TOWARDS A GENERIC COMPRESSION SOLUTION FOR DENSELY AND SPARSELY SAMPLED LIGHT FIELD DATA

Waqas Ahmad, Roger Olsson, Mårten Sjöström

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

ABSTRACT

Light field (LF) acquisition technologies capture the spatial and angular information present in scenes. The angular information paves the way for various post-processing applications such as scene reconstruction, refocusing, and synthetic aperture. The light field is usually captured by a single plenoptic camera or by multiple traditional cameras. The former captures a dense LF, while the latter captures a sparse LF. This paper presents a generic compression scheme that efficiently compresses both densely and sparsely sampled LFs. A plenoptic image is converted into sub-aperture images, and each sub-aperture image is interpreted as a frame of a multi-view sequence. In comparison, each view of the multi-camera system is treated as a frame of a multi-view sequence. The multi-view extension of high efficiency video coding (MV-HEVC) is used to encode the pseudo multi-view sequence. This paper proposes an adaptive prediction and rate allocation scheme that efficiently compresses LF data irrespective of the acquisition technology used.

Index Terms— Light field, plenoptic, Multi-camera, MV-HEVC

1. INTRODUCTION

The capturing of the spatial and angular information in a scene enables various post-processing applications such as reconstructing a three-dimensional scene model, refocusing at different depth planes, changing the depth of field, etc. Light field acquisition technologies are used to capture the spatial and angular information of light rays in a scene. A light field can be captured by a single plenoptic camera or by multiple traditional cameras. The former captures a densely sampled LF, while the latter records a sparse LF. Ren Ng at Lytro [1] used the proposal of Lippmann [2] to develop the first commercial plenoptic camera in 2006. Based on the optical design, which places a lenslet array in front of the sensor, this camera multiplexes the spatial and angular information of a scene onto a single image. In multi-camera-based LF acquisition, each camera captures the scene from a slightly different perspective. Additional angular information is acquired at the expense of increased total information size. Traditional image encoders do not exploit the correlation present in LF data, since such encoders are built on assumptions drawn from natural images. JPEGs recent attempts to develop a new compression standard (JPEG Pleno) reflect the importance of a compression scheme that is capable of addressing LF properties [3].

This paper presents the design of a generic compression system for LF data. The captured LF is converted into multi-view pseudo-video sequences and provided as input to MV-HEVC encoder. The prediction scheme categorizes the frames into different prediction levels based on their utilization as predictors for other frames. Compression efficiency is improved by estimating the quantization parameter for each frame based on its assigned prediction level. This paper is novel in that it introduces a prediction model that adapts according to the properties of the incoming LF; in addition, it defines the central frame as the base frame and estimates the quantization parameter for each frame by considering its distance, prediction level, and decoding order with respect to the base frame. The paper is organized as follows: state-of-the-art compression schemes are explained in Section 2, the proposed compression system is described in Section 3, test and evaluation criteria are discussed in Section 4, the results are discussed in Section 5, and Section 6 serves as the conclusion.

2. STATE-OF-THE-ART ALGORITHMS FOR LIGHT FIELD COMPRESSION

In the recent past, various studies have reported on the compression of LF data. These compression schemes can be divided into two major groups based on the two different LF acquisition technologies: the plenoptic camera and the multi-camera system. The plenoptic images captured by the Lytro camera have received significant attention from the

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.

research community due to this cameras availability in the consumer market and the fact that Lytro datasets are used in various competitions [3, 4]. In compression solutions for Lytro images, two approaches are noticeable: either novel representation of the plenoptic image is used, or novel tools are incorporated in compression standards. In the first approach, the plenoptic image is converted into sub-aperture images, and standard video compression tools are used, as proposed by Olsson et al., for the compression of integral imaging as a pseudo-video sequence using H.264 [5]. Recently, Liu et al. converted the plenoptic image into sub-aperture images that were input as frames into the HEVC [6]. Ahmad et al. extended this concept by treating sub-aperture images as frames of multi-view sequences and encoding them using MV-HEVC [7]. In the second approach, standard encoders are updated with novel tools to make better use of the correlation present in plenoptic images. Li et al. introduced a bi-prediction capability within the HEVC intra-prediction framework by taking references from already encoded blocks. For each block, 33 intra-modes and an additional bi-prediction mode are evaluated, and the best mode is selected [8]. Following the same principle, Monteiro et al. introduced two novel tools to the HEVC image compression framework. The prediction of the current block is performed based on local linear embedding (LLE) and self-similarity (SS) operators [9]. The current block is estimated as a linear combination of already encoded k-nearest blocks. The rate distortion optimization selects the best candidate mode. Similarly, Conti et al. also proposed a SS-based prediction scheme [10].

