Compression scheme for sparsely sampled light field data based on pseudo multi-view sequences

Waqas Ahmad, Mårten Sjöström, and Roger Olsson

Mid Sweden University, Department of Information Systems and Technology,
Homolgatan-10, Sundsvall, Sweden

ABSTRACT

With the advent of light field acquisition technologies, the captured information of the scene is enriched by having both angular and spatial information. The captured information provides additional capabilities in the post processing stage, e.g. refocusing, 3D scene reconstruction, synthetic aperture etc. Light field capturing devices are classified in two categories. In the first category, a single plenoptic camera is used to capture a densely sampled light field, and in second category, multiple traditional cameras are used to capture a sparsely sampled light field. In both cases, the size of captured data increases with the additional angular information. The recent call for proposal related to compression of light field data by Joint Picture Expert Group (JPEG), also called JPEG Pleno, reflects the need of a new and efficient light field compression solution. In this paper, we propose a compression solution for sparsely sampled light field data. Each view of multi-camera system is interpreted as a frame of multi-view sequences. The pseudo multi-view sequences are compressed using state-of-art Multiview-extension of High Efficiency Video Coding (MV-HEVC). A subset of four light field images from Stanford dataset are compressed, on four bit-rates in order to cover the low to high bit-rates scenarios. The comparison is made with state-of-art reference encoder HEVC and its real-time implementation x265. The rate distortion analysis shows that the proposed compression scheme outperforms both reference schemes in all tested bit-rate scenarios for all the test images. The average BD-PSNR gain of 1.36 dB over HEVC and 2.15 dB over x265 is achieved using the proposed compression scheme.

Keywords: Light field, MV-HEVC, Compression, Plenoptic, Multi-Camera

1. INTRODUCTION

The information of light in a observable space is entirely represented by the seven-dimensional (7D) plenoptic function.[1] The plenoptic function takes into account the spatial position (3D position in space), direction (2D angular information), time (1D) and wavelength (1D) of all the light rays within the space to be observed. The physical limitations and capabilities of present technology requires to approximate the plenoptic function into lower dimensional space. A set of carefully developed assumptions are constructed in order to reduce the dimensions of plenoptic function. One of the mostly used approximation of plenoptic function is 4D light field.[2] In light field, the observable scene is anticipated as static that eliminates the need to capture the time information. The wavelength is discrtized with red, green and blue channels and finally the scene is assumed as free of occluders that enables the capturing of incoming light rays on a 2D plane. The captured 4D light field contains angular and spatial information of the scene that enables 3D scene reconstruction and various post-processing applications, i.e. refocusing, synthetic aperture, 3D scene reconstruction.

Intially, light field acquisition was performed by using a system of multiple traditional cameras.[2] Each camera, captures a single angular information (perspective) of the scene and the number of cameras in the multi-camera system defines the angular resolution of the captured light field. In recent past, new technologies are introduced that captures the light field using single camera referred to as plenoptic camera. The idea of plenoptic capturing was first reported by Gabriel Lippmann in 1908.[3] In 2006, Ren Ng at Lytro introduced the first commercial

Further author information: (Send correspondence to M.S)
W.A: E-mail: waqas.ahmad@miun.se, Telephone: +46-010-142 87 22
M.S: E-mail: Marten.Sjostrom@miun.se, Telephone: +46-010-142 88 36
R.O: E-mail: Roger.olsson@miun.se, Telephone: +46-010-142 86 98

model of a plenoptic camera[4] that use an array of small lenses between main lens and photographic plate to capture spatial and angular information onto a single image. The main lens and micro-lens array of plenoptic camera are configured in such a way that each micro-lens image records the angular information of a point in space.[4]The overall resulting image is referred to as a plenoptic image. The multi-camera system captures sparsely sampled light field and on the other hand plenoptic camera captures densely sampled light field. The capturing of spatial and angular information provide the possibility of various post-processing applications but on the expense of increased in the data size.

