INTERPRETING PLENOPTIC IMAGES AS MULTI-VIEW SEQUENCES FOR IMPROVED COMPRESSION

Waqas Ahmad, Roger Olsson, Mårten Sjöström

Mid Sweden University
Department of Information Systems and Technology
SE-851 70 Sundsvall Sweden

ABSTRACT

Over the last decade, advancements in optical devices have made it possible for new novel image acquisition technologies to appear. Angular information for each spatial point is acquired in addition to the spatial information of the scene that enables 3D scene reconstruction and various post-processing effects. Current generation of plenoptic cameras spatially multiplex the angular information, which implies an increase in image resolution to retain the level of spatial information gathered by conventional cameras. In this work, the resulting plenoptic image is interpreted as a multi-view sequence that is efficiently compressed using the multi-view extension of high efficiency video coding (MV-HEVC). A novel two-dimensional weighted prediction and rate allocation scheme is proposed to adopt the HEVC compression structure to the plenoptic image properties. The proposed coding approach is a response to ICIP 2017 Grand Challenge: Light field Image Coding. The proposed scheme outperforms all ICME-contestants, and improves on the JPEG-anchor of ICME with an average PSNR gain of 7.5 dB and the HEVC-anchor of ICIP 2017 Grand Challenge with an average PSNR gain of 2.4 dB.

***Index Terms*—** Light field, plenoptic, MV-HEVC

1. INTRODUCTION

The visible light with a observable space is completely represented by the seven-dimensional plenoptic function [1] that takes into consideration spatial position and direction, wavelength, and time for all light rays within this space. The 7D light information may be uniquely reduced into 4D referred to as a light field [2]. The 4D light field contains both angular and spatial information and the resulting light ray dataset enables 3D scene reconstruction and various post-processing effects. The light field can be acquired in different ways. Using a system of multiple traditional cameras [2] or with a lenslet array attached to a single camera where each lenslet projection records angular information through a specific spatial point, as first reported by Gabriel Lippmann in 1908 [3]. The approach to use an array of small lenses on top of photographic plate was later used in 2006, when Ren Ng at Lytro introduced the first commercial model of a plenoptic camera [4]. For this type of light field the image created by each micro lens is referred to as an *elementary image* and the overall resulting image is referred to as a *plenoptic image*. This way of acquiring angular information implies a tradeoff between spatial and angular resolution. To retain the same amount of spatial information as a conventional cameras an increase in sensor resolution is required, which increase the resulting image size. Moreover, the captured plenoptic image implies two types of correlations, i.e. angular correlation within each micro-lens image, and spatial correlation between micro-lens images.

Generally image compression techniques aim to de-correlate the data by exploiting redundancies in the image by e.g. employing multi-resolution and prediction models. The JPEG 2000 digital image compression standard achieved compression by means of bi-orthogonal wavelet transform [5]. However, conventional image encoders are developed for natural two-dimensional images and do not efficiently compress plenoptic images due to their correlation properties, and could be compressed more efficiently if transformed into pseudo video sequences [6]. The high efficiency video codec (HEVC/H.265) uses discrete cosine transform for compressing video frames [7]. For video, additional prediction models are used to address the temporal redundancy, minimizing the difference between consecutive frames by means of motion compensation and estimation techniques.

The novelties of this paper are: 1) A coding scheme where the multi-view extension of HEVC (MV-HEVC) is used to compress the plenoptic image in the form of a multi-view sequence. 2) A two-dimensional level based prediction scheme that controls the prediction structure within the frames of the multi-view sequence. 3) A method to calculate individual quantization parameter for each frame by using the parameters accessible in the multi-view encoder. The paper is orga-

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.

nized as follows: state-of-art in plenoptic image compression is discussed in section 2. In section 3, the selected representation of plenoptic image is explained, section 4 explains the proposed compression scheme, section 5 gives the details of the experimental setup, and section 6 presents the results of the proposed scheme. The presented work is concluded in section 7.

