [\![1,q]\!]}$. The thresholds are defined to split in equal parts the range of depth values between the minimum and maximum depth in D. Each depth plane p is characterized by a mask $M^p = \mathbb{1}_{]\sigma_p, \sigma_{p+1}]}(D)$, where $\mathbb{1}$ is the pixel-wise indicator function.

We apply one homography to each depth plane p of the sub-aperture image I_i. In what follows, we note h_i^p the vector containing the corresponding homography parameters. Blending

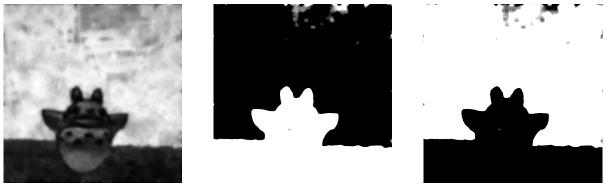


Fig. 3. Separation on two depth planes: depth map (the left column) and the weights w^1 and w^2 associated to each depth plane (the middle and the right columns).

is required to naturally mix the depth planes. Here, instead of blending the pixel values, we blend the homographies, which yields less artifacts at the frontier of depth planes. For that purpose, at each pixel coordinate (x, y), we define the series of weights $\{w^p_{(x,y)}\}_{p \in [\![1,q]\!]}$ that determine the importance of the homography of each depth plane p for this pixel:

$$w^p_{(x,y)} = \begin{cases} 1, \text{if } D_{(x,y)} \in [\sigma_p + \delta, \sigma_{p+1} - \delta]\,; \\ \frac{D_{(x,y)} - (\sigma_p - \delta)}{2\delta}, \text{if } |D_{(x,y)} - \sigma_p| < \delta; \\ \frac{(\sigma_{p+1} + \delta) - D_{(x,y)}}{2\delta}, \text{if } |D_{(x,y)} - \sigma_{p+1}| < \delta; \\ 0, \text{otherwise}, \end{cases} \tag{12}$$

with δ a shallow neighborhood where the blending is applied. Fig. 3 shows the weights w^1 and w^2 associated to each depth plane for the LF “TotoroWaterfall”.

The warping (Eq. (3)) is then modified as follows:

$$[x_H, y_H, w_H]^\top = \sum_{p=1}^{q} w^p_{(x,y)} H^p_i \cdot [x, y, 1]^\top\,. \tag{13}$$

Once the warped images $I \circ h$ are obtained by applying Eq. (13) for each sub-aperture image and each depth plane, we compute B and C at each iteration exactly as described in Section IV. Note that we must now determine nq homographies. The size of h is then $8 \times nq$. Similar to Eq. (10), each vector of homography parameters h^p_i is updated by adding Δh^p_i computed as:

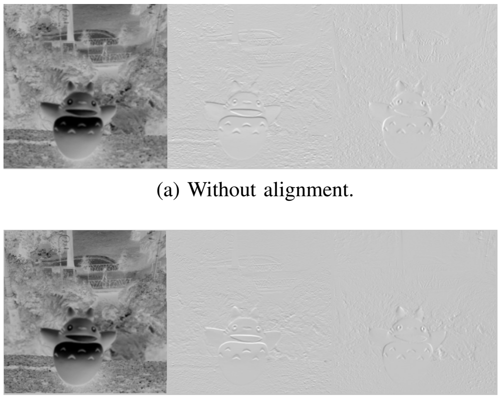
$$\forall i, p,\ \Delta h^p_i = J^{p\dagger}_i \left[(BC - I \circ h)\epsilon_i \odot \text{vec}(M^p)\right], \tag{14}$$

with M^p the corresponding binary mask of depth plane p and $\odot$ the Hadamard product.

VI. Compression algorithm

A. Algorithm overview

The different steps of the complete compression algorithm are shown in Fig. 5. The low rank representation is compressed by encoding the columns of the matrix B using HEVC Intra coding. However, any encoder could be used to compress the proposed homography-based low rank representation. The columns of the matrix B are first quantized on 16 bits before being encoded using HEVC-Intra coding. The first three columns of the matrix B for the LF “TotoroWaterfall” are shown in Fig. 4. The first column represents low frequency information, whereas the others contain data with high frequency. One can observe that by using homographies to align sub-aperture views, the average image in the first column becomes sharper, and there is less high frequency information remaining in the following columns.



(a) Without alignment.

(b) With alignment, 1 homography per view, $k = 5$.

Fig. 4. First three columns of the matrix B (“TotoroWaterfall” LF).

Besides the matrix B, additional elements need to be transmitted. The coefficients of the matrix C of size $k \times n$, where k and n are the approximation rank and the number of views, are encoded using a scalar quantization on 16 bits and Huffman coding. The $8 \times n \times q$ homography parameters, with q the number of depth planes per view, are encoded the same way. In the case where multiple homograhies are applied, depth-planes need to be segmented. So that the decoder can find the depth planes, as explained in Section V, one depth map is encoded using 8 bit quantization followed by HEVC-Intra (we used QP=32 in the experiments). Percentages of bitrate cost for the different elements are analyzed in Section VII-A4.

B. Model-based coding parameters prediction

For a given target bit-rate and a given input light field, the PSNR performance of the compression scheme depends on two key parameters: the rank k of the approximation and the HEVC quantization parameter (QP). To automatically select the best pair of parameters (k, QP), we train a model represented by a function f of a set of input features:

$$(k, \text{QP}) = f(\text{LF features}, \text{Target bitrate}) \tag{15}$$

To generate training data labels, we encode at first several light fields with different values of k and QP, and then labels are extracted by taking the (k, QP) pairs only corresponding to the data points on the envelope of the PSNR-rate points.