Over the span of last two decades, various compression standards are proposed for image and video data. The image compression techniques exploits the spatial redundancies present in the image by employing multi-resolution and predictive models. JPEG 2000 standard, achieved image compression by means of bi-orthogonal wavelet transform.[5] The transform coefficients are quantized and entropy coded. The state-of-art High efficiency video coding (HEVC) Intra compresses images by using novel prediction modes along with Discrete Cosine Transform (DCT) and residual coding.[6] The video compression schemes also take into account the temporal correlation present in the data by using motion estimation and compensation techniques. Moreover, Multi-view extension of HEVC (MV-HEVC) provides additional capability to exploit inter-view correlation.[7] However, conventional image/video compression standard are developed for natural two-dimensional images and are not capable to exploit the properties of light field for efficient compression. The recent call for proposal by JPEG also referred to as JPEG Pleno reflects the requirement of new compression tools for light field data.[8] In this paper, we present a compression scheme for light field data capture with multi-camera system. Each view of captured LF is interpreted as frame of multi view sequence and pseudo multi-view sequence is given as input to MV-HEVC. A two dimensional prediction and rate-allocation scheme is used in order to better exploit the properties offered by LF data to improve compression efficiency.

The novelties of this paper are: 1) A coding scheme where the multi-view extension of HEVC (MV-HEVC) is used to compress LF images captured with multi-camera system. 2) A two-dimensional prediction and rate allocation scheme for LF data captured with multi-camera system. The paper is organized as follows: state-of-art in LF compression is discussed in section 2. In section 3, the proposed compression scheme is presented, and section 4 reports the results of the proposed scheme. The presented work is concluded in section 5.

2. LIGHT FIELD COMPRESSION

In recent past, light field compression has received greater attention from research community. Intially, the ICME grand challenge on plenoptic image compression yields different compression solutions for plenoptic images.[9] Later on, JPEG categorizes the light field compression activities in two groups, i.e. plenoptic image compression and multi-camera based light field compression. Recently, JPEG has issued a call for light field compression proposal also named as JPEG Pleno to further advance the state-of-art methods.[8] The compression solution for plenoptic image can be classified in two groups. In the first group, the captured plenoptic image is treated as a single 2D image and novel tools are introduced in HEVC image compression standard to efficiently exploit the correlation present in the plenoptic image. Li et al. has introduced the Bi-prediction capability within the framework of HEVC intra prediction.[10] In this way, each mirco-lens image perform motion estimation and compensation with already encoded neighbouring micro blocks in addition to 33 intra-modes. A similar strategy to take prediction from already encoded blocks was introduced by Monteiro et al. by including Local Linear Embedding-based (LLE) and Self-Similarity (SS) compensated based prediction tools in HEVC Intra coding framework.[11] In second group, the sub-aperture representation of plenoptic image was adopted to provide a more suitable input to the video compression tools. The idea of compressing integral images as pseudo video sequence[12] is extended for the sub-aperture images. Most prominently, Liu et al. interpreted sub-aperture images as frames of pseudo video sequence and compressed using HEVC.[13] Each frame was assigned with a specific quanitization parameter in order to achieve better compression efficiency. Ahmad et al. interpreted sub-aperture images as frames of pseudo multi-view sequence and performed compression using MVHEVC.[14] A 2D prediction and rate allocation scheme that exploits the light field properties was introduced in the frame work of MV-HEVC to obtain better compression efficiency.

Recently, few studies have been reported on compression of light field data captured with multi-camera system. Hawary et al. utilized the sparsity in the angular fourier domain of the captured light field to achieve

better compression efficiency.[15] A subset of views are encoded in base layer of HEVC and corresponding decoded version of the subset is used to predict the intermediate views. Firstly, the spatial frequencies of each decoded view are estimated using fourier transform and in the second stage, each spatial frequency among the views is taken as discrete line and again fourier transform is computed for each such line. A voting mechanism is used by following the projection slice theorem and integer frequencies are estimated. The underlying angular spectrum of light field is recovered by performing further refinement for the non-integer frequencies. In the work of Jiang et al., a homography-based low rank approximation is used for compression of light field data.[16]The central view is taken as a reference and remaining views are aligned using homography transformation. Iteratively, homographies are estimated that minimizes the low rank approximation error for a target rank k. The low rank representation of captured LF is encoded using HEVC intra compression tool.

