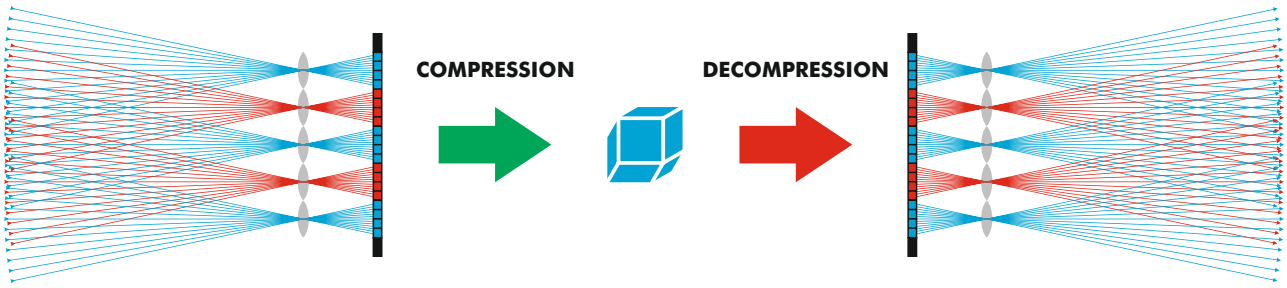
BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2019
http://excel.fit.vutbr.cz

4D-DCT Based Light Field Image Compression

Drahomír Dlabaja*



Abstract
This paper proposes a light field image encoding solution based on four-dimensional discrete cosine transform and quantization. The solution is an extension to JPEG baseline compression. A light field image is interpreted and encoded as a four-dimensional volume to exploit both intra and inter view correlation. Solutions to 4D quantization and block traversal are introduced in this paper. The experiments compare the performance of the proposed solution against the compression of individual image views with JPEG and HEVC intra in terms of PSNR. Obtained results show that the proposed solution outperforms the reference encoders for light images with a low average disparity between views, therefore is suitable for images taken by lenslet based light field camera and images synthetically generated.

Keywords: Light field — Lossy compression — JPEG — Transform coding — Plenoptic representation — Quality assessment

Supplementary Material: Downloadable Code

*xdlaba02@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

With the development of new technologies for scene capture, there are also growing demands for storage and transmission of these captures. An example of such technology is light field photography. So far, the light field has been an experimental field of computer graphics, but with the rise of more powerful hardware in combination with expanding availability of light field cameras such as Lytro and Raytrix[1], light field images are spreading to the general public.

Such images contain not only the angular information of light distribution but also spatial information, which creates a more complete representation of the scene. In practice, however, this means that memory requirements of light field image are several times larger in comparison to a classic image. Therefore, effective compression is a need.

In this paper, a new light field encoding method based on 4D-DCT is proposed. The method is extending JPEG baseline compression standard and introducing solutions to 4D quantization and block traversal. The method is exploiting both intra and inter view correlation by interpreting light field image as 4D volume. The proposed method is evaluated and compared with other solutions on different types of plenoptic images.

This paper could be crucial for future work on light field compression methods, as the knowledge that inter view correlation is exploitable by JPEG baseline extended to four dimensions can be transferred to other 2D transformation-based compression methods.

[1] Lytro, Raytrix – https://raytrix.de/

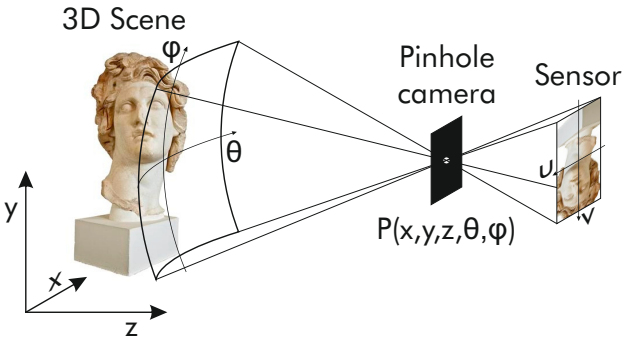


Figure 1. Amount of light passing through the pinhole at given position x, y, z and angle θ, ϕ correspond to value of the plenoptic function $P(x, y, z, \theta, \phi)$. This image is vectorized version of Figure 1 in [1].