1) Feature space: The parameter prediction is regarded as a classification problem with the following input features:

- Disparity indicators of original light field:
  - proportion of singular values of the matrix I which contain at least 95% of the energy of I;
  - decay rate of singular values of the matrix I which is defined as the ratio between the first and the second singular value.
- Disparity indicators of aligned light field: same indicators as above for the matrix $I \circ h$.
- Texture indicators: same indicators as above computed on the matrix in which each column is a vectorized version of each 8×8 block of the central view.
- Bitrate of the encoded light field for a certain pair (k, QP). This feature gives some indication of the bitrate range for

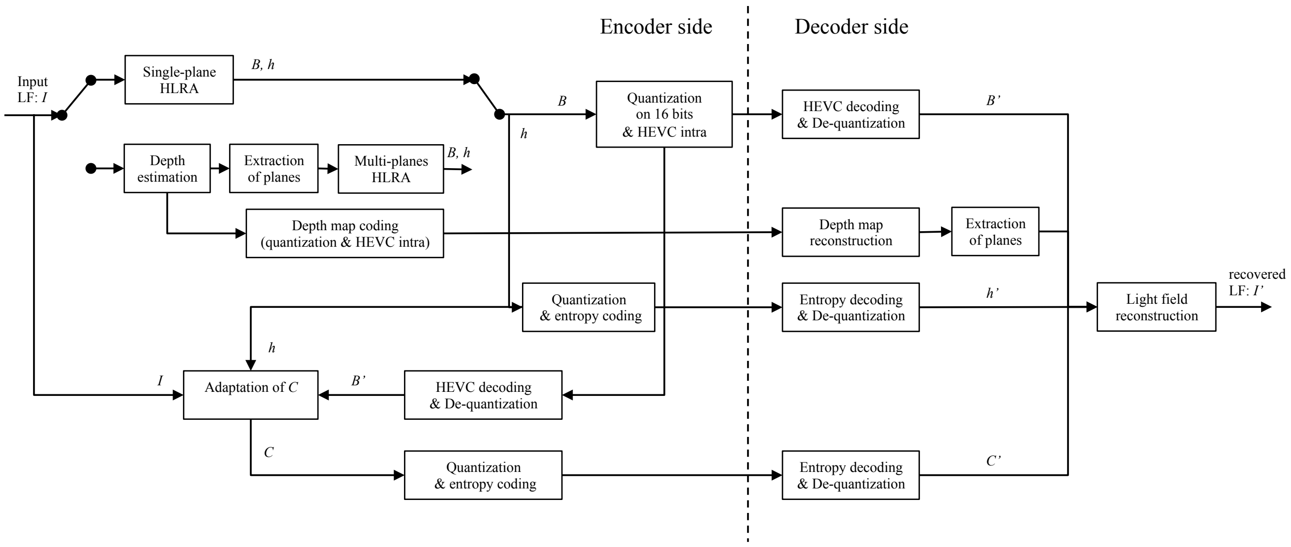

Encoder side
Decoder side
Input LF: I
Single-plane HLRA
B, h
Depth estimation
Extraction of planes
Multi-planes HLRA
B, h
B
h
Quantization on 16 bits & HEVC intra
Depth map coding (quantization & HEVC intra)
HEVC decoding & De-quantization
B'
Depth map reconstruction
Extraction of planes
Quantization & entropy coding
Entropy decoding & De-quantization
h'
Light field reconstruction
recovered LF: I'
h
I
Adaptation of C
B'
HEVC decoding & De-quantization
C
Quantization & entropy coding
Entropy decoding & De-quantization
C'

Fig. 5. Coding and decoding scheme overview.

the input light field. In the experiments, we take $k = 15$, QP $= 14$ and encode the original LF (without alignment for a fast computation). Note that it requires nevertheless a supplementary encoding pass.

- Target bitrate.

2) Decision trees: The model input features are a mixture of continuous (*e.g.* bitrates and decay rate of singular values) and discrete (*e.g.* proportion of singular values keeping 95% of the energy) variables. Contrary to SVM or Logistic Regression which are only efficient with continuous inputs, Decision Trees (DT) are good candidates for dealing with data of different nature and do not need variable scaling. DT is also known to be robust to noisy data (outliers). Among models of DT ensemble, we use Random Forest [23] as classifier. The number of trees is 150, which is tuned by 10-fold cross-validation with our dataset.

3) Chain-based multi-output classification: The (k, QP) prediction task can be considered as a problem of Multi-output classification (MOC), a supervised learning problem where an instance is associated with a set of discrete labels, rather than with a single label. A classical way is to predict separately each label with a different classifier by assuming that these labels are independent. In our case, however, k and QP are strongly correlated. In order to improve the MOC performance, we model the label dependencies by a modified Classifier Chain (CC) at the expense of an increased computational cost.

As in [52], a CC model involves several classifiers, which are linked along a chain where each classifier deals with a classification problem associated with a different label. The predictions of the different classifiers are cascaded as additional features. In other words, the feature space of each link is extended with the label associations of all previous links. We have observed that a competitive chain scheme is experimentally better than simple unidirectional chains. In such a scheme, the values of k and QP are at first separately predicted by two independent Random Forests, each taking the feature space defined in Section VI-B1. We then choose the prediction (k or QP) for which the classifier gets a higher probability and add it into the new feature space. A third Random Forest is then employed to predict the other label with the augmented feature space.

VII. Performance analysis

For simulations, we consider real light fields captured by plenoptic cameras using an array of micro-lenses, coming from different sources: 1/- the INRIA dataset [24] which contains LFs captured either by a first generation Lytro camera (63 LFs, 11×11 views of 379×379 pixels per LF) or a second generation Lytro Illum camera (46 LFs, 15×15 views of 625×434 pixels per LF); 2/- the ICME 2016 Grand Challenge dataset [25] containing 12 Lytro Illum LFs.

Lytro LFs are decoded by the Matlab Light Field Toolbox v0.4 [53]. With the INRIA dataset, we only consider the 9×9 central views in order to alleviate the strong vignetting and distortion problems on the peripheral views, which comparatively impact more the performance of the HEVC-based reference schemes, e.g. [12], [13]. Note, however, that there are still variations of light intensity, to a lesser extent, in the truncated light fields. With the ICME 2016 Grand Challenge dataset, we take 13×13 central views as defined by the challenge testing conditions. The test light fields in this section are shown in Fig. 6 and Fig. 7.

For the INRIA dataset, the method in [49] has been used to estimate a depth map of the central view in order to test our method with multiple homographies per view as described in Section V. For the ICME 2016 Grand Challenge dataset, the provided depth maps have been used.

In experiments, the bitrate and PSNR are given for the luminance component. The PSNR is derived from the MSE (mean square error) computed on the whole light field.

A. Analysis of HLRA-based compression

1) Joint homography and low rank optimization: We first assess the benefit of the joint optimization of the homographies

Fig. 6. Test light fields in INRIA dataset. First line: Real Light Fields captured by a Lytro 1G camera (From left to right: TotoroWaterfall, Beers, Flower and TapeMeasure); Second line: Real Light Fields captured by a Lytro Illum camera (From left to right: Fruits, Bench, BouquetFlower1 and Toys).