The plenoptic image compression schemes utilizing the sub-aperture representation shows better compression efficiency compared to other schemes.[9] The sub-aperture images reflect the similar information as 2D natural images. The assumptions and constraints used for video compression suits better to sub-aperture representation of plenoptic image. In compression of LF captured with multi-camera system, novel but computational intensive tools in video compression frame work are introduced to achieve better compression efficiency. In this paper, we extended our previously proposed plenoptic image compression scheme for multi-camera system based LF data. The input LF views are treated as frames of pseudo multi-view sequences and already developed frame work of MV-HEVC is used for efficient compression.

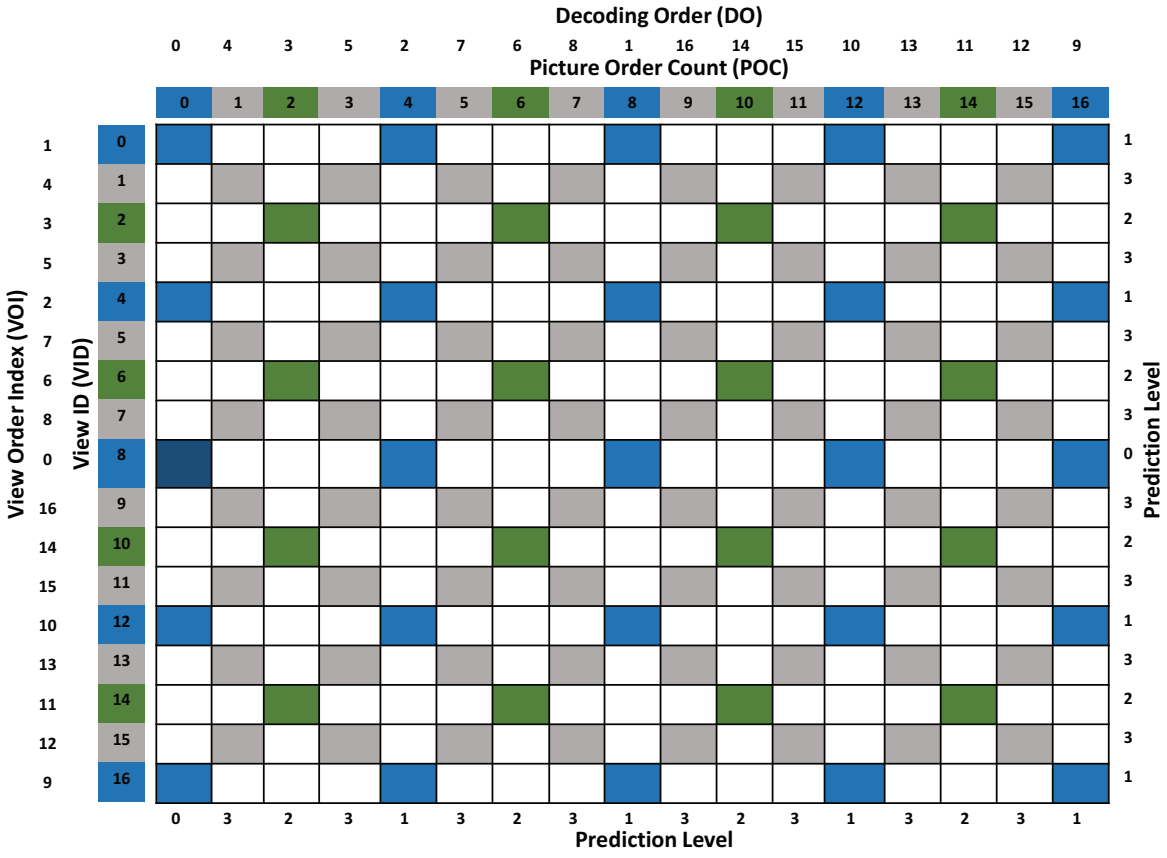

Figure 1. The 17x17 grid of the captured light field. The parameters of MV-HEVC are labelled and assigned prediction level for each frame is represented by blue, green and gray colors.