In the compression of a LF captured by multiple cameras, Hawary et al. utilized the sparseness present in the angular Fourier transform of the LF [11]. A subset of views are encoded in the base layer of the HEVC, and a corresponding decoded version of the subset was used to predict the intermediate views. Xian et al. has proposed a homography based LF compression method [12]. The low-rank approximation was applied jointly with the alignment of views by finding the homographies that reduce the low-rank error.

Significant compression improvements have been reported in plenoptic image compression between interpretation of the plenoptic image as a single image and as a set of sub-aperture images. In this manner, standard video compression schemes are modified to obtain efficient plenoptic image compression. Similarly, a generic compression scheme that can efficiently compress LF captured with plenoptic cameras and multi-camera systems could further streamline LF compression efforts.

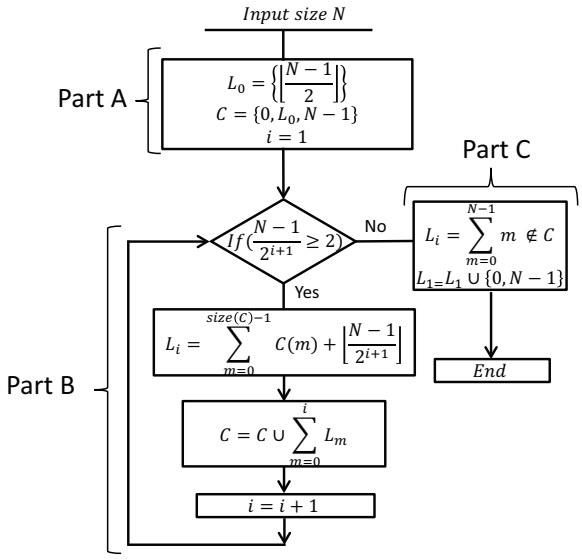

Fig. 1: Prediction level assignment is shown as a flow diagram for a single view with N Frames.

3. PROPOSED METHOD

3.1. Light Field Representation

Plenoptic Camera: The raw LF image captured with the plenoptic camera was first de-mosaiced and de-vignetted using the Matlab toolbox [13]. Thereafter, the preprocessed plenoptic image was converted into a set of 13x13 sub-aperture images, each of which has a resolution of 624x432. Border views were discarded due to vignetting noise. Each sub-aperture image depicts the scene from a slightly different perspective and can therefore be thought of as a single camera view. Next, the sub-aperture images were interpreted as frames of multiple pseudo-video sequences. As such, a set of 13x13 sub-aperture images were treated as 13 pseudo-video sequences, with each sequence having 13 frames.

Multi-Camera System: The LF captured with multiple cameras inherently captures each perspective view in a separate image. Each perspective view was interpreted as a frame of a pseudo-video sequence, and the complete grid of MxN views was treated as a multiple pseudo-video sequence. The proposed LF representation allowed the MV-HEVC [14] encoder to achieve better compression efficiency.