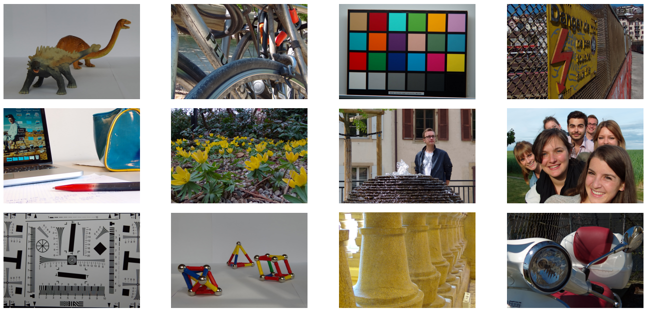


Fig. 7. Thumbnails of the ICME 2016 Grand Challenge dataset. From left to right, First line: Ankylosaurus_&_Diplodocus_1, Bikes, Color_Chart_1, Danger_de_Mort; Second line: Desktop, Flowers, Fountain, Friends; Third line: ISO_Chart_12, Magnets, Stone_Pillars_Outside, Vespa.

TABLE I
PSNR OBTAINED WITH DIFFERENT VALUES OF RANK k FOR HOMOGRAPHY OPTIMIZATION AND LOW RANK APPROXIMATION.

aligning	approximation	PSNR (dB)	
rank	rank	TotoroWaterfall	Toys
5	5	34.44	34.67
30	30	45.98	41.94
30	5	33.04	33.12
5	30	40.23	38.33

TABLE II
EXACT MATCHING RATE OF DIFFERENT CLASSIFICATION SCHEMES.

	independent	CC	CC	competitive
	classifiers	k –> QP	QP –> k	CC
Lytro 1G	64.2%	65.7%	66.1%	67.1%
Lytro Illum	59.2%	62.0%	62.0%	64.1%

and of the B and C matrices. Table I shows the PSNR obtained with different values of rank k for homography search and low rank approximation. The same value of k in both columns means that the same rank is used for computing the homographies and the transmitted matrices B and C. By comparing the first and the third row on one hand, and the second and fourth row on the other hand, for both light fields, one can see that for a given approximation rank, a joint optimization of homographies and of the approximation brings a significant gain.

2) *Alignment gain:* Fig. 8 shows the interest of view alignment for compression for three LFs: “TotoroWaterfall”, “Toys” and “Ankylosaurus_&_Diplodocus_1”. The gain of applying homographies to light field views is mostly significant at low bitrates and with a low approximation rank. Note that for plenoptic cameras using arrays of micro-lenses, disparity across views is relatively limited, hence alignment by one homography per view is usually sufficient to satisfy low rank assumption. When the disparity significantly varies across the scene (*e.g.* “TotoroWaterfall”), multiple homographies can further improve the compression performance, despite the additional bitrate cost for transmitting the depth map.

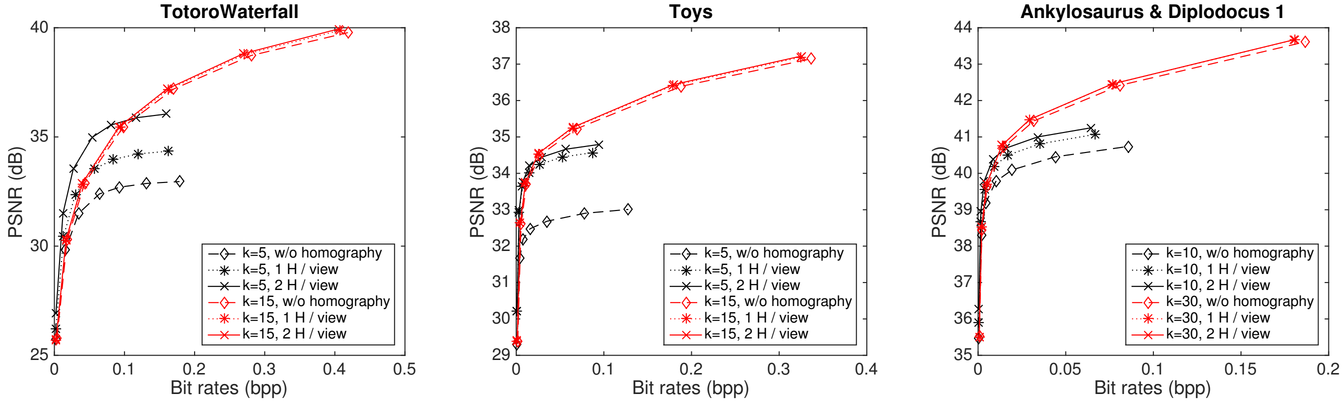
3) *Accuracy of the parameters prediction model:* Fig. 8 also suggests that the PSNR-rate performance depends on the values of the approximation rank k and of the HEVC coder quantization parameter (QP). For example, it appears that a smaller approximation rank k is preferred at low bit-rate. It can also be observed that the best pair of parameters (k, QP) depends on the input light field, hence the need to model the relationship between the target bit-rate, input light field features and the best values for these parameters.

The model proposed in subsection VI-B for predicting the best pair of (k, QP) has been trained using light fields in the INRIA dataset [24]. The dataset contains both indoor and outdoor captures and the light fields are taken with variable focal length. The test content is not similar to the training content. For each type of camera (Lytro first generation or Illum), we chose 4 light fields for the test (*cf.* Fig. 6). The remaining light fields are used for training and they correspond to different types of scenes. Training and test LFs are available for download on the website[1]. For each LF, compression is performed with the HLRA scheme for a combination of different values of k and QP ($k \in \{5, 15, 30\}$ and QP $\in$ $\{2, 6, 10, 14, 20, 26, 38\}$). Finally, training data labels consist of (k, QP) pairs only corresponding to the data points on the envelope of the PSNR-rate points.

Table II shows the exact matching rates of the competitive classifier chain (competitive CC) compared to other classification methods: 1/- independent classifiers: k and QP are independently predicted; 2/- CC k –> QP: classifier chain with k predicted before QP; 3/- CC QP –> k: classifier chain with QP predicted before k. Note that in competitive CC scheme, with two labels to classify, there are four classifiers to train and three of them are employed at the test time, whereas in a classical CC model, the number of classifiers is two at both the training and test phase. Fig. 9 shows in solid lines the true envelope after testing all possible (k, QP) pairs, and in dashed lines the PSNR-rate curves corresponding to the predicted (k, QP) pairs. Both one (black curves) and two homographies per view (red curves) are investigated. Although the exact match of (k, QP) values is not achieved for about $1/3$ of cases (Table II), we observe that the predicted PSNR-rate curve is very close to the true envelope.

