

HOMOGRAPHY-BASED LOW RANK APPROXIMATION OF LIGHT FIELDS FOR COMPRESSION

Xiaoran Jiang¹, Mikael Le Pendu¹, Reuben A. Farrugia², Sheila S. Hemami³, Christine Guillemot¹

¹ INRIA, Campus de Beaulieu, Rennes, France

² University of Malta, Malta

³ Northeastern University, Boston, USA

ABSTRACT

This paper studies the problem of low rank approximation of light fields for compression. A homography-based approximation method is proposed which jointly searches for homographies to align the different views of the light field together with the low rank approximation matrices. We first consider a global homography per view and show that depending on the variance of the disparity across views, the global homography is not sufficient to well-align the entire images. In a second step, we thus consider multiple homographies, one per region, the region being extracted using depth information. We first show the benefit of the joint optimization of the homographies together with the low-rank approximation. The resulting compact representation is then compressed using HEVC and the results are compared with those obtained by directly applying HEVC on the light field views re-structured as a video sequence. The experiments using different data sets show substantial PSNR-rate gain of our method, especially for real light fields.

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

1. INTRODUCTION

Light fields are densely sampled high-dimensional signals containing information about the light rays interacting with the scene. They yield a very rich description of a 3D scene which enables advanced creation of novel images, with e.g., refocusing, extended depth of field, different viewpoint rendering, from a single capture [1–3]. Many acquisition and sampling techniques have been recently designed to capture light fields, going from arrays of cameras capturing the scene from slightly different viewpoints [3], to single cameras mounted on moving gantries and plenoptic cameras. Plenoptic cameras use arrays of microlenses placed in front of the photosensor to obtain angular information about the captured scene [2, 4].

Light fields consist of very large volumes of high dimensional data, which appear to be a major downside for practical

use. Compression schemes have been proposed for light fields based on vector quantization [5], transform coding [6, 7], statistical representations [8], or using multiview video compression and disparity compensation techniques [9, 10]. Intra coding modes have also been proposed in [11] for light fields compression using HEVC, and scalable coding has been recently considered for light fields [12].

This paper investigates low rank approximation methods exploiting data geometry for dimensionality reduction of light fields. Low rank approximations have already been explored in [13] and [14] for video compression. In this paper, an approximation method is proposed in which homographies and the rank approximation model are jointly optimized. A global homography per view is first considered to align each view on the central one. The homographies are searched in order to align linearly correlated sub-aperture images in such a way that the batch of views can be approximated by a low rank model. Note that a low rank approximation model has been considered in the RASL method [15] for aligning correlated views. While RASL approximates the input warped matrix by the sum of a minimum rank matrix plus a sparse error matrix, the rank constraint is here expressed in a factored form where one matrix B contains basis vectors and where the other one C contains weighting coefficients. The basis vectors and weighting coefficients can be compressed separately exploiting their respective characteristics. The optimization hence proceeds by iteratively searching for the homographies and the factored model of the input set of sub-aperture images (views), which will minimize the approximation error. The approximation is further improved by updating the coefficient matrix C given the encoded matrix B so that the transmitted matrix C accounts for the quantization errors in B .

We first show the advantage of using the joint alignment and low-rank optimization rather than first aligning independently of the targeted rank. This homography-based low rank model allows us to exploit the coherence between the sub-aperture images. The component matrices of the model can then be compressed using a standard encoder (HEVC has been used in the experiments). The compression performances are assessed against those obtained by applying

The work has been partly funded by a Google faculty research award.

HEVC-based inter-coding on the sequences of images formed by extracting the sub-apertures images following a lozenge scan order starting at the central view. The experiments using different data sets show that a substantial PSNR-rate gain compared with HEVC encoding for light fields can be obtained with only one homography per view, when the disparity between views is limited or does not significantly vary across the scene (scenes with limited depth of field). Multiple homographies (one homography per depth plane) have then been considered and shown to improve the PSNR-rate performance, when the disparity significantly varies from one depth plane to another. To cope with artifacts at the frontier of two depth planes, a blending of homographies is performed.

2. LIGHT FIELDS: NOTATIONS

We consider the simplified 4D representation of light fields called 4D light field in [1] and lumigraph in [16], 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. The light field can be seen as capturing an array of viewpoints (called sub-aperture images in particular in the case of micro-lenses based capturing devices) of the imaged scene with varying angular coordinates u and v . The different views will be denoted here $I_{u,v} \in \mathbb{R}^{X,Y}$, 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) .

In the following, the notation $I_{u,v}$ for the different views (or sub-aperture images) will be simplified as I_i with a bijection between (u, v) and i . The complete light field can hence be represented by the matrix $I \in \mathbb{R}^{m,n}$:

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

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.