3.2. Prediction Scheme

An adaptive prediction scheme was introduced to compress both densely and sparsely sampled LF data. The prediction scheme categorized the frames into multiple prediction levels (L), and the rate allocation scheme used this information when assigning quality to each frame. The best available quality was assigned to the first prediction level, and, as the prediction level increased, slightly lower quality was as-

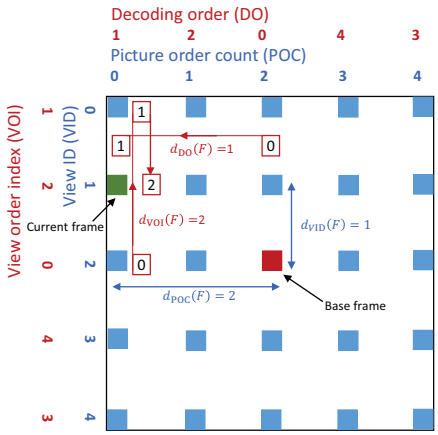

Fig. 2: The rate allocation scheme is shown for a grid of 5x5 views. The base frame (POC=2,VID=2) is shown in red color and current frame (POC=0,VID=1) is shown in green color.

signed at each successive prediction level. The frames placed in higher prediction levels are allowed to take prediction from the frames placed in lower prediction levels; however, the frames placed in the last prediction level were not used for the prediction of any other frames, since they were assigned the lowest quality. In this way, better-quality frames were used to predict other frames in order to increase overall compression efficiency.

In Figure 1, the prediction level estimation process is depicted in the form of a flow diagram. For simplicity, the process is shown for a single view ($M = 1$) with N frames, ranging from 0 to $N-1$. In part A, the position of the central frame, also referred to as the base frame, was estimated and assigned with prediction level L_0. Moreover, the first and last frames of the input sequence and the central frame were updated in the candidate list (C). In part B, it is estimated that weather another partition is possible. In the case of a true outcome, frames were assigned with the next prediction level, and the next iterations were performed. In the case of a false outcome (part C) when further partition is not possible, the remaining frames were labeled with the final prediction level. In the final step, the prediction level L_1 was updated with border frames. In this way, all the frames were assigned with a prediction level.

3.3. Rate Allocation

The rate allocation scheme used MV-HEVC parameters picture order count (POC), view identification (VID), decoding order (DO), and view order index (VOI) in order to estimate the quantization parameter for each frame. The horizontal and vertical axes of the frame position are identified by POC and VID, respectively. Decoding order and VOI represent the decoding order of each frame's horizontal and vertical axis, as shown in Figure 2. Initially, the base frame was assigned the base quantization parameter (Q_{B}). The remaining frames added the quantization offset (Q_{o}) to the Q_{B} to obtain the quantization parameter Q. Equations (1) through (6) explain the rate allocation methodology. In Equation (1), the frame distance between the current frame and the base frame in the POC axis is estimated and represented by d_{POC}. Similarly, in Equation (2), the frame distance in the VID axis is estimated and represented by d_{VID}:

$$d_{\mathrm{POC}}(x) = n_{\mathrm{POC}}(x) - B_{\mathrm{POC}} \tag{1}$$

$$d_{\mathrm{VID}}(y) = v_{\mathrm{VID}}(y) - B_{\mathrm{VID}} \tag{2}$$

Here, POC is denoted by n_{POC}, VID is denoted by v_{ID}, B_{POC} and B_{VID} represents the position of base frame, x indexes the horizontal axis (from 0 to $N-1$) and y indexes the vertical axis (from 0 to $M-1$).

The decoding distances in POC and VID axis between the current frame and the base frame are estimated in equation (3) and (4) and are represented by d_{DO} and d_{VOI} respectively.

$$d_{\mathrm{DO}}(x) = \begin{cases} k_{\mathrm{DO}}, & n_{\mathrm{POC}}(x) \le B_{\mathrm{POC}} \\ k_{\mathrm{DO}}(x) - B_{\mathrm{POC}}, & n_{\mathrm{POC}}(x) > B_{\mathrm{POC}} \end{cases} \tag{3}$$

$$d_{\mathrm{VOI}}(y) = \begin{cases} i_{\mathrm{VOI}}(y), & v_{\mathrm{VID}}(y) \le B_{\mathrm{VID}} \\ i_{\mathrm{VOI}}(y) - B_{\mathrm{VID}}, & v_{\mathrm{VID}}(y) > B_{\mathrm{VID}} \end{cases} \tag{4}$$

Here, k_{DO} and i_{VOI} represent DO and VOI.