2. Related work

Light field image, the same as any other type of scene capture, is just another slice of a plenoptic function [2]. The function measures a radiance of light traveling along a ray. It is defined as $P(x, y, z, \theta, \phi)$, where x, y, z are spatial coordinates and θ, ϕ are angular coordinates of a light ray. In some publications, t as a time coordinate and λ as a wavelength are another parameters in a plenoptic function, but for the sake of simplicity, these variables will not be taken into account in this paper. Plenoptic function is illustrated in Figure 1.

With an assumption that light always travels straight along a ray and the path is free of occlusions, the plenoptic function can be reduced to $L(x, y, \theta, \phi)$, a function called 4D light field. With light field, it is possible to capture the complete light representation of a scene in a convex hull. This can be pictured as a pinhole image of a scene described by angular coordinates θ, ϕ captured at every position of a plane x, y.

To capture a light field, a sampling of the light field function is needed. Light fields are often sampled quite densely in the angular domain. This is directly proportional to the number of microlenses in a light field camera or to the resolution of a camera in camera array. Compared to that, sampling in the spatial domain is quite sparse. It is indicated by a number of pixels capturing every microlens in a light field camera or number of cameras in a camera array.

A number of compression methods have been developed for light field images. Available image coding standard solutions are compared on 4D light field images without taking advantage of the specific light field data structure in [3]. Standard solutions are also compared on a raw data format from lenslet based plenoptic camera in [4]. Compression of demosaiced, devignetted and sliced raw lenslet images with JPEG 2000 is addressed in [5]. Light field compression exploiting the correlation between adjacent views based on 3D-DCT is suggested in [6]. Objective and subjective performance evaluation of a few state-of-the-art algorithms for light field image compression is performed in [7].

There are also some new approaches to the light field image compression. Compression based on convolutional neural networks and linear approximation is introduced in [8]. Codec with disparity guided sparse coding over a learned perspective-shifted light field dictionary based on selected structural key views is proposed in [9]. Homography-based low-rank approximation light field compression is evaluated in [10]. Novel pseudo sequence based 2D hierarchical reference structure for the light field image compression is presented in [11]. Plenoptic image compression via simplified subaperture projection is performed in [12].

Although most of these advanced solutions successfully exploits the inter view correlation inside light field image in one way or another, there is little to no documentation about a compression approach based on the 4D-DCT.

3. Proposed Coding Solution

This encoding solution is an extension to the JPEG baseline compression, so this section is based on ITU-T Recommendation T.81. To fully understand the solution, it is recommended to first consult the text.

An input to the encoder is a light field image as a 2D array of subaperture views. This is a native format for camera array, but preprocessing is needed for lenslet images from light field camera such as Lytro. The first part of the compression chain is a conversion to YCbCr or other color space suitable for compression. The components of the image are compressed separately. Part 1 in Figure 2 is representing this process. It is desired that the values of samples are level shifted to the signed representation. This is done by subtracting 2^{P-1} from all samples, where P bit precision of the sample. When $P = 8$, the level shift is by $2^{8-1} = 128$.

Image component is partitioned into blocks of $8 \times 8 \times 8 \times 8$ samples. This process is represented by part 2 in Figure 2. If the block exceeds the image component boundary, the rest of the block shall be filled with nearest edge sample values.

The 4D forward discrete cosine transform is calculated on the block. This is done by performing 1D-DCT in Eq. (1) to all four dimensions of the block. This is represented by part 3 in Figure 2. The DCT transforms the block from spatial domain to frequency