3. 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

$$I_j(x, y) = (I_i \circ h_i)(x, y) = I_i(h_i(x, y)). \quad (2)$$

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

$$[x_H, y_H, w_H]^\top = H_i \cdot [x, y, 1]^\top. \quad (3)$$

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 $[h_1 \mid \dots \mid h_n]$ where h_1, \dots, h_n are vectors of size 8×1 whose elements are the homography parameters. The low rank optimization problem is then formulated as

$$\arg\min_{h, B, C} \|I \circ h - BC\|_F^2, \quad (4)$$

where $\|\cdot\|_F$ 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, \dots, h_n and can be written as

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

3.1. 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:

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

where $\Delta h = [\Delta h_1 \mid \dots \mid \Delta h_n]$, 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.

The minimization problem in Eq. (4) is iteratively solved by updating alternatively the matrices B and C and the homographies h_1, \dots, 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, and using the approximation in Eq. (6). Therefore, the updated homographies are given by $h' = h + \Delta h$, where

$$\Delta h = \arg\min_{\Delta h} \|I \circ h - BC + \sum_{i=1}^n J_i \Delta h_i \epsilon_i^\top\|_F^2 \quad (7)$$

This is a linear least squares problem with solution:

$$\forall i, \Delta h_i = J_i^\dagger (BC - I \circ h) \epsilon_i \quad (8)$$

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

3.2. Recalculate 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



Fig. 1: Light Fields used in the tests: Synthetic (three left images called Buddha, Butterfly, StillLife) and Real Light Fields (TotoroWaterfall and Beers captured by a Lytro camera) and Watch captured by a Raytrix camera.

compression (i.e. quantization) errors on the light field reconstruction, the matrix C is recalculated to account for these quantization errors, as follows:

$$C' = \underset{C}{\operatorname{argmin}} \|I \circ h - B'C\|_F^2 = (B')^\dagger (I \circ h). \quad (9)$$

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).

4. LOW RANK APPROXIMATION USING MULTIPLE HOMOGRAPHIES

When the disparity varies from one depth plane to another, the performances can be improved by using multiple homographies. Depth map D can be normalized between 0 and 1. A number q of depth planes is then obtained by uniformly quantizing D with quantization thresholds $\{\sigma_p\}_{p \in \llbracket 1, q \rrbracket}$, each depth plane p being characterized by a mask $M^p = \mathbb{1}_{[\sigma_p, \sigma_{p+1}]}(D)$, with $\mathbb{1}$ the pixel-wise indicator function.

We apply a different homography h_i^p to depth plane p of the sub-aperture image I_i . Blending 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. To do this, for each coordinates (x, y) , we define a series of weights $\{w_{(x,y)}^p\}$ that take into account the importance of each homography h_i^p to this pixel:

$$w_{(x,y)}^p = \begin{cases} 1, & \text{if } D_{(x,y)} \in [\sigma_p + \delta, \sigma_{p+1} - \delta]; \\ \frac{D_{(x,y)} - (\sigma_p - \delta)}{2\delta}, & \text{if } |D_{(x,y)} - \sigma_p| < \delta; \\ \frac{(\sigma_{p+1} + \delta) - D_{(x,y)}}{2\delta}, & \text{if } |D_{(x,y)} - \sigma_{p+1}| < \delta; \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

with δ a shallow neighborhood where the blending is applied. Eq. (3) is then modified as following:

$$[x_H, y_H, w_H]^\top = \sum_{p=1}^q w_{(x,y)}^p H_i^p \cdot [x, y, 1]^\top. \quad (11)$$

Once the warped images $I \circ h$ are obtained by applying Eq. (11) for each sub-aperture image and each depth plane, we compute B and C at each iteration exactly as described in Section 3. Note that h contains now $n \times q$ homographies. Similar to Eq. (8), each homography h_i^p is updated as:

$$\forall i, p, \Delta h_i^p = J_i^{p\dagger} [(BC - I \circ h)\epsilon_i \odot \operatorname{vec}(M^p)], \quad (12)$$

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

5. EXPERIMENTAL RESULTS

We consider both synthetic light fields (9×9 views of 768×768 pixels) from the dataset in [17] and real light fields captured by a first generation Lytro camera [18] and a Raytrix camera. Views of the test light fields are shown in Fig. 1. The real light fields are 11×11 views of 379×379 pixels (except for “watch”, each view of which has 992×628 pixels).

5.1. Joint homography-low rank optimization

We first assess the benefit of the joint optimization of the homographies and the transmitted B and C matrices. Table 1 shows the PSNR obtained with different values of rank k for homography search and low rank optimization. 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.