In order to avoid multiple trainings for different numbers of homographies, the training has been done with a single homography. However, one can see in Fig. 9 that the prediction

[1] https://www.irisa.fr/temics/demos/lightField/LowRank2/datasets/datasets.html



(a) (b) (c)

Fig. 8. PSNR-rate performance of HLRA: with (1 or 2 homographies per view) or without alignment of light field views.

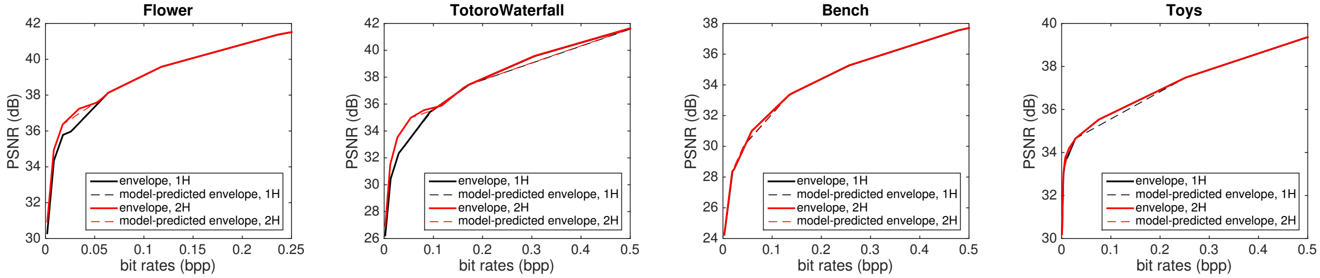


Fig. 9. Performance evaluation of model-based parameter prediction for 4 of the test LFs. The PSNR-rate curves corresponding to model-predicted coding parameters are shown by dashed lines, and the envelope obtained by all possible combinations of coding parameters are shown by solid lines. Both one (black curves) and two homographies per view (red curves) are investigated.

TABLE III

PERCENTAGE OF THE BITRATE ALLOCATED TO EACH ELEMENT. THE RESULTS ARE AVERAGED OVER TEST LIGHT FIELDS IN THE ICME 2016 GRAND CHALLENGE DATASET AND INRIA LYTRO DATASET, RESPECTIVELY.

Datasets	Nb. of H per view	Traget bitrates (bpp)	B	C	h	D
ICME Grand Challenge	1H	4.5×10^{-3} (k=10, QP=26)	81.9%	11.5%	6.6%	-
	1H	9.8×10^{-2} (k=60, QP=6)	95.7%	4.0%	0.3%	-
	2H	5.1×10^{-3} (k=10, QP=26)	67.6%	9.5%	11.1%	11.8%
	2H	9.9×10^{-2} (k=60, QP=6)	94.8%	4.0%	0.5%	0.7%
INRIA Lytro Dataset	1H	1.0×10^{-2} (k=5, QP=26)	92.4%	3.3%	4.3%	-
	1H	3.0×10^{-1} (k=60, QP=6)	99.1%	0.8%	0.1%	-
	2H	1.2×10^{-2} (k=5, QP=26)	77.5%	2.7%	7.5 %	12.3%
	2H	3.0×10^{-1} (k=30, QP=6)	98.4%	0.8%	0.3%	0.5%

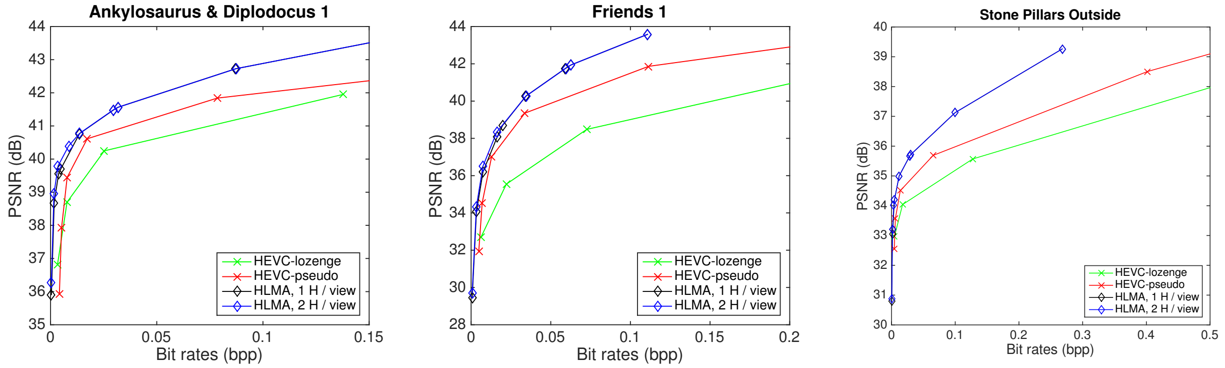
model learned on light fields for one homography remains valid for multiple homographies.

4) Bitrate cost percentage analysis: The percentage of the bitrate allocated to each element is detailed in Table III. A dominant part of bits are allocated to encoding the matrix B. The homography parameters in h and the depth map D require a fixed cost that does not vary in function of the target bitrate. As a consequence, the percentage of their cost becomes negligible at high bitrates.