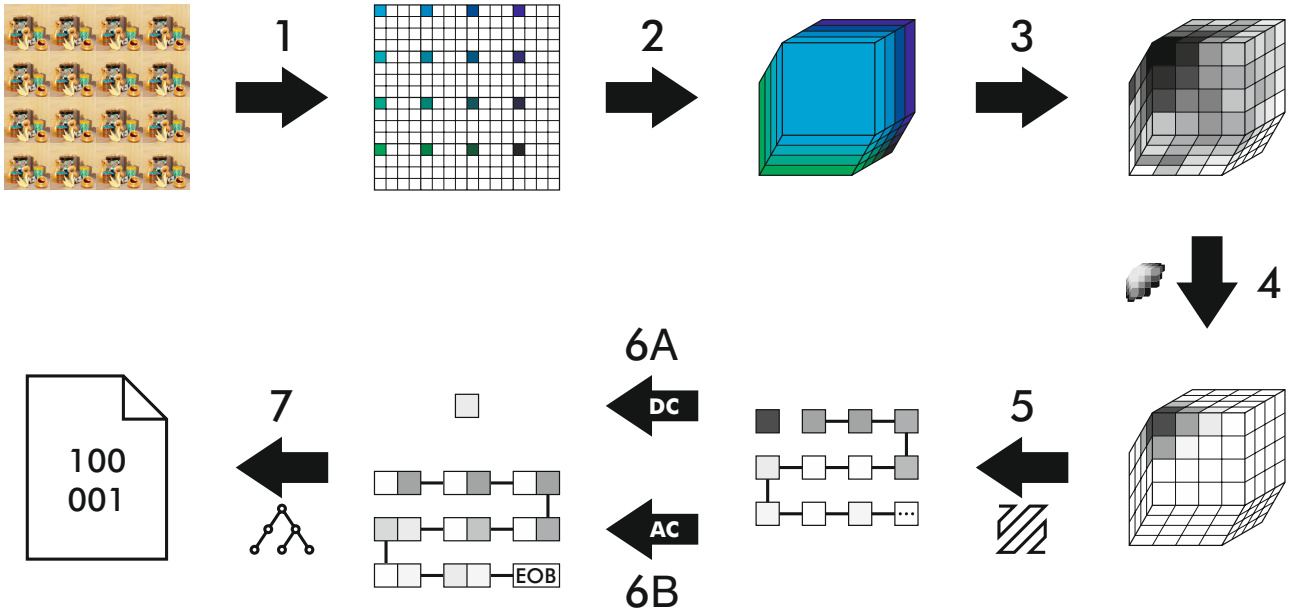


Figure 2. Processing chain of proposed light field encoder. 1) An input image is converted to suitable color space (YCbCr) and value range. The encoder then encodes every component separately. 2) Image component is partitioned into 4D blocks. The colored squares are a representation of 2D blocks of 4×4 samples. Blocks at corresponding positions from 4×4 adjacent views are combined into one 4D block. In this figure, the edge of the block is 4 samples for the sake of simplicity. In a real implementation, the blocks are 8 samples wide in each dimension. 3) 4D-DCT is performed on the block. Energy is concentrated around the DC coefficient. 4) Coefficients are quantized with 4D quantization matrix. 5) DC coefficient is treated separately. AC coefficients are traversed with a zig-zag or other 4D traversal algorithms to create a 1D array of samples. 6A) DC coefficient is DPCM encoded. 6B) AC coefficients are run-length encoded according to a number of consequent zeroes. 7) DC coefficients and AC pairs are entropy encoded to the final bitstream by Huffman encoder.

domain and concentrates the energy around the DC coefficient.

$$X_k = \sqrt{\frac{2}{N}} c_k \sum_{n=0}^{N-1} x_n \cos\left[\frac{(2n+1)k\pi}{2N}\right] \tag{1}$$

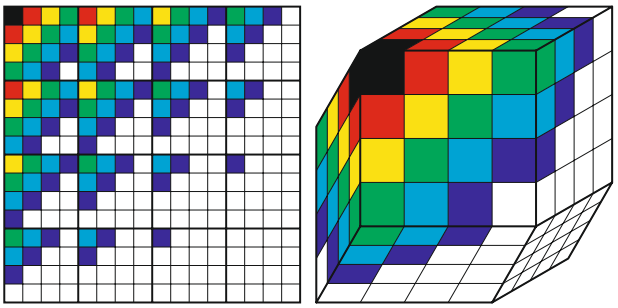
$$c_k = \begin{cases} 1/\sqrt{2} & \text{if } k = 0 \\ 1 & \text{otherwise} \end{cases}$$

The lossy compression is based on the fact that the human eye is more sensitive to low frequencies and less sensitive to high frequencies in an image. This means that higher frequencies can be reduced with no significant impact on quality. This is done in the quantization step. A quantization matrix is applied to coefficients by the formula in Eq. (2), where X is the block of DCT coefficients, Q is quantization matrix, X_q is the block of quantized coefficients and $\mathbf{k}$ is an index of a coefficient. Quantization step is represented by part 4 in Figure 2.

$$X_q(\mathbf{k}) = \text{round}\left[\frac{X(\mathbf{k})}{Q(\mathbf{k})}\right] \tag{2}$$