5.2. PSNR-rate performance

Fig. 2 shows the compression performance obtained with the homography-based low rank approximation compared with a direct encoding of the views as a video sequence (following a lozenge scanning order starting at the central view) using HEVC Inter coding (version HM-16.10). For our homography-based scheme, the low rank representation is compressed by encoding the columns of the matrix B using HEVC Intra coding. Intra mode is chosen since little correlation resides in the columns of B . The coefficients of the

Table 1: PSNR obtained with different values of rank k for homography optimization and low rank approximation.

aligning rank	approximation rank	PSNR (dB)	
		Still-life	Watch
15	15	27.83	49.56
60	60	35.62	64.48
60	15	25.52	43.67
15	60	33.23	53.28

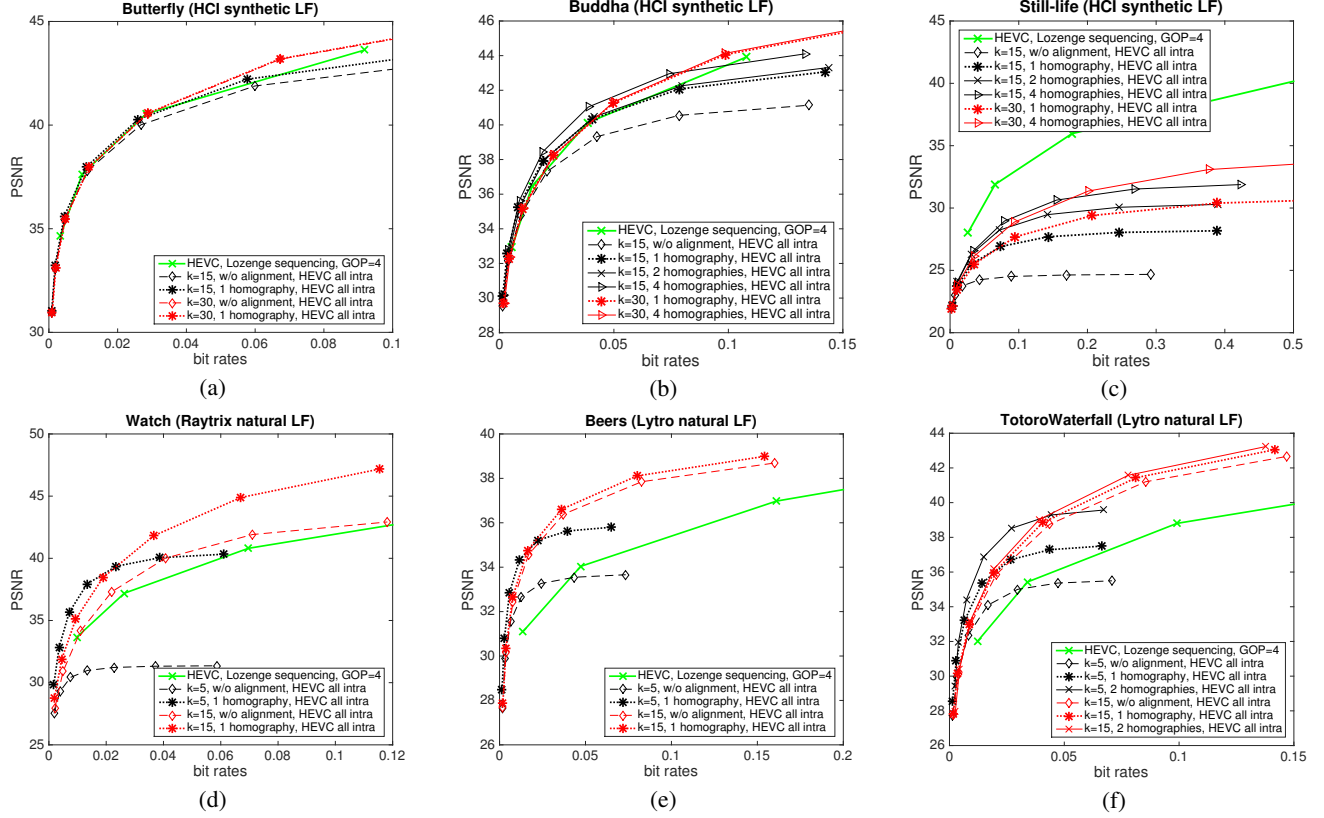


Fig. 2: PSNR-rate performance without and with view alignment for synthetic light fields (Fig. 2a - 2c) or for real light fields (Fig. 2d - 2f) compared with direct HEVC inter coding (green curves).