3. PROPOSED COMPRESSION SCHEME

The light field data captured with multi-camera system inhertily possess high correlation that can be categorized as spatial and angular (inter-view) correlation. The spatial correlation is predicted from the same view on the other hand the inter-view prediction is taken from 2D (horizontal and vertical) neighbouring views. In the proposed compression scheme, each view of the captured LF is interpreted as a frame of a pseudo multi-view sequence. In this way, the captured light field with MXN views is represented by M pseudo video sequences with each having N frames and compression is performed using MV-HEVC.[7] The framework of MV-HEVC allows each frame to take spatial, temporal and inter-view prediction. In our proposed scheme we cast the inter-view prediction into temporal prediction in order to have 2D inter-view prediction available for LF data. In LF, each camera captures the scene from slightly different perspective and hence high correlation is present in the neighbouring views.

3.1 Prediction and Rate allocation Scheme

A 2D prediction and rate allocation scheme is used to efficiently compress the light field data captured with multi-camera system. The figure 1 shows the representation of input LF along with MV-HEVC parameters. The Picture Order Count (POC) and view ID (VID) represent the position of each frame in the grid. The Decoding Order (DO) and View Order Index (VOI) reflects the decoding order of each frame in horizontal and vertical axis of the grid respectively. The POC is represented by variable n_{POC}, VID is represented by v_{VID}, DO is represented by k_{DO} and VOI is denoted by i_{VOI}. Each frame of pseudo multi-view sequences is assigned with a specific prediction level. The frame with ($v_{\mathrm{VID}} = 8$ and n_{POC}=0) is taken as a reference frame for the entire prediction scheme and assigned with the first prediction level ($PL = 0$). The remaining frames are categorized into further three prediction levels. The frames having $v_{\mathrm{VID}} = \{0, 4, 8, 12, 16\}$, and $n_{\mathrm{POC}} = \{0, 4, 8, 12, 16\}$ (shown in blue color) are assigned with second prediction level ($PL = 1$). These frames are encoded immediately after the reference frame. Moreover, the frames with $v_{\mathrm{VID}} = \{2, 6, 10, 14\}$, and $n_{\mathrm{POC}} = \{2, 6, 10, 14\}$ (shown in green color) are assigned with third prediction level ($PL = 2$). The frames with $v_{\mathrm{VID}} = \{1, 3, 5, 7, 9, 11, 13, 15\}$, and $n_{\mathrm{POC}} = \{1, 3, 5, 7, 9, 11, 13, 15\}$ (shown in gray color) are assigned with the last prediction level ($PL = 3$). In this way, both POC and VID axis are completely assigned with the prediction levels. The assigned prediction levels are shown in figure 1 (at the bottom and on the right side). In figure 1, frames depicted in white color are assigned with two different prediction levels in the horizontal and vertical direction. The prediction scheme is device in such a way that each frame can take prediction from the frames placed in higher prediction levels and also from the already encoded frames of the current prediction level. The first and second prediction level frames (shown by blue and green color) provide 2D prediction available for the remaining frames. In this way, compression scheme better handles the occluded regions present in the captured LF image and hence compression efficiency is improved.

The rate allocation process make use of already assigned prediction levels while distributing quality among the frames. The frames placed in higher prediction level are assigned with best available quality and at each successive level a slightly lower quality is assigned to the corresponding frames. In this way, the better quality frames are mainly utilized for prediction that improves the overall compression efficiency. An initial Quantization Parameter (QP) is assigned to the reference frame and referred to as Q_R. All the remaining frames add a quantization offset with Q_R to obtain their quantization parameter. The quantization offset of a frame is estimated based on its distance and view-wise decoding order relative to the reference frame. The methodology was previously proposed for plenoptic image compression[14] and in this work it is extended for LF captured with multi-camera system . The equation(1) presents the quantization offset estimation process for each frame.

$$Q_{\mathrm{o}}(x, y) = \lfloor \frac{|n_{\mathrm{POC}}(x) - R_{\mathrm{POC}}| + |v_{\mathrm{VID}}(y) - R_{\mathrm{VID}}|}{W} \rfloor + \lfloor \frac{(i_{\mathrm{VOI}} \mod R_{\mathrm{VID}})}{W} \rfloor \quad (1)$$