2. PLENOPTIC IMAGE COMPRESSION

In recent past, various studies have reported on compression of plenoptic images, as presented by [8]. Liu et al. convert the plenoptic image to perspective or sub-aperture images that are considered as frames in a pseudo video sequence compressed using HEVC single layer compression [9]. Neighboring frames are assigned different temporal id's in order to use them for the prediction . The frames are divided into layers, each having a specific compression ratio governed by a selected quantization parameter value. In the study of Li et al., compression is performed on the plenoptic image directly using a modified HEVC encoder [10]. Inter- and bi-prediction capability is provided within the intra prediction module based on references taken from already encoded parts of the image. Monteiro et al. have used Local Linear Embedding-based (LLE) and Self-Similarity (SS) compensated prediction as additional tools in HEVC for coding plenoptic images [11]. The LLE method estimates the current block as a linear combination of k nearest neighbor patches. In addition, the best match between current block and already reconstructed blocks is estimated and signaled as SS vector. The LLE and SS mode predictions are compared on the basis of Rate Distortion Optimization (RDO) and the most efficient approach is used. Conti et al. have also utilized the SS-based scheme for compressing plenoptic images [12]. Perra and Assuncao have interpreted the plenoptic image as a single layer pseudo sequence [13]. The plenoptic image is partitioned into non-overlapping tiles and the tiles are then considerd as consecutive frames of a pseudo video sequence and compressed using HEVC.

A rate-distortion analysis of the above schemes supports the conclusion that the sub-aperture based compression approach is efficient, especially in low bitrate scenarios [8]. The other presented schemes show significant compression efficiency in high bit rates. Important to note that the sub-aperture based plenoptic image compression scheme has not utilized the tools available in multi-view extension of HEVC [14]. The other presented compression schemes doesn't take into account the full spatial correlation structure in the plenoptic image. These compression schemes rely on block based prediction which is restricted to a search window containing previously compressed neighbouring blocks, which reduce the compression efficiency possible to achieve.

In our proposed method, the plenoptic image is converted to perspective images and interpreted as frames in a multi-view sequence compressed using MV-HEVC [15] . A two dimensional prediction and rate allocation scheme is proposed to efficiently assign a prediction structure within the multi-view sequence, and distribution of quantization parameters to achieve a favorable distribution of rate and distortion within the image set.

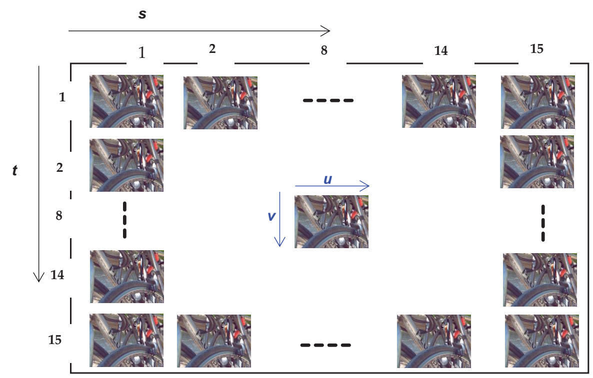

Fig. 1: A plenoptic image represented by a set of 15-by-15 sub-aperture images. s and t represent the horizontal and vertical position, u and v are the pixel coordinates in sub-aperture image plane.

3. PLENOPTIC IMAGE REPRESENTATION

The Matlab Lytro toolbox is used to convert the plenoptic image (with a resolution of 7728x5368 px) to a set of 15-by-15 perspective images, as shown in Fig. 1 [16]. From hereafter, we denote perspective images as *sub-aperture images*. Each sub-aperture image (with a resolution of 625x434 pixels) depicts the scene from a slightly different position. In the proposed method, only the central 13-by-13 sub-aperture images are used for compression.

To facilitate comparison with state-of-the-art the proposed compression scheme has been evaluated with image input format and compression ratios of both the ICME 2016 and the ICIP 2017 grand challenge [17, 18]. Explicit results from this comparison is shown later in the paper as well as online [19].