Equation (5) calculates the required quantization offset (Q_{o}) for each frame.

$$\begin{aligned} Q_{\mathrm{o}}(x,y) = \lfloor \frac{|d_{\mathrm{POC}}(x)|}{W} \rfloor + \lfloor \frac{|d_{\mathrm{VID}}(y)|}{W} \rfloor + \lfloor \frac{|d_{\mathrm{DO}}(x)|}{W} \rfloor + \\ = \lfloor \frac{|d_{\mathrm{VOI}}(y)|}{W} \rfloor \end{aligned} \tag{5}$$

Here, W represents the weightage parameter that controls the extent of Q_{o} by normalizing the frame and decoding distances. The same values for W are used as described in [7]. Frames with lower prediction levels were assigned large weightage values, and frames with higher prediction levels were assigned smaller weightage values. Finally, Equation (6) estimates the quantization parameter (Q) for each frame.

$$Q(x,y) = \begin{cases} Q_{\mathrm{B}} + L_{\mathrm{Max}}(x,y), & \text{if } x = B_{\mathrm{POC}} \text{ or } y = B_{\mathrm{VID}} \\ Q_{\mathrm{B}} + Q_{\mathrm{o}}(x,y), & \text{otherwise} \end{cases} \tag{6}$$

Here, Q_{B} represents the base quantization parameter, and L_{Max} represents the maximum prediction level of each frame in horizontal and vertical axes. The frames placed in base POC or in base VID were assigned quantization offsets equal to their maximum prediction level ($L_{\mathrm{MAX}}(x,y) = \max(s_{\mathrm{POC}}(x), t_{\mathrm{VID}}(y))$). The prediction level in the POC axis is represented by s_{POC}, and the prediction level in the VID axis is represented by t_{VID}. The remaining frames estimate the quantization offset, as explained in (5).

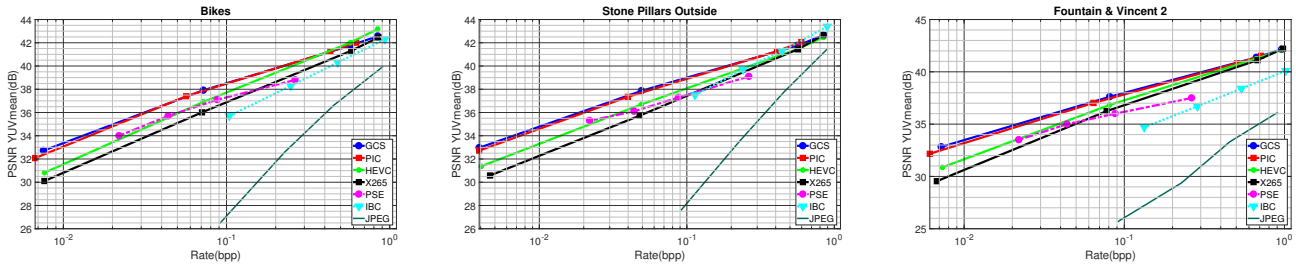

Fig. 3: *Plenoptic Camera*: The Rate Distortion analysis of Generic Compression Scheme (GCS) is performed with the state-of-art compression methods; i.e., plenoptic image coder (PIC) [7], Pseudo sequence encoder (PSE) [6], Image B-coder (IBC) [8] and with two anchor schemes HEVC And X265.

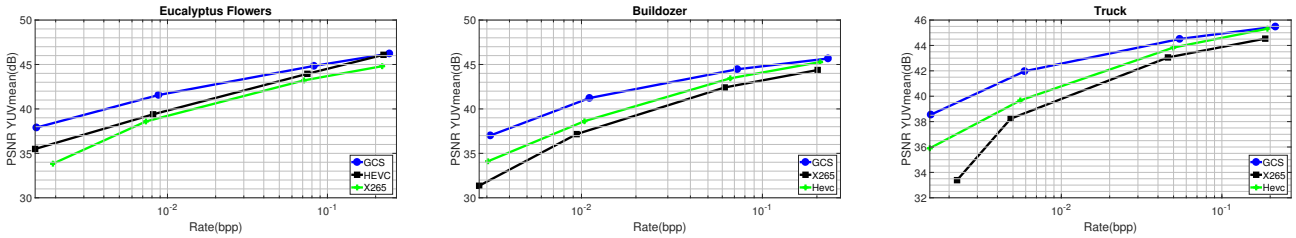

Fig. 4: *Multi-Camera*: The Rate Distortion analysis of Generic Compression Scheme (GCS) is performed with HEVC and X265 anchor schemes.

