OPTIMIZED PREDICTIVE CODING OF 5D LIGHT FIELDS

Harini Priyadarshini Hariharan, *StMIEEE* and Thorsten Herfet, *SMIEEE*
Saarland Informatics Campus
Saarland University, Telecommunications Lab
Saarbrücken, Germany
hariharan@nt.uni-saarland.de, herfet@cs.uni-saarland.de

ABSTRACT

With the emergence of Light Field (LF) technology, the number of dimensions representing light has once again increased. $4D$ light fields captured with additional temporal information per ray or as assemblies of rays include the 5th dimension, namely time and thus produce $5D$ light fields. This is very crucial when we have moving objects in the scene. In the recent years, research has paved way to several ideas on efficient $4D$ light field compression. However, techniques for compression and storage for higher dimensions is still an open challenge. In this paper we have introduced a low-complexity predictive coding of $5D$ light fields by automatic generation of per frame customized coding structure exploiting both spatial and temporal neighbors. Evaluations with HEVC codec shows an increase of more than $1.4\ dB$ gain in quality.

***Index Terms*— 3.2 Video Coding and Processing; 3.7 3*D* and Multiview Video; 3.12 Scalable video coding & Content adaptation**

1. INTRODUCTION

The conceptual foreground on light fields has been evolving since 1996 [1]. Two widely acknowledged capturing techniques are lenslet based cameras [2] and camera arrays [3]. With lenslet based cameras being restricted to narrow baselines, low resolution and lesser possibilities of capturing the temporal information, camera arrays are gaining precedence. Light field videos become extremely important when the scene is not static but has moving components. Also, with captured temporal information, the number of options for post-processing increases manifold. Consumers are constantly attracted to more and more effortless immersive experience which increases the expectations on the lower end mobile manufactures to the media and production industries. Since acquiring Lytro, Google has also been actively researching on capturing, processing, compression and rendering of light fields [4].

Figure 1: Light Field array [3]

Light field representation of higher dimensions includes information beyond intensity and direction. Light fields are commonly represented as $4DLF = (t, s, v, u, c)$ where, st is the angular domain, uv the spatial domain and c the color information. Capturing light field videos comprises temporal information and to avoid ambiguity with respect to the time we have decided to use $\alpha\beta$ for angular, xy for spatial and t for temporal resolution respectively. The overall representation of $5DLF = (t, \beta, \alpha, y, x, c)$.

In the last decade, the state-of-the-art image and video compression committees have been working on standardizing light field compression. MPEG-I[1] is a collection of standards to digitally represent immersive media and the committee is engaged on both multi-dimensional audio and video representations. JPEG Pleno[2] standardizes all plenoptic modalities such as light fields [5], point clouds and holographic representations. On the other hand, several research ideas have been put forward on $4D$ light field compression [6]. Pre-processing of light field data before coding using H.264 [7], using HEVC [8], multiview coding structure [9] and by pseudo-temporal reordering [10] have been well evaluated. Convolutional neural networks based approach [11] and adversarial network based view synthesis [12] for light field compression are the recent advances.

[1] https://mpeg.chiariglione.org/standards/mpeg-i
[2] https://jpeg.org/jpegpleno/

978-1-7281-2150-5/19/$31.00 ©2020 IEEE

However, with increased dimensions the challenge of efficient compression is still unexplored. Pre-processing the different dimensions cleverly to exploit both spatial and temporal redundancies would be well suited to adapt the light field data for the available state-of-the-art codecs. In this research work we have introduced a technique for automatic generation of predictive coding structure for $5D$ lights fields. The next chapter gives an overview on $5D$ light field data and adaptability. Further, the proposed technique and results are discussed.

2. 5D LIGHT FIELDS

The camera array [3], shown in Figure 1 consists of 8x8 synchronized cameras arranged with constant distances both vertical and horizontal. Images are generated at 40fps with a resolution of $1920x1200$. The rig is electronically controlled and the different cameras can be configured to trigger at varied time instances enabling the temporal behavior. In this paper we have used the HaToy dataset [13] generated using the above mentioned camera array. As stated