Where, x (0 to 16) and y (0 to 16) represent the index of each frame in the grid. The variable n_{POC} represents the POC of each frame, R_{POC} represents the POC of reference frame, v_{VID} represents the VID of each frame, R_{VID} represents the VID of reference frame, i_{VOI} represents the VOI of each frame, W represents the weightage parameter, Q_{o} indicates the estimated qunatization offset for each frame. The parameter W controls the extent

Table 1. The assigned weights to each prediction level

Predictor Levels	**Picture Order Count**			
View ID	**0**	**1**	**2**	**3**
0	Q_R	3	3	3
1	3	3	3	2
2	3	3	3	2
3	3	2	2	1.5

of quantization offset by normalizing it using the frame prediction level. Frames placed in higher prediction level are assigned with a large weightage value as compared to frames placed in lower prediction levels. The table 1 shows the assigned weightages for each prediction level used in the proposed compression scheme. Finally, the equation (2) estimates the quantization parameter (Q) for each frame.

$$Q(x,y)=\begin{cases} Q_{\mathrm{R}} + P_{\mathrm{Max}}(x,y), & \text{if } x=R_{\mathrm{POC}} \text{ or } y=R_{\mathrm{VID}} \\ Q_{\mathrm{R}}+Q_{\mathrm{o}}(x,y), & \text{otherwise} \end{cases} \tag{2}$$

All the frames estimate the quantization offset as explained in equation (1). However, the frames having either reference POC or reference VID are assigned a quanitzation offset equivalent to maximum of their prediction level($P_{\mathrm{MAX}}(x,y) = \max(s_{\mathrm{POC}}(x), t_{\mathrm{VID}}(y))$). The prediction levels in POC axis and VID axis are represented by s_{POC} and t_{VID}.