4. PROPOSED METHOD

In the proposed compression scheme, each row of sub-aperture images (as shown in figure 1) is interpreted as a single view of a multi-view video sequence. In this way 13 views (with 13 frames each) are used as input for the MV-HEVC encoder. The Multi-View extension of HEVC allows each image to use temporal and inter-view prediction to efficiently exploit the correlation within each image and among neighboring sub-aperture images [15]. In plenoptic camera, scene is captured from different perspectives on a single image sensor that results in minor perspective change

Table 1: The weights for each frame based on its predictor level

Predictor Levels (PL)	**Picture Order Count**			
View ID	PL_0	PL_1	PL_2	PL_3
PL_0	QP_B	3	3	3
PL_1	3	3	3	2
PL_2	3	3	3	2
PL_3	3	2	2	1.5

among neighbouring views and this fact is efficiently utilized by multi-view extension of HEVC.

4.1. Prediction Scheme

A two-dimensional prediction scheme, shown in Fig. 2, is devised to classify sub-aperture images as frames to be compressed using MV-HEVC. The prediction scheme relies on two parameters, View ID (V_{ID}) and Picture Order Count (POC) that together uniquely identify a frame. Each axis (ViewID and POC) is further categorized into four predictor levels. In horizontal axis, the POC with index 0 is defined as predictor level-0 (PL_0). The POC with indexes {4,8,12} are defined as level-1 predictors (PL_1), which mainly use PL0 and other available level-1 predictor frames for prediction. Similarly, the POC indexes {2,6,10} are defined as level-2 predictors (PL_2), which use PL_1 and PL_0 for prediction. The remaining POC indexes are referred to as leaf frames with predictor level-3 (PL_3) and may predict from any of the previous levels, but are not used as prediction references themselves. The same prediction scheme is followed for vertical axis as well, the V_{ID} with index 6 is defined as predictor level-0 (PL_0). The V_{ID} with indexes {2,10} are defined as predictor level-1 (PL_1) and indexes {0,4,8,12} are chosen as predictor level-2 (PL_2). The remaining V_{ID} indexes are marked as leaf frames with predictor level-3 (PL_3).

The POC and V_{ID} categorization into predictor levels, results into 16 possible combinations that a frame can have as shown in table 1. The proposed prediction scheme provides two main advantages. The leaf frames can handle the occlusion problem more effectively since immediate neighbor frames are available for prediction. Secondly, the rate allocation takes in to account the predictor level while assigning the quality to each frame and give better quality to frames placed in higher predictor level. The weights for each prediction level are estimated after empirical testing and are shown in table 1.

4.2. Rate Allocation

The POC = 0 and V_{ID} = 6 is defined as base frame and a base Quantization Parameter (QP) is assigned to this frame,

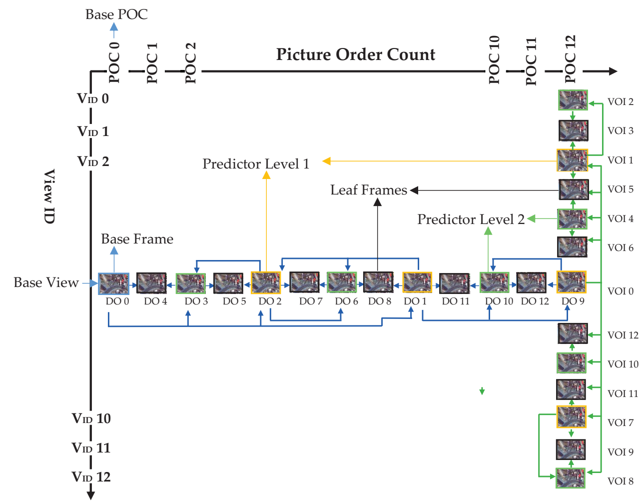

Fig. 2: Proposed prediction and rate allocation scheme. Each row represents the views of a multi-view sequence, each column represents the frames within each view.

referred to as QP_B. All other frames are assigned a specific QP (QP_F), which is set relative to QP_B. The relative quantization offset is estimated by considering each frame's V_{ID}, POC, View Order Index (VOI), and its prediction level. The frames included in base view or in base POC as shown in Fig.2 are given quantization offset equal to maximum of their predictor level ($PL_{Max} = max[PL(POC), PL(V_{ID})]$). For example, the frame with $POC = 0$ and $V_{ID} = 10$ is assigned quantization offset of 1 since POC index 0 is defined in PL_0 and V_{ID} index 10 is defined in PL_1 . The QP (QP_F) for remaining frames is calcu1ated by estimating their distance from the base frame and their view-wise decoding order relative to base view. Equations (1) and (2) defines this assignment explicitly.