B. Comparative assessment of compression performance

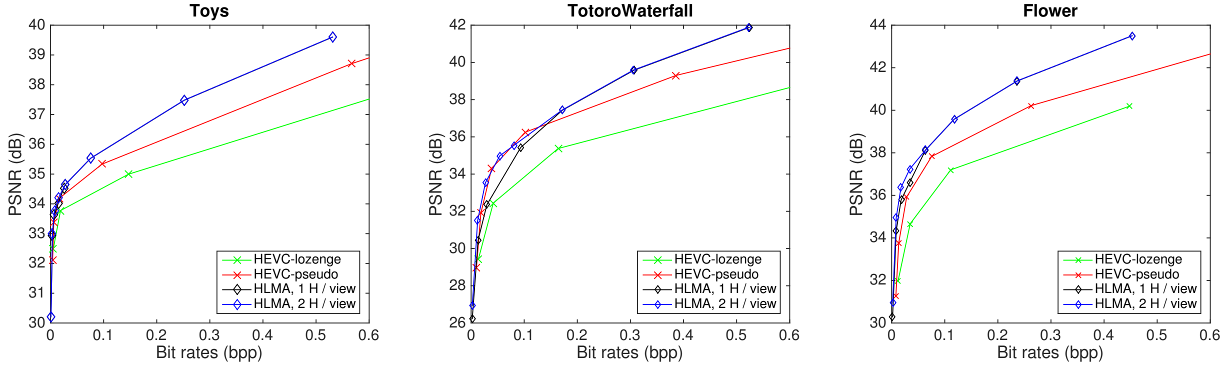
We assess the compression performance obtained with the homography-based low rank approximation against two schemes: direct encoding of the views as a pseudo-video sequence according to a lozenge order (HEVC-lozenge) [12] and according to the scanning order proposed in [13] (HEVC-pseudo). In simulations, the base QPs of HEVC-pseudo are set to QP_B =8, 14, 20, 26, 32 and 38, and the views at hierarchical layers 2, 3, 4, 5, 6 respectively have QPs equal to $\text{QP}_B + 8, \text{QP}_B + 9, \text{QP}_B + 10, \text{QP}_B + 11$ and $\text{QP}_B + 12$, as described in [13]. For HEVC-lozenge, the base QPs are set to 20, 26, 32, 38 and a GOP of 4 is used. The HEVC version used in the tests is HM-16.10.

In Figs. 10 and 11, both HLRA with 1H (one homography per view) and HLRA with 2H (two homographies per view) are investigated against HEVC-lozenge and HEVC-pseudo.



(a) (b) (c)

Fig. 10. PSNR-rate performance comparisons with three images from the ICME 2016 Grand Challenge dataset.



(a) (b) (c)

Fig. 11. PSNR-rate performance comparisons with three images from the INRIA dataset.

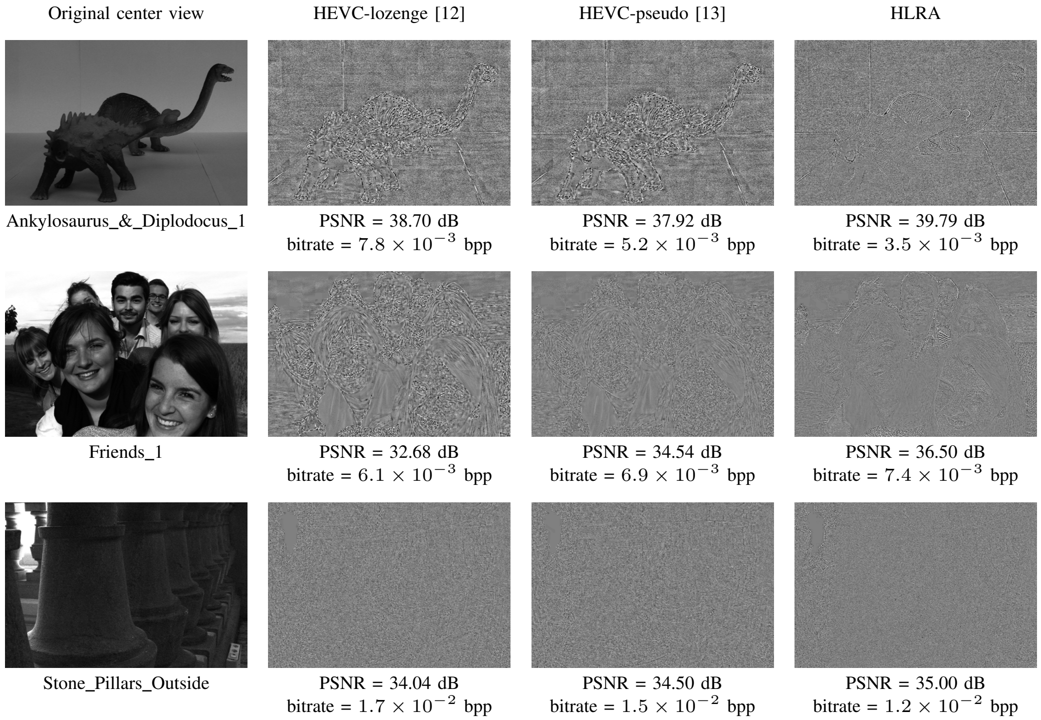

Original center view
HEVC-lozenge [12]
HEVC-pseudo [13]
HLRA
Ankylosaurus_&_Diplodocus_1
PSNR = 38.70 dB
bitrate = 7.8 × 10^{-3} bpp
PSNR = 37.92 dB
bitrate = 5.2 × 10^{-3} bpp
PSNR = 39.79 dB
bitrate = 3.5 × 10^{-3} bpp
Friends_1
PSNR = 32.68 dB
bitrate = 6.1 × 10^{-3} bpp
PSNR = 34.54 dB
bitrate = 6.9 × 10^{-3} bpp
PSNR = 36.50 dB
bitrate = 7.4 × 10^{-3} bpp
Stone_Pillars_Outside
PSNR = 34.04 dB
bitrate = 1.7 × 10^{-2} bpp
PSNR = 34.50 dB
bitrate = 1.5 × 10^{-2} bpp
PSNR = 35.00 dB
bitrate = 1.2 × 10^{-2} bpp

Fig. 12. The approximation error of the center view obtained with HEVC-lozenge, HEVC-pseudo and HLRA compression schemes. Similar bitrates are chosen for the three schemes.

TABLE IV
BD-PSNR GAINS WITH RESPECT TO HEVC-LOZENGE SCHEME [12]. THE GAINS ARE SHOWN FOR THE HEVC-PSEUDO [13] AND FOR OUR HLRA SCHEME WITH ONE OR TWO HOMOGRAPHIES PER VIEW.