For the purposes of this encoder, a new 4D quantization schemes were developed to match coefficient distribution in the block. The easiest quantization scheme is the uniform quantization where the coefficients in a block are divided by a constant value. This scheme does not exploit human vision in any way, so the performance of the compression is not optimal. The other quantization scheme is an application of 2D quantization matrix to every 2D slice of the 4D block. This exploits the inter, but not the intra correlation. The most convenient scheme for the solution is to calculate average quantization value for each diagonal in the existing 2D matrix and fill this value to the corresponding diagonal in the 4D quantization matrix. This will successfully exploit intra and inter view correlation. Figure 3 shows example of diagonals in a 4D volume. Even better quantization scheme could be achieved by measuring psychovisual thresholds for 4D hypervolume compression, but this is out of the scope of this paper.

The quantized AC coefficients are traversed in an effective way to concentrate as much energy as possible near the DC coefficient in a one-dimensional structure. This is represented by part 5 in Figure 2. This paper proposes two effective approaches. The computationally inexpensive approach is to use 4D zig-zag sequence. The zig-zag sequence for 4D block works



(a) Orthographic view. **(b)** The same block unfolded to a 2D plane.

Figure 3. Diagonals in a 4D block. Samples in each diagonal have different color. Black sample is the DC coefficient, red samples are the first diagonal, yellow samples are the second diagonal, etc.

the same way as zig-zag for 2D blocks. For every diagonal in 4D volume, the coefficients are scanned by a 3D zig-zag algorithm. Scans are then connected in such a way that diagonals are in increasing order and outer coefficients of every diagonal scan are adjacent in the 4D block. An alternative to zig-zag is a scanning sequence based on a reference block. The reference block is averaged from absolute values from a few selected blocks or from all blocks of the image. A scan that would sort values from a reference block in descending order is then performed on all blocks. This approach is computationally expensive due to one extra compression pass to construct the reference block but results in a slightly better compression ratio.

The DC coefficient $X_q(0,0,0,0)$ is treated separately from the other 4095 AC coefficients. The value that shall be encoded is the difference between the quantized DC coefficient of the current block and that of the previous block. This step is called differential pulse-code modulation (DPCM) and is represented by part 6A in Figure 2.

After AC coefficients are scanned into a one-dimensional structure, the run-length encoding is performed. This is done in the same manner as in JPEG baseline, so the description here will be brief. The output of the run-length encoding is a list of pairs (run-length, amplitude), where run-length is a number of consequent zeroes before the nonzero coefficient and amplitude is said nonzero coefficient. Two special cases exist. The pair (15, 0) serves as a substitute for 16 consequent zeroes. This is needed to fit run-length value into 4 bits. The pair (0, 0), also called End Of Block (EOB), serves to indicate that remaining coefficients in a block are zero. For example, consider AC coefficient sequence -3, 0, -3, -2, 0, 0, 0, -4, 0, 1, 0, 0, 0... and assume that remaining coefficients to the end of the block are zero. The output of a run-length encoding would be (0, -3), (1, -3), (0, -2), (3, -4), (1, 1), (EOB). Run-length encoding is represented by part 6B in Figure 2.

The last step is Huffman encoding. Sizes of both DC and AC symbols are 8 bits. The symbol for the AC coefficient consists of run-length value in 4 most significant bits and the minimum number of bits needed to represent encoded amplitude in 4 least significant bits. The symbol for the DC coefficient consists entirely of DPCM encoded amplitude bit size as there is no run-length encoding performed. Huffman encoding is represented by part 7 in Figure 2.

Optimal Huffman table can be constructed from the input component(s) in an extra pass, or suboptimal precomputed Huffman table can be used. In the proposed solution, separate Huffman tables exist for DC and AC coefficients. Amplitudes shall be encoded to the final bitstream by writing Huffman codeword followed by the coefficient value in the number of bits specified in the encoded symbol.

The decoding is done by reversing encoding process. Input bitstream is Huffman decoded. The DC coefficients are DPCM decoded with the use of DC value of the previous block of the same component. AC coefficients are run-length decoded and scanned in reversed manner. Coefficients in a block are dequantized by the same matrix used for quantization as in Eq. (3).

$$X(\mathbf{k}) = X_q(\mathbf{k}) \times Q(\mathbf{k}) \tag{3}$$

The block is inverse cosine transformed to spatial domain. Samples are decomposed from four dimensional block into two dimensional component and inverse shifted by adding 2^{P-1}. The color space conversion is performed if needed.