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 dimension of the matrix C is such that its coding rate is quite negligible, as well as the matrix h ($8 \times n$ coefficients). In Fig. 2, the curves obtained with the homography-based low rank approximation (solid or dotted lines) are compared with those obtained without homography (dashed lines). In most cases, significant gains are observed for the homography-based low rank approximation compared to direct HEVC inter coding of the sub-aperture view sequences (illustrated in green), especially for real light fields. When the disparity across views is high, the gain brought by homographies can be significant, provided that the low rank assumption holds after alignment (c.f. “buddha” vs. “butterfly”). Lower is the approximation rank, more noticeable is the gain. It also appears that a smaller approximation rank k is preferred at low bit-rate. However, when the disparity significantly varies across the scene, which is the case for “still-life” (Fig. 2c), the use of a unique homography per view is not sufficient, hence the consideration of multiple homographies.

5.3. Performance with multiple homographies per view

The gain brought by the use of multiple homographies per view has been measured both for synthetic and real data. Syn-

thetic light fields are provided with a ground truth depth map for each view. The depth map used to “segment” the depth planes is the average of the depth maps for all the views. For real light fields, the depth map can be estimated using methods as proposed in [19–21] for the central view. This depth map has been encoded with HEVC intra (QP=32) in the experiments. The encoding cost of the set of h of homographies is negligible ($8 \times n \times q$ coefficients, with q generally smaller than 5). Fig. 2b, 2c and 2f show that the use of multiple homographies significantly improves the PSNR-rate performance even if for “still-life”, the method still falls behind HEVC. More simulation results can be found on the web page <https://www.irisa.fr/temics/demos/lightField/LowRank/LRcompression.html>.

6. CONCLUSION

In this paper, we have proposed a compact representation of light fields based on a joint optimization of homographies and of a low rank approximation, considering both a single homography per view and per depth plane. A blending of homographies has also been proposed for a smooth merging of the border of the warped (and inversely warped) depth planes. Experimental results show very good PSNR-rate performances compared with HEVC encoding for most tested light fields, and especially for real ones.

7. REFERENCES

- [1] Marc Levoy and Pat Hanrahan, “Light field rendering,” in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 1996, SIGGRAPH ’96, pp. 31–42, ACM.
- [2] R. Ng, *Light Field Photography*, Ph.D. thesis, Stanford University, 2006.
- [3] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy, “High performance imaging using large camera arrays,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, July 2005.
- [4] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine, “Superresolution with the focused plenoptic camera,” 2011.
- [5] Andrew C. Beers, Maneesh Agrawala, and Navin Chaddha, “Rendering from compressed textures,” in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, 1996, SIGGRAPH ’96, pp. 373–378, ACM.
- [6] G. Miller, Rubin S., and D. Ponceleon, “Lazy decomposition of surface light fields for precomputed global illumination,” in *Rendering Techniques ’98*, 1998, pp. 281–292.
- [7] M. Magnor, A. Endmann, and B. Girod, “Progressive compression and rendering of light fields,” in *Vision, Modelling and Visualization*, 2000, pp. 199–203.
- [8] Dan Lelescu and Frank Bossen, “Representation and coding of light field data,” *Graph. Models*, vol. 66, no. 4, pp. 203–225, July 2004.
- [9] M. Magnor and B. Girod, “Data compression for light-field rendering,” *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 10, no. 3, pp. 338–343, Apr. 2000.
- [10] Chuo-Ling Chang, Xiaoqing Zhu, P. Ramanathan, and B. Girod, “Light field compression using disparity-compensated lifting and shape adaptation,” *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 793–806, April 2006.
- [11] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, “Efficient intra prediction scheme for light field image compression,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 539–543.
- [12] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, “Scalable coding of plenoptic images by using a sparse set and disparities,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 80–91, Jan 2016.
- [13] 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.
- [14] J. Hou, L.-P. Chau, N. Magnenat-Thalmann, and Y. He, “Slrma: Sparse low-rank matrix approximation for data compression,” *IEEE Trans. on Circuits and Systems for Video Technology*, Dec. 2015.
- [15] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma, “Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition, CVPR*, 2010.
- [16] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, “The lumigraph,” in *Proc. SIGGRAPH*, 1996, pp. 43–54.
- [17] Sven Wanner, Stephan Meister, and Bastian Goldluecke, “Datasets and benchmarks for densely sampled 4d light fields,” in *VMV Workshop*, 2013, pp. 225–226.
- [18] “Lytro image dataset,” <http://www.irisa.fr/temics/demos/lightField/index.html>, Accessed: 2016-08-20.
- [19] Michael W. Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” Dec. 2013.
- [20] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon, “Accurate depth map estimation from a lenslet light field camera,” in *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [21] Ting-Chun Wang, Alexei Efros, and Ravi Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.