Datasets	LF Names	HEVC-pseudo [13]	HLRA with 1H	HLRA with 2H
ICME Grand Challenge $13 \times 13 \times 625 \times 434$	Bikes	2.53	2.87	**2.91**
	Danger_de_Mort	2.74	**3.60**	3.46
	Flowers	3.16	3.22	**3.33**
	Stone_Pillars_Outside	0.70	1.47	**1.48**
	Vespa	1.79	1.79	**2.00**
	Ankylosaurus_&_Diplodocus_1	0.58	1.30	**1.35**
	Desktop	2.00	1.95	**2.14**
	Magnets_1	0.56	1.48	**1.49**
	Fountain_&_Vincent_2	**2.08**	1.51	1.82
	Friends_1	2.40	3.37	**3.47**
	Color_Chart_1	0.84	**1.73**	1.67
	ISO_Chart_12	**1.97**	1.90	1.77
	Average	1.78	2.18	**2.24**
INRIA Lytro Illum $9 \times 9 \times 625 \times 434$	BouquetFlower1	1.78	2.12	**2.13**
	Toys	0.70	1.55	**1.56**
	Bench	2.47	2.76	**2.77**
	Fruits	**2.72**	1.92	2.13
	Average	1.92	2.09	**2.15**
INRIA Lytro 1G $9 \times 9 \times 379 \times 379$	Beers	1.55	2.24	**2.26**
	TotoroWaterfall	1.85	1.50	**2.08**
	Flower	1.47	2.36	**2.52**
	TapeMeasure	1.26	1.63	**1.69**
	Average	1.53	1.93	**2.14**
Overall average		1.76	2.11	**2.20**

Table IV, gives the PSNR gain (using the Bjontegaard measure) of the proposed method with one or two homographies in comparison with the method HEVC-pseudo [13]. The reference considered for computing the Bjontegaard measure is HEVC-lozenge [12]. Furthermore, the approximation error of the center view obtained with HEVC-lozenge, HEVC-pseudo and HLRA is given in Fig. 12 for comparison.

Substantial gains in favor of HLRA schemes are observed for most of the test light fields. In Table IV, in most cases, best results are obtained with HLRA using two homographies per view, though on average, one homography per view is sufficient to outperform the HEVC-pseudo scheme. HLRA with two homographies are especially interesting when the scene contains several depth planes, such as “Desktop”, “TotoroWaterfall” and “Flower”. Note that there is strictly only one depth plane in “Color_Chart_1” and “ISO_Chart_12”, and the inaccuracy in depth map estimation explains the degradation when using 2 homographies per view for these two LFs.

For real world LFs captured by plenoptic cameras, global variations of light intensity are present between views. Although this degrades the performance of HEVC inter coding, our compression scheme is little affected since the rank of a matrix remains constant when its columns are multiplied by different factors.

Note that the PSNR of the central views (the 7×7 views in the center of the light field) is quite stable. However, the PSNR of the reconstructed views at the periphery of the light field varies. The gap can be up to 2 - 3 dB for real LFs between the central views and the views at the periphery. This variation is due to 2 reasons: 1/- a peripheral view requires a more significant homography transformation in order to align with other views, and the error due to forward and inverse warping is consequently more important. 2/- peripheral views suffer more severely from noise and distortion than central views. These artifacts are removed by low rank approximation, which causes the drop of PSNR, but not necessarily the degradation of visual quality. One may refer to the website[2] to see the visual quality of reconstructed views. Furthermore, this variation of PSNR between views is compensated in the two layer scheme presented in Section VIII by transmitting the residue.

C. Computational complexity analysis

The computational complexity of HLRA scheme mainly resides in two parts: finding the set of homographies to align the light field views, and encoding the matrix B. Searching for homographies is an iterative procedure, each iteration containing successive steps: 1/- warping involves multiplying each pixel coordinate in each light field view $[x, y, 1]$ by a 3×3 matrix, the complexity being $\mathcal{O}(nm)$; 2/- computing the Jacobian J_i for $i \in [|1, n|]$ implies $\mathcal{O}(nm)$ operations; 3/- the complexity for the SVD of the matrix $I' = I \circ h$ is $\mathcal{O}(n^2 m)$; 4/- and finally for $i \in [|1, n|]$, computing Δh_i involves computing the pseudo-inverse of the Jacobian matrix J_i of size $m \times 8$, which has a complexity of $\mathcal{O}(m)$, followed by a matrix multiplication with complexity $\mathcal{O}(m)$. Overall, each iteration requires $\mathcal{O}(n^2 m)$ arithmetic operations.

[2] https://www.irisa.fr/temics/demos/lightField/LowRank2/LRcompression.html

TABLE V
RUNTIMES. THE RESULTS ARE AVERAGED OVER THE TEST LIGHT FIELDS IN DIFFERENT DATASETS. THE NUMBER OF REQUIRED ITERATIONS (NB. ITERS) AND RUNTIME (T) ARE DETAILED. THE RUNTIME IS MEASURED AT $\text{QP}_B = 20$ BOTH FOR HLRA AND HEVC-PSEUDO [13].

Datasets		HLRA			HEVC-pseudo [13]
		Searching H		Encoding B	
		Nb. iters	t (min)	t (min)	t (min)
ICME Grand	$k = 10$	29.9	12.5	0.6	17.0
Challenge	$k = 30$	7.3	3.2	1.7	
$13 \times 13 \times 625 \times 434$	$k = 60$	2.7	2.7	3.3	
INRIA	$k = 5$	40.0	8.9	0.4	8.3
Lytro Illum	$k = 15$	7.8	1.8	0.9	
$9 \times 9 \times 625 \times 434$	$k = 30$	6.3	1.5	1.7	
INRIA	$k = 5$	20.8	2.2	0.2	5.0
Lytro 1G	$k = 15$	5.0	0.6	0.5	
$9 \times 9 \times 379 \times 379$	$k = 30$	4.8	0.5	0.9	

In our experiments, we consider that the algorithm has converged when the PSNR gain is less than 0.002 dB between two successive iterations. Logically, an approximation with smaller rank requires more iterations to converge. In fact, in the extreme case where the approximation rank is equal to the number of light field views, no alignment (hence zero iteration) is needed and the decomposition in B and C is reduced to a simple SVD. The encoding time of the matrix B is proportional to the number of columns (the approximation rank), since each column is intra coded by HEVC. Considering their small size, the encoding time of C, h and the depth map is negligible.

The number of iterations and the consumed time for finding homographies and encoding B are detailed in Table V. Simulations have been carried out with a Macbook Pro with a 2.8GHz Intel Core i7 processor. In spite of the additional cost for the iterative alignment, HLRA consumes less time in total than the HEVC-pseudo scheme because the intra coding of the columns of B is much faster than inter coding of all the views in HEVC-pseudo. Note that our implementation of the iterative alignment is written in Matlab and could be further optimized in the future.

On the decoder side, only k images need to be decoded with HEVC intra, and the remaining steps are the matrix multiplication BC (complexity $\mathcal{O}(mnk)$) and inverse warpings (complexity $\mathcal{O}(mn)$). Note that the decoding process is not iterative since the homographies and the matrices B and C are directly transmitted to the decoder.