$$F_D = \lfloor \frac{|POC_F - POC_B|}{W} \rfloor + \lfloor \frac{|V_{ID_F} - V_{ID_B}|}{W} \rfloor \qquad (1)$$

$$V_D = \begin{cases} VOI_F, & VOI_F \leq V_{ID_B} \\ VOI_F - V_{ID_B}, & VOI_F > V_{ID_B} \end{cases} \qquad (2)$$

$$QP_F = \begin{cases} QP_B + PL_{Max}, \text{ if (POC=0 or } V_{ID}\text{= 6)} \\ QP_B + F_D + V_D, \text{ otherwise} \end{cases} \qquad (3)$$

The variable POC_F represents the POC of the current frame, POC_B represents the POC of base frame. Similarly, V_{ID_F} and V_{ID_B} indicates the view ID of current and base frame. The parameter W is used to assign weights to each frame based on its utilization as a predictor. The variable VOI_F represents the view order index of each frame.

The equation (1) calculates the POC and V_{ID} difference between current and the base frame. The parameter W ensures that high level predictor receive better quality as compared to low level predictors. Hence, the variable F_D stores

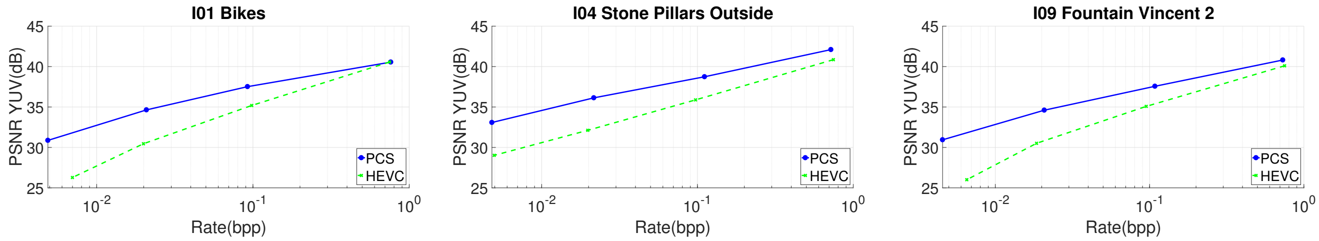

Fig. 3: The Rate Distortion analysis of proposed compression scheme (PCS) is performed with HEVC anchor by following the ICIP 2017 Grand challenge experimental Setup.

the predictor level aware distance between current and base frame. The equation (2) estimates the view wise decoding distance between base frame and current frame. In this way, better quality is assigned to views encoded first since they are utilized as predictor for future views. The equation (3) summarizes the proposed QP selection process for each frame. The base view and base POC are assigned quanization offset equal to their predictor level. The remaining frames are assigned quantization offset by considering their prediction level, distance and view wise decoding order relative to base frame.

5. TEST ARRANGEMENT AND EVALUATION CRITERIA

A subset of 5 images from EPFL database, provided in the Grand Challenge are encoded on four specified bit rates corresponding to compression ratios (R1 = 0.75, R2 = 0.1, R3 = 0.02, and R4 = 0.005 bits per pixel) [18]. The MV-HEVC reference software HTM-16.2 is used for implementing the proposed compression scheme [20]. An analysis of the first and last column of each sub-aperture image reveals that half of the pixels in these columns contain a zero value. In our proposed scheme, these zero-values are discarded and the first and last columns of each sub-aperture image are multiplexed together as a single column, thereby reducing the width of the sub-aperture image from 625 to 624 pixels. This enables the width of the sub-aperture images to be exactly divisible by minimum CU size (8 pixels) of the MV-HEVC encoder and hence does not require additional image padding. Following the guidelines of competition, the decoded plenoptic image is converted to the reference light field structure and mean Peak Signal-to-Noise Ratio (PSNR) is calculated by using the script provided in the competition.

