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 on Light Field Image Coding and the compression results are put in contrast to the state-of-art in plenoptic image compression presented at of the ICME 2016 Grand Challenge on Light-field Image Compression. The proposed scheme outperforms all contestants in the ICME Grand Challenge with a significant improvements in compression efficiency, i.e. with an average PSNR gain of 7.7 dB over reference JPEG image compression.

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 dataset this function defines 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 effectse. 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 were each lenslet projection records angular information through a specific spatial point, as first reported by Gabrial Lippman 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 decorrelate the data by exploiting redundancies in the image by e.g. employing multi-resolution and prediction models. The JPEG2000 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 does not efficiently compress plenoptic images due to their correlation properties, and could be more 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 organized 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. This compression approach is referred as to as pseudo sequence encoder (PSE) and it gives up to 4.5 dB improvement as compared to the reference JPEG compression scheme. 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. Their proposed encoding scheme is referred to as an image B-coder (IBC) and up to 5.5 dB gain is reported over reference JPEG. 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. The results shows 4.8 dB gain over reference JPEG. Conti et al. have also utilized the SS-based scheme for compressing plenoptic images and show up to 5.1 dB gain over reference JPEG [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. This approach shows less compression gains compared to the previously described schemes.

An rate-distortion analysis of the above schemes supports the conclusion that the psuedo sequence based compression approach is efficient, especially in low bitrate scenarios [8]. The other presented schemes show significant compression
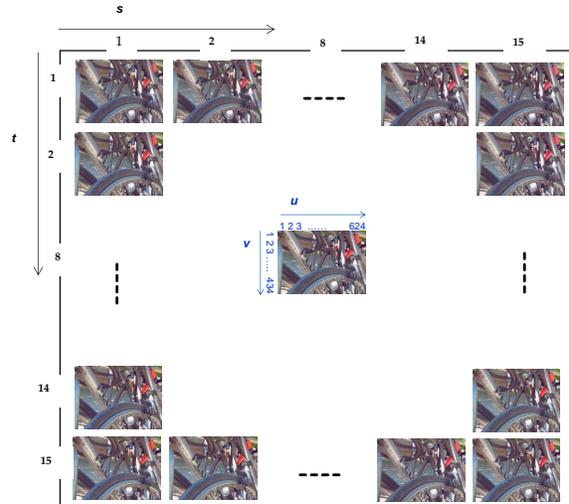

Fig. 1: A plenoptic image represented by a set of 15-by-15 sub-aperture images. s and t represent the horizontal and vertical position, u and v are the pixel coordinates in sub-aperture image plane.

efficiency in high bit rates. Important to note that the pseudo sequence based plenoptic image compression scheme has not utilized the tools available in multi-view extension of HEVC [14]. On the other hand the image B-coder, the LLE and SS based compression schemes doesn't take into account the full spatial correlation structure in the plenoptic image. since block prediction is restricted to a search window containing previously compressed neighbouring blocks, which reduce the compression efficiency possible to achieve. Similarly, when encoding a pseudo video sequence with frame from single layers, only a portion of the angular information of the scene is considered for redundancy reduction.

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.

3. PLENOPTIC IMAGE REPRESENTATION

The Matlab Lytro toolbox is used to convert the plenoptic image (YUV420 8-bits per pixel, 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. The

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

Levels	Level 0	Level 1	Level 2	Leaf frame
Level 0	QP_B	3	3	3
Level 1	3	3	2	2
Level 2	3	2	3	3
Leaf frame	3	2	3	1.5

same plenoptic image representation is used in that adopted for ICME 2016 Grand Challenge to facilitate comparisons with state-of-the-art [17]. However, , the proposed compression scheme is equally applicable for the input image format and compression ratios defined by the ICIP 2017 Grand Challenge and explicit results for those setting are available online at [18].

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 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 for how they will be compressed as frames using MV-HEVC. The two parameters View ID (V_{ID}) and and Picture Order Count (POC) together uniquely identifies a frame within the set. A single frame is defined as a base frame ($V_{ID} = 0$ and $POC = 0$) from which all predictions stem. The other frames are defined as predictor frames and divided into three categories. The frames with $V_{ID} = \{2, 10\}$, and $POC = \{4, 12\}$ are defined as level 1 predictors mainly use the base frame for prediction. The level 1 predictors are encoded immediately after the base frame. Similarly, the frames with $V_{ID} = \{0, 4, 8, 12\}$ and $POC = \{2, 6, 10\}$ are defined as level 2 predictors and use level 1 predictors or the base frame for prediction. The remaining frames are referred to as leaf frames, may predict from any of the previous levels including the base frame, but are not used as prediction references themselves. The proposed prediction scheme provides two main advantages. The leaf frames are most efficiently encoded since the neighbor
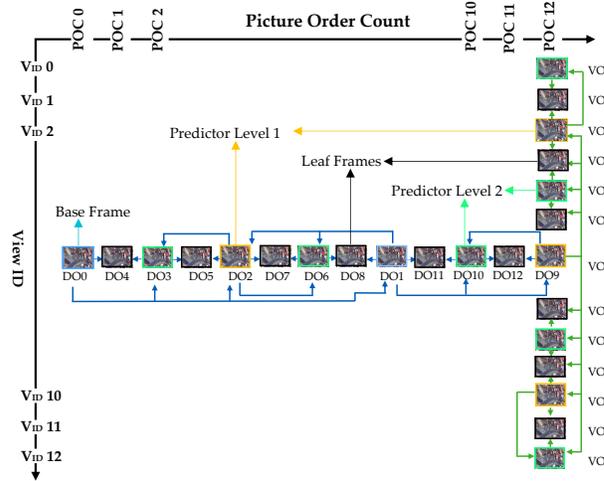