4. TEST ARRANGEMENT AND EVALUATION CRITERIA

The proposed compression scheme was evaluated on LF images captured with a Lytro camera [15] and a multi-camera system [16]. The input LF images were converted into the YUV 420 format with 8 bit per pixel(bpp) depth. Two benchmark anchors, HEVC [17] and its real-time implementation X265 [18], were used to evaluate the proposed compression scheme. The sub-aperture images decomposed from the plenoptic image and the views captured by the multi-camera system were converted into a single pseudo-video sequence by following the method explained in [3] and input into both anchor schemes. The rate-distortion (R-D) curve was plotted on four different bit-rates in order to cover the range from low bit-rates to high bit-rates. The BD-PSNR [19] performance measure was used to show the compression efficiency of the proposed compression scheme relative to anchor schemes.

5. RESULTS AND ANALYSIS

The proposed compression scheme was evaluated on a subset of LF images taken from New dataset [15]. As presented in Figure 3, the R-D analysis demonstrates that the proposed scheme performed better overall than the state-of-the-art schemes, mainly because the proposed compression scheme allows the input LF data to seek predictions in both dimensions. Moreover, comparison between the proposed scheme and the previously developed plenoptic image compression scheme [7] demonstrates that the selection of a suitable base frame and modification in the rate allocation scheme improves overall compression efficiency. An average BD-PSNR increase of 1.59 dB and 0.91 dB was obtained when the proposed scheme was compared with pseudo-sequence based X265 and HEVC anchors, respectively. The performance of the proposed scheme was significantly better in low bit-rates than in high bit-rates. The sensitivity of the human vision system to the compression artifacts in low bit-rates [20] favors the proposed compression scheme over the anchor schemes. The proposed compression scheme was also applied to a subset of LF images chosen from the Stanford dataset [16]. Again, the R-D analysis in Figure 4 demonstrates higher compression efficiency in the proposed scheme than the anchor schemes. An average BD-PSNR gain of 2.36 dB and 1.51 dB was obtained when the proposed scheme was compared to pseudo-sequence based X265 and HEVC anchors, respectively.

6. CONCLUSION

This paper presented a generic compression scheme for densely and sparsely sampled LFs. Both the sub-aperture image of a plenoptic camera and the single view captured by a multi-camera system were interpreted as a frame of a multi-view sequence. The multi-view extension of high efficiency video coding was used to encode the pseudo multi-view sequence. A modification in the previously proposed prediction and rate allocation scheme was performed to make the compression scheme generic. The proposed compression scheme efficiently compresses LF data irrespective of acquisition technology.

7. REFERENCES

[1] R. Ng, M. Levoy, B. Mathieu, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

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

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

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

[5] R. Olsson, M. Sjöström, and Y. 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.

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

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

[8] Y. Li, R. Olsson, and M. 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.

[9] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. 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.

[10] C. Conti, P. Nunes, and L.D. 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.

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

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

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

[14] HTM, "MV-HEVC reference software, [online]," `https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-16.2/`, Accessed = 2018-02-01.

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

[16] V. Vaish and A. Adams, "The new stanford light field archive, [online]," `http://lightfield.stanford.edu/lfs.html`, Accessed = 2018-02-01.

[17] HM, "HEVC reference software, [online]," `https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9/`, Accessed = 2018-02-01.

[18] Multicoreware, "X265 hevc encoder, [online]," `https://bitbucket.org/multicoreware/x265/`, Accessed = 2018-02-01.

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

[20] D. Lin and P. Chau, "Objective human visual system based video quality assessment metric for low bit-rate video communication systems," in *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006, pp. 320–323.