D. Limitations of the method

For synthetic light fields in HCI dataset [54], the HLRA scheme performs in general better than HEVC-lozenge but worse than HEVC-pseudo in terms of PSNR-rate performance (*cf.* Table VI). Two main reasons may explain this degradation. First, while the baseline of the light fields captured with plenoptic cameras is limited by the aperture size of the camera, there is no such limitation with synthetic light fields which may then have much higher disparities between views. In these conditions, global homography projections less accurately compensate for the inter-view disparities than the block-matching of HEVC inter. With large baselines, the first columns of the matrix B contain considerably more information, even after view alignment, and are therefore more expensive to encode. This is the reason why, in Table VI, the gap between our method and HEVC-pseudo is more important for light fields with larger baselines (“StillLife” and “Buddha”) than for those with smaller baselines (“Butterfly” and “MonasRoom”). Secondly, synthetic light fields are free of imperfections such as noise and variations of light intensity between views. The real world LF imperfection degrades considerably the performance of HEVC inter coding while HLRA compression scheme is little affected. In fact, a global change of light intensity in one view simply leads to multiplying the corresponding coefficients in the matrix C by different factors, while the noise is mostly dropped by the low rank model.

TABLE VI
BD-PSNR GAINS (WITH RESPECT TO HEVC-LOZENGE SCHEME [12]) EVALUATION FOR SYNTHETIC LIGHT FIELDS IN HCI DATASET. THE GAINS ARE SHOWN FOR THE HEVC-PSEUDO [13] AND FOR OUR HLRA SCHEMES.

LF Names	HEVC-pseudo [13]	HLRA
Buddha	4.72	1.00
Butterfly	2.52	0.17
MonasRoom	3.90	1.08
StillLife	4.37	-4.28

VIII. SCALABLE LIGHT FIELD CODING

Thanks to the matrix factorization used in our approach, our coding scheme naturally presents an interesting scalability property. Although the encoder transmits a fixed number of columns of the matrix B corresponding to the rank k, the decoder can choose to decode less than k columns. Since the matrices B and C are obtained by a singular value decomposition which sorts the singular values and their corresponding singular vectors, the first columns of B contain most of the energy of the light field signal. Therefore, a fast approximation of the encoded light field can already be performed on the decoder side by decoding only those first columns. The decoder may then progressively refine the low rank approximation by decoding additional columns.

However, in addition to the errors caused by the low rank approximation, we have identified two other types of errors either caused by the quantization in the HEVC compression, or by the forward and inverse homography warping. These errors are not corrected by encoding additional columns of B. Therefore, we propose a scalable extension of our method where the residual between the original and the decoded light field is encoded as an enhancement layer, as illustrated in Fig. 13.

For our experiment, the base layer is computed by encoding the original light field with HLRA using a single homography. For the residual layer, four coding schemes are tested:

- HEVC-lozenge encoding [12].
- HEVC-pseudo encoding [13].
- HLRA without alignment (only BC factorization).
- HLRA with one homography.

Fig. 14 shows the PSNR-rate performance of the scalable light field coding. For comparing the different residual layer coding schemes, the base layer is encoded at a fixed target bitrate of 0.01 bpp. The corresponding k and QP parameters are automatically predicted by our classification model decribed in section VI-B. For the HLRA encoding of the residual (either with or without homography alignment), results are shown for different values of the approximation rank k_r. Each curve is generated by varying the QP parameter in the HEVC encoding of the residual B matrix.

We first note that all the variants of our HLRA scheme perform significantly better than the HEVC-lozenge and HEVC-pseudo schemes applied to the residual layer.

Unlike the base layer (*cf.* Fig. 8), aligning the residual and encoding with a low rank k_r does not significantly improve the results at low or medium bitrates compared to a higher rank encoding. The best performance at any bitrate is then obtained by choosing the highest rank k_r. In this case, homography alignment only brings very little gains which may not justify the added complexity of the optimization procedure for determining homographies for the residue. The BC factorization computed by a single SVD step is then sufficient for the residual matrix encoding.

Finally, in comparison to the single layer approach (red curve), encoding a residual layer with a high approximation rank ($k_r = 30$) only results in a negligible loss compared to single layer HEVC-based coding while providing scalability. Note that similarly to the base layer, the residual layer encoded with the BC factorization can be decoded by progressive refinement when decoding the successive columns of the B matrix.

IX. Conclusion

In this paper, we have proposed a new compression scheme for light field images. In our method, each view of the light field is first warped using either one global homography, or a model consisting of a homography per depth plane with a smooth transition between depth planes. Considering the matrix formed by concatenating each warped and vectorized view, a joint optimization of homography parameters and low rank matrix approximation is performed. For an approximation rank k, the resulting rank k matrix can be factorized as the product of a matrix containing k basis vectors and a smaller

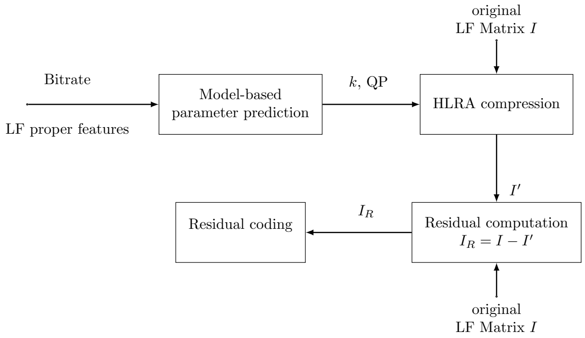


Fig. 13. Scalable light field coding chain.

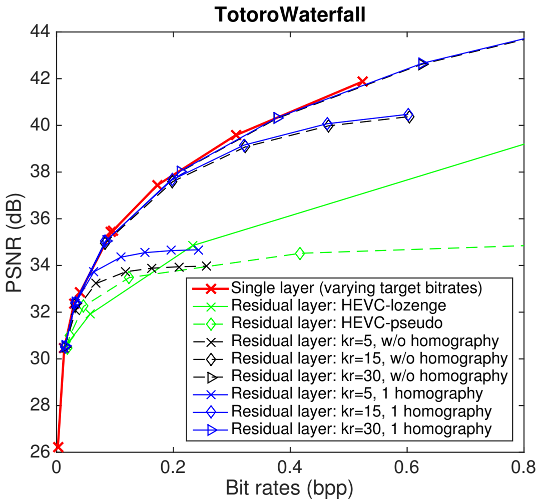


Fig. 14. Scalable coding performance for a base layer encoded with the single homography HLRA. The red curve shows the single layer result at varying target bitrates. The other curves show the scalable performance for different residual encoding methods, where the base layer is encoded at a target bitrate of 0.01 bpp ($k = 5$, QP$= 26$).

matrix of coefficients. Our method then consists in encoding each of the k basis vectors using HEVC intra, along with the coefficients and homography parameters. In the case of multiple depth planes, the depth map must be encoded as well using HEVC intra. For most tested light fields, experimental results show substantial performance gains compared to the state-of-the-art methods which encode the views as a pseudo-sequence with HEVC inter. The method is particularly well suited for real light fields captured with plenoptic cameras which have limited disparity, but may contain imperfections such as variations of light intensity across views.