Fig. 2: Proposed prediction and rate allocation scheme. Each row represents the views of a multi-view sequence, each column represents the frames within each view.

frames are already encoded . Secondly, the rate allocation takes in to account the predictor level while assigning the quality to each frame. In this way better quality is distributed among different frames with in 13x13 sub-aperture images. The weight's for each prediction level are estimated after empirical testing and are shown in table 1.

4.2. Rate Allocation

A base Quantization Parameter (QP) is assigned to the first encoded frame as (QP_B). All other frames are assigned a specific QP (QP_F), which is set relative to QP_B and considering each frame's V_{ID}, POC, and View Order Index (VOI), and its prediction level. Equations (1) and (2) defines this assignment explicitly.

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

$$QP_F = QP_B + F_D + (VOI_F \mod V_{ID_B}) \quad (2)$$

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 equation (1) calculates the POC and V_{ID} difference between current and the base frame and it estimates the offset F_D for providing rate allocation based on distance between current and base frame. The equation (2) estimates the final QP (QP_F) for each frame by adding QP of base frame, offset F_D and current frame view order index with respect to base frame. In this way, each frame is assigned a QP based on its
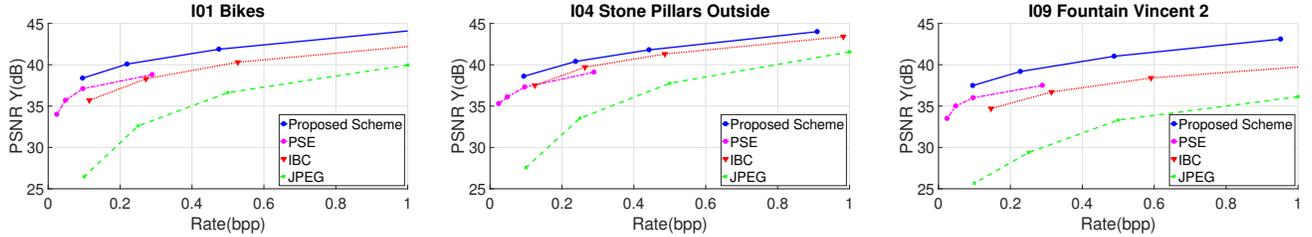
Fig. 3: The Rate Distortion analysis of proposed scheme with JPEG, Pseudo sequence encoder (PSE) and Image B-coder (IBC).

distance from the base frame, its predictor level, and its view decoding order with respect to the base view.

5. TEST ARRANGEMENT AND EVALUATION CRITERIA

In this paper the proposed compression scheme is compared with JPEG anchors, Pseudo Sequence based encoder and Image B-coder published in the ICME Grand challenge [17]. A subset of 6 images from EPFL database, provided in the Grand Challenge are encoded on four specified bit rates corresponding to compression ratios (R1 = 1, R2 = 0.5, R3 = 0.25, and R4 = 0.1 bits per pixel) [19]. 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 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 JPEG anchors, Pseudo sequence based encoder and Image B-coder [9, 10]. In all the test cases, the proposed scheme completely outperforms the other schemes in the low, medium and high bit-rate scenarios. The Table 2 reports the gain in PSNR (mean) with respect to reference JPEG 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 7.7 dB over reference JPEG anchors. The gain is minimum for image 4 with PSNR improvement of over 6.4

Table 2: Comparison of proposed scheme with reference JPEG anchors

Image ID	Base QP				BD-PSNR (dB)	
	R1	R2	R3	R4	Y	YUV
1	8	13	17	21	7.9	7.3
2	12	15	19	23	8.0	7.7
3	10	15	18	23	7.2	7.0
4	9	13	16	20	6.4	6.5
9	10	14	18	22	10.6	9.3
10	7	12	15	19	8.1	7.5
Average					**8.0**	**7.5**

dB and it is maximum for image 9 with PSNR improvement of 10 which reflects the scene dependence on compression scheme. The proposed scheme provide better compression efficiency in comparison to the work presented in ICME 2016, Grand challenge for plenoptic image coding as discussed in section 2.

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 comparison is made with state-of-art and proposed scheme outperforms both Image B-coder and pseudo sequence based plenoptic image encoding schemes. The PSNR improvements for the proposed method can reach over 7.7 dB compared with reference JPEG.

8. REFERENCES

[1] Edward H Adelson and James R Bergen, "The plenoptic function and the elements of early vision," 1991.

[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] 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).

[19] Martin Rerabek and Touradj Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, number EPFL-CONF-218363.

[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).