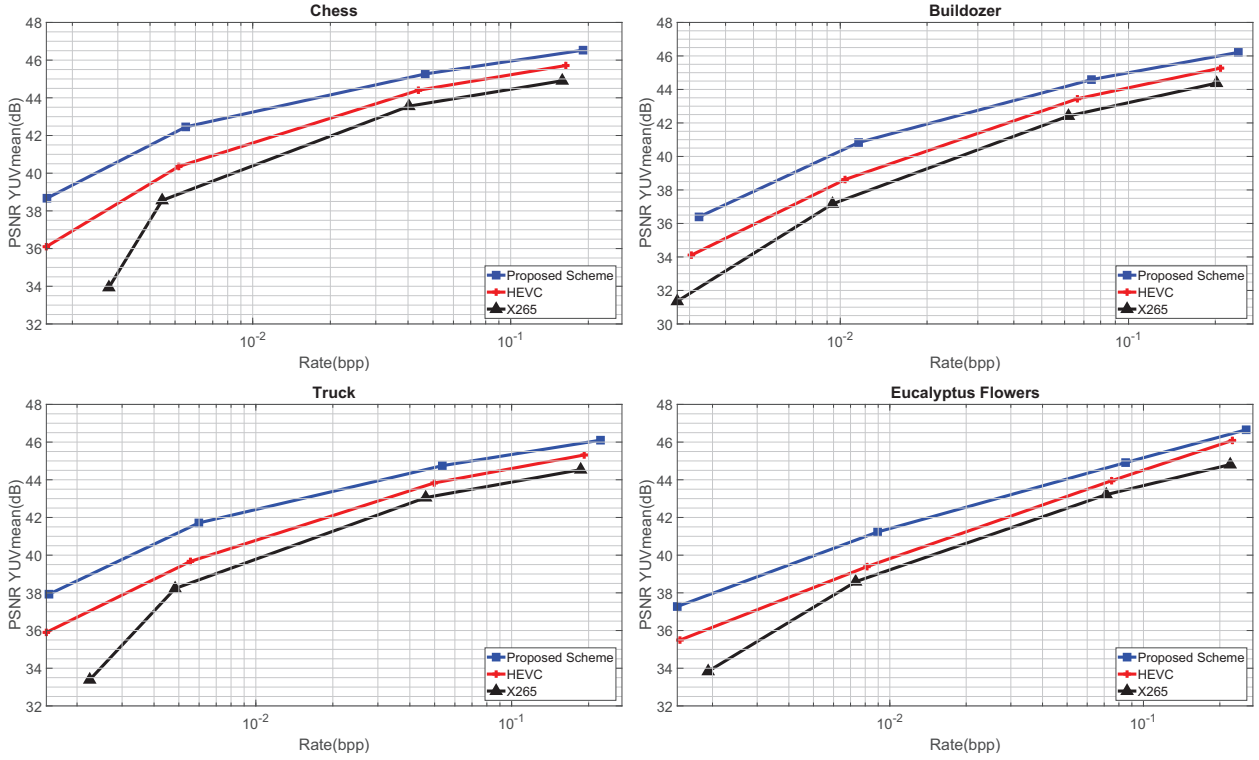

Figure 2. Rate distortion analysis for the proposed compression scheme compared to HEVC and x265 anchor schemes.

Table 2. BD-PSNR analysis of proposed scheme compared to HEVC and x265 benchmark schemes

Image ID	**BD-PSNR-YUV(dB)**	
	HEVC	x265
Chess	1.39	1.93
Lego Bulldozer	1.43	2.50
Lego Truck	1.34	2.14
Eucalyptus Flowers	1.28	2.02
Average	2.15	1.36

4. RESULTS AND ANALYSIS

The light field dataset captured with multi-camera system provided by the Stanford university is used to evaluate the proposed compression scheme.[17] The dataset contains 12 LF images with each having 17x17 views. The images contains various characteristic,e.g. Truck contains complex geometry and Bunny contains texture information. The 17x17 input views are converted into 17 pseudo video sequences with each having 17 frames and compressed using MV-HEVC. The proposed compression scheme is evaluated on four different bit-rates in order to cover low, medium and high bit-rates scenarios. The comparison is made with two benchmark compression schemes HEVC and its real time implementation x265 using the BD-PSNR metric.[18] The input LF of 17x17 views is converted into a single pseudo video sequence of 289 frames and given as a input to the anchor schemes. Figure 2 shows the rate distortion curves for a subset of four LF images taken from stanford dataset.[17] The proposed compression scheme performs significantly better compared to the anchor schemes in all the test bit-rates. The compression efficiency is significantly higher in low bit-rates and the efficiency reduces in high bit-rates. The table 2 reports the BD-PSNR gain of the proposed scheme over the anchor schemes on four LF images. The average PSNR gain of 2.15 dB over x265 and 1.36 over HEVC is achieved using the proposed compression scheme.

Table 3. Compression details of 12 LF images compressed using proposed scheme

Image ID	**Rate (Bytes)**				**PSNR-YUV(dB)**			
	R1	R2	R3	R4	R1	R2	R3	R4
Chess	47720407	1887381	222509	64488	46.5	45.2	42.4	38.6
Bulldozer	15365523	4754190	739989	206846	46.2	44.5	40.8	36.3
Truck	9879988	2378392	265759	68959	46.1	44.7	41.7	37.9
Eucalyptus Flowers	18041808	6038023	635705	103158	46.6	44.9	41.2	37.2
Amethyst	8496744	2740824	285598	61914	45.8	44.3	40.8	36.7
Bracelet	9771763	4056819	401004	67446	45.5	43.7	40	35.8
The Stanford Bunny	5755549	966780	105215	39827	46.3	45.4	43.1	39.8
Jelly Beans	1754545	453705	74025	28139	47.3	46.7	44.7	41.1
Lego Knights	9518414	3384609	589781	165466	45.8	44.4	40.8	36.2
Tarot Cards and Crystal Ball	30195013	12896580	1694254	409137	44.2	42	37.6	32.7
Treasure Chest	26223020	10496585	1211768	217551	45.7	44	40	35.5
Lego Gantry Self Portrait	8276839	3033339	426015	87670	45.2	43.6	40	35.9

The complete LF dataset from the Stanford University[17] is compressed using the proposed compression scheme and experimental details are presented in table 3. The first column shows the name of each LF image, next four columns (from 2 to 5) show the size of compressed LF image at each bit-rate in Bytes (B). The columns 6 to 9 show the corresponding PSNR value (for YUV channels) for each bit-rate.