Our method is dependent on two encoding parameters: the rank k and the QP parameter in HEVC. Therefore, we have also proposed a prediction scheme for determining the best couple (k, QP) as a function of a target bitrate and additional features computed on the input light field. Our experiments show that the predicted parameters always result in close to optimal coding performance. Furthermore, the model remains valid in broader conditions than what it was trained for (*i.e.* multiple homographies).

Finally, a scalable extension has been proposed where a residual layer is encoded. We have shown that homography warping is not necessary for the residual. A simpler coding scheme only based on matrix factorization is sufficient and it substantially outperforms the HEVC inter encoding of the pseudo-sequence of view residuals.

REFERENCES

[1] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[2] R. Ng, "Light field photography," Ph.D. dissertation, Stanford University, 2006.

[3] T. Georgiev, G. Chunev, and A. Lumsdaine, "Super-resolution with the focused plenoptic camera," *Proc. SPIE*, 2011.

[4] M. Levoy and P. Hanrahan, "Light field rendering," in *23rd Annual Conf. on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. ACM, 1996, pp. 31–42.

[5] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. of Selected Topics in Signal Processing. Special Issue on Light Field Image Processing*, Oct. 2017.

[6] P. Lalonde and A. Fournier, "Interactive rendering of wavelet projected light fields," in *Int. Conf. on Graphics Interface*, 1999, pp. 107–114.

[7] I. Peter and W. Straßer, "The wavelet stream - progressive transmission of compressed light field data," in *IEEE Visualization 1999 Late Breaking Hot Topics*. IEEE Computer Society, 1999, pp. 69–72.

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

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

[10] M. Magnor, A. Endmann, and B. Girod, "Progressive compression and rendering of light fields," in *Vision, Modelling and Visualization*, 2000, pp. 199– 203.

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

[12] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on focus stack and extended focus images," in *24th European Signal Processing Conf. (EUSIPCO)*, Aug. 2016, pp. 898–902.

[13] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, Jul. 2016.

[14] X. Jiang, M. L. Pendu, R. A. Farrugia, S. S. Hemami, and C. Guillemot, "Homography-based low rank approximation of light fields for compression," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 1313–1317.

[15] C. Conti, P. Nunes, and L. D. Soares, "New HEVC prediction modes for 3d holoscopic video coding," in *IEEE Int. Conf. on Image Processing (ICIP)*, Sept. 2012, pp. 1325–1328.

[16] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient intra prediction scheme for light field image compression," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 539–543.

[17] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, Jul. 2016.

[18] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, Jul. 2016.

[19] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, Jul. 2016.

[20] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW)*, Jul. 2016.

[21] T. Sakamoto, K. Kodama, and T. Hamamoto, "A study on efficient compression of multi-focus images for dense light-field reconstruction," in *Visual Communications and Image Processing*, Nov. 2012, pp. 1–6.

[22] A. Levin and F. Durand, "Linear view synthesis using dimensionality gap light field prior," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2010, pp. 1831–1838.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] "INRIA Lytro image dataset," https://www.irisa.fr/temics/demos/lightField/LowRank2/datasets/datasets.html.

[25] "ICME 2016 Grand Challenge dataset," http://mmspg.epfl.ch/EPFL-light-field-image-dataset.

[26] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *23rd Annual Conf. on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. ACM, 1996, pp. 43–54.

[27] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[28] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Advances in Neural Information Processing Systems (NIPS)*, 2010, pp. 937–945.

[29] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Trans. on Information Theory*, vol. 56, no. 9, pp. 4402–4416, Sep. 2010.

[30] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. on Information Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.

[31] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *28th Int. Conf. on Machine Learning (ICML)*, 2011, pp. 33–40.

[32] J. Tanner and K. Wei, "Normalized iterative hard thresholding for matrix completion." *SIAM J. Scientific Computing*, vol. 35, no. 5, 2013.

[33] Y. D. Kim and S. Choi, "Weighted nonnegative matrix factorization," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2009, pp. 1541–1544.

[34] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.

[35] Z. Lin, M. Chen, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," *arXiv:1009.5055*, Sep. 2010.

[36] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, Jun. 2012.

[37] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[38] J. Wright, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[39] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2010, pp. 763–770.

[40] C. Chen, J. Cai, W. Lin, and G. Shi, "Incremental low-rank and sparse decomposition for compressing videos captured by fixed cameras," *J. of Visual Communication and Image Representation*, vol. 26, pp. 338–348, Jan. 2015.

[41] J. Hou, L. P. Chau, N. Magnenat-Thalmann, and Y. He, "Sparse low-rank matrix approximation for data compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 27, no. 5, pp. 1043–1054, May 2017.

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

[43] L. Lucas, C. Conti, P. Nunes, L. Soares, N. Rodrigues, C. Pagliari, E. Silva, and S. Faria, "Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC," in *22nd European Signal Processing Conf. (EUSIPCO)*, Sept. 2014, pp. 11–15.

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

[45] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Trans. on Image Processing*, vol. 25, no. 1, pp. 80–91, Jan. 2016.

[46] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Conf. on*

Computer Vision and Pattern Recognition (CVPR), Jun. 2013, pp. 1027–1034.

[47] S. Kundu, "Light field compression using homography and 2d warping," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 1349–1352.

[48] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[49] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2013, pp. 673–680.

[50] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1547–1555.

[51] T. C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *IEEE Int. Conf. on Computer Vision (ICCV)*, Dec. 2015, pp. 3487–3495.

[52] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333–359, 2011.

[53] D. Dansereau, "Light Field Toolbox for Matlab," 2015.

[54] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *VMV Workshop*, 2013, pp. 225–226.