4. Experimental Results

The dataset used for the experimental evaluation of the proposed method consists of different types of light field images from three internet archives. Thumbnails of images from each category are in Figure 4. First three images are part of The (New) Stanford Light Field Archive[2]. These images were captured with lego gantry as 17×17 views with high to medium average disparity. Next three images are from HCI dataset[3] [13]. These images were synthetically generated as 9×9 views with medium to low disparity. Last three images are from EPFL Light Field Image Dataset[4] [14]. These images were captured by Lytro Illum B01

[2]Stanford dataset – https://lightfield.stanford.edu/

[3]HCI dataset – http://hci-lightfield.iwr.uni-heidelberg.de/

[4]EPFL dataset – http://mmspg.epfl.ch/EPFL-light-field-image-dataset/

Figure 4. Images used to test proposed encoding solution. From left to right, top to bottom: Treasure Chest, Chess, Amethyst, Origami, Herbs, Kitchen, Bikes, Flowers and Friends 1.

(10-bit) camera and converted to 15×15 views with low disparity. The color depth was reduced to 8 bits per channel.

A reference implementation of the proposed method (lfif4D) is compared with its 3D alternative (lfif3D), which exploits inter view correlation in only one dimension to measure the gain of higher dimensional compression. Both of these implementations can be found as supplementary material for this paper. The solution is implemented as a C++ library and serves as a proof of concept proposed in this paper. The proposed method is compared with JPEG implemented as mozjpeg[5] library and HEVC intra encoding implemented as x265[6] library.

Peak signal-to-noise ratio (PSNR) metric is used for evaluation. It is calculated from the average mean square error (MSE) of each channel from the original and encoded RGB image. PSNR is presented versus compressed image bitrate. Results for each light field image type were interpolated with a spline and averaged.

Results of tests performed on lego gantry images are in Figure 5. The results show that HEVC intra compression is superior with high disparity images. The average disparity is high enough to make a correlation between views in a block scope almost non-existent. This results in worse compression performance due to compression of uncorrelated data together in one block. The proposed method is becoming less effective with increasing bitrate to be surpassed by JPEG and 3D method around two bits per pixel.

[5]mozjpeg library – https://github.com/mozilla/mozjpeg/
[6]x265 library – http://x265.org/

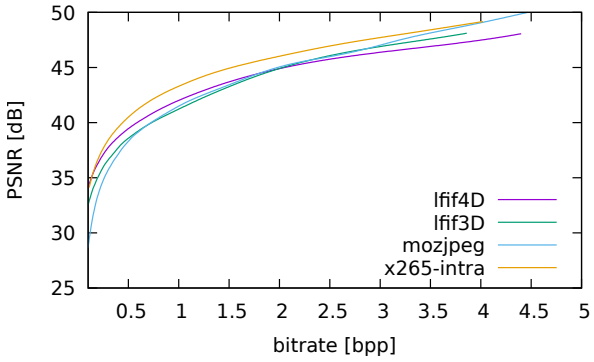


Figure 5. Results for high average disparity images.

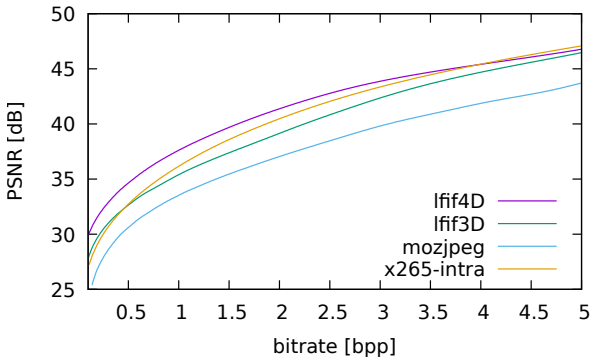


Figure 6. Results for medium average disparity images.

Results for synthetically generated images are in Figure 6. The results show that the proposed method exceeds the HEVC intra for most of the bitrate range. This is because the correlation in a block scope is strong enough to be exploited to surpass the HEVC intra predictions. The 3D method is slightly better than HEVC intra for bitrates as low as 0.5 bits per pixel. The JPEG is the worst possible solution for the bitrate range. This is expected because JPEG is the most simple method in the tests.