5. CONCLUSION

In this paper, we present a compression scheme for sparsely sampled light field data captured with multi-camera system. The views of captured LF are interpreted as frames of multi-view sequences and are compressed using MV-HEVC. The multi-view extension of HEVC provides inter-view correlation (in vertical direction) and temporal correlation (in horizontal direction). In this way, MV-HEVC frame work provides an opportunity to each LF frame to take prediction from already encoded 2D neighbouring frames to improve compression efficiency. The Stanford dataset (containing 12 LF images) is compressed using proposed compression scheme and Rate-distortion results are reported in the paper. The BD-PSNR based comparison with two benchmark compression schemes HEVC and x265 is also presented for a subset of four LF images. The proposed compression scheme provides an average gain of 1.36 dB over HEVC and 2.15 dB over x265.

ACKNOWLEDGMENTS

The work in this paper was funded from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 676401, European Training Network on Full Parallax Imaging.

REFERENCES

[1] Adelson, E. H. and Bergen, J. R., "The plenoptic function and the elements of early vision," (1991).

[2] Levoy, M. and Hanrahan, P., "Light field rendering," in [*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*], 31–42, ACM (1996).

[3] Lippmann, G., "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.* **7**(1), 821–825 (1908).

[4] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P., "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR* **2**(11), 1–11 (2005).

[5] Skodras, A., Christopoulos, C., and Ebrahimi, T., "The jpeg 2000 still image compression standard," *IEEE Signal processing magazine* **18**(5), 36–58 (2001).

[6] Sullivan, G. J., Ohm, J., Han, W.-J., and Wiegand, T., "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology* **22**(12), 1649–1668 (2012).

[7] Tech, G., Chen, Y., Müller, K., Ohm, J.-R., Vetro, A., and Wang, Y.-K., "Overview of the multiview and 3d extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology* **26**(1), 35–49 (2016).

[8] Ebrahimi, T., Foessel, S., Pereira, F., and Schelkens, P., "Jpeg pleno: Toward an efficient representation of visual reality," *Ieee Multimedia* **23**(4), 14–20 (2016).

[9] Rerabek, M., Bruylants, T., Ebrahimi, T., Pereira, F., and Schelkens, P., "Icme 2016 grand challenge: Light-field image compression," *Call for proposals and evaluation procedure* (2016).

[10] Li, Y., Olsson, R., and Sjöström, M., "Compression of unfocused plenoptic images using a displacement intra prediction," in [*Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*], 1–4, IEEE (2016).

[11] Monteiro, R., Lucas, L., Conti, C., Nunes, P., Rodrigues, N., Faria, S., Pagliari, C., da Silva, E., and Soares, L., "Light field hevc-based image coding using locally linear embedding and self-similarity compensated prediction," in [*Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*], 1–4, IEEE (2016).

[12] Olsson, R., Sjostrom, M., and Xu, Y., "A combined pre-processing and h. 264-compression scheme for 3d integral images," in [*Image Processing, 2006 IEEE International Conference on*], 513–516, IEEE (2006).

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

[14] Ahmad, W., Olsson, R., and Sjöström, M., "Interpreting plenoptic images as multi-view sequences for improved compression," in [*Image Processing, 2017 IEEE International Conference on*], 513–516, IEEE (2017).

[15] Hawary, F., Guillemot, C., Thoreau, D., and Boisson, G., "Scalable light field compression scheme using sparse reconstruction and restoration," in [*ICIP 2017*], (2017).

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

[17] Vaish, V. and Adams, A., "The (new) stanford light field archive," *Computer Graphics Laboratory, Stanford University* (2008).

[18] Bjontegaard, G., "Calculation of average psnr differences between rd-curves," *ITU SG16 Doc. VCEG-M33* (2001).