6. RESULTS AND ANALYSIS

The Fig. 3 shows the Rate Distortion (RD) analysis of proposed scheme with reference HEVC anchors. In all the test cases, the proposed scheme completely outperforms the provided anchors. The Table 2 reports the gain in PSNR YUV (mean) with respect to reference HEVC anchors. The base quantization parameter for each image is also explained in the table to obtain desired compression ratio. The proposed scheme shows an average improvements of 2.4 dB over reference JPEG anchors. The gain is minimum for image 10 with PSNR improvement of over 0.7 dB and it is maximum for image 4 with PSNR improvement of 2.95 which reflects the scene dependence on compression scheme. The proposed scheme also provide better compression efficiency in comparison to the work presented in ICME 2016, Grand challenge for plenoptic image coding [9–13] and the comparison results are available online [19].

Table 2: Comparison of proposed scheme with HEVC anchor. Corresponding table for JPEG anchor presented online [19]

Image ID	**Base QP** (QP_B)				**BD-PSNR (dB)**
	R1	R2	R3	R4	YUV$_{mean}$
1	10	21	29	39	2.74
2	12	23	31	40	2.77
4	10	19	26	35	2.95
9	11	21	29	39	2.77
10	8	18	24	33	0.71
Average					**2.39**

7. CONCLUSION

In this paper we have used sub-aperture based representation of plenoptic image as a multiview sequence and have proposed two dimensional prediction and rate allocation scheme. The encoding is performed by utilizing multiview extension of HEVC. In plenoptic camera, perspective change between neighbouring sub-aperture images is very small that enables the multiview extension of HEVC to achieve efficient compression. The PSNR improvements for the proposed method can reach over 2.4 dB compared with HEVC anchors provided in the ICIP 2017 Grand challenge on plenoptic image compression.

8. REFERENCES

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

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

[3] Gabriel Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[4] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[5] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal processing magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[6] Roger Olsson, Marten Sjostrom, and Youzhi Xu, "A combined pre-processing and h. 264-compression scheme for 3d integral images," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 513–516.

[7] Gary J Sullivan, Jens Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[8] Irene Viola, Martin Řeřábek, Tim Bruylants, Peter Schelkens, Fernando Pereira, and Touradj Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Picture Coding Symposium (PCS), 2016*. IEEE, 2016, pp. 1–5.

[9] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng, "Pseudo-sequence-based light field image compression," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

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

[11] Ricardo Monteiro, Luís Lucas, Caroline Conti, Paulo Nunes, Nuno Rodrigues, Sérgio Faria, Carla Pagliari, Eduardo da Silva, and Luís Soares, "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*. IEEE, 2016, pp. 1–4.

[12] Caroline Conti, Paulo Nunes, and Luís Ducla Soares, "Hevc-based light field image coding with bi-predicted self-similarity compensation," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[13] C Perra and P Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–4.

[14] Li Zhang, Ying Chen, Xiang Li, and Marta Karczewicz, "Advanced inter-view residual prediction in multiview or 3-dimensional video coding," Sept. 17 2013, US Patent App. 14/029,696.

[15] Gerhard Tech, Ying Chen, Karsten Müller, Jens-Rainer Ohm, Anthony Vetro, and Ye-Kui Wang, "Overview of the multiview and 3d extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, 2016.

[16] Donald G Dansereau, "Light field toolbox for matlab," *software manual*, 2015.

[17] Martin Rerabek, Tim Bruylants, Touradj Ebrahimi, Fernando Pereira, and Peter Schelkens, "Icme 2016 grand challenge: Light-field image compression," *Call for proposals and evaluation procedure*, 2016.

[18] Call for Proposals on Light Field Coding, "Jpeg pleno," *ISO/IEC JTC 1/SC29/WG1N74014, 74th Meeting, Geneva, Switzerland*, January 15-20, 2017.

[19] ICIP-2017, *Results for input format and target bit rate set out by ICIP 2017, [Online]*, Available:http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-30740 (Accessed: 2017-05-17).

[20] JCT-3V, *MV-HEVC and 3D-HEVC Reference Software, [Online]*, Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-16.2/, (Accessed: 2017-05-07).