Results for images acquired with the Lytro Illum camera are in Figure 6. The results show that the proposed method is superior for this type of images. The disparity is low enough to make correlation very noticeable even in one block scope. This correlation is then fully exploited by the 4D-DCT. The 3D method is worse than the 4D method, which means there is a noticeable gain for higher dimensional compression. HEVC intra is slightly worse than the 3D method and JPEG is the worst of the compared solutions as expected.

5. Conclusions

In this paper, a new light field compression method based on 4D-DCT was proposed. The method is extending the JPEG baseline compression standard to four dimensions to exploit both intra and inter view

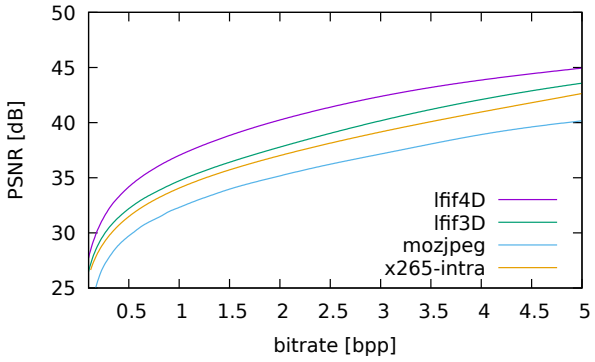


Figure 7. Results for low average disparity images.

correlation in light field image. Solutions to 4D quantization and block traversal were presented in this paper.

The proposed solution was compared with its 3D alternative, JPEG and HEVC intra encoders. The results show that the proposed solution is superior with light field images with a lower disparity between views, such as images captured by Lytro camera or images synthetically generated. The inter view correlation was successfully exploited by the 4D-DCT, therefore it is possible that other transformation-based compression methods extended to the fourth dimension would behave the same way. This knowledge can be crucial for future work on advanced light field compression schemes.

The proposed solution can be further improved by developing new 4D quantization matrices that would more precisely exploit human eye properties based on experiments with psycho-visual thresholds.

Acknowledgements

I would like to thank my supervisor Ing. David Bařina, Ph.D. for his help.

References

[1] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, and Peter Schelkens. JPEG Pleno: Toward an efficient representation of visual reality. *IEEE MultiMedia*, 23(4):14–20, 2016.

[2] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[3] Gustavo Alves, Fernando Pereira, and Eduardo A. B Da Silva. Light field imaging coding: Performance assessment methodology and standards benchmarking. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2016.

[4] Rogério Seiji Higa, Roger Fredy Larico Chavez, Ricardo Barroso Leite, Rangel Arthur, and Yuzo Iano. Plenoptic image compression comparison between JPEG, JPEG 2000 and SPITH. 2013.

[5] C. Perra and D. Giusto. JPEG 2000 compression of unfocused light field images based on lenslet array slicing. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pages 27–28, Jan 2017.

[6] A Aggoun. A 3D DCT compression algorithm for omnidirectional integral images. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II. IEEE, 2006.

[7] Irene Viola, Martin Rerabek, and Touradj Ebrahimi. Comparison and evaluation of light field image coding approaches, 2017.

[8] Nader Bakir, Wassim Hamidouche, Olivier Deforges, Khouloud Samrouth, and Mohamad Khalil. Light field image compression based on convolutional neural networks and linear approximation. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1128–1132. IEEE, 2018.

[9] Jie Chen, Junhui Hou, and Lap-Pui Chau. Light field compression with disparity-guided sparse coding based on structural key views. *IEEE Transactions on Image Processing*, 27(1):314–324, 2018.

[10] Xiaoran Jiang, Mikael Le Pendu, Reuben A Farrugia, and Christine Guillemot. Light field compression with homography-based low-rank approximation. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1132–1145, 2017.

[11] Li Li, Zhu Li, Bin Li, Dong Liu, and Houqiang Li. Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1107–1119, 2017.

[12] Haixu Han, Jin Xin, and Qionghai Dai. *Plenoptic Image Compression via Simplified Subaperture Projection: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part II*, pages 274–284. 09 2018.

[13] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4D light fields. In *Asian Conference on Computer Vision*. Springer, 2016.

[14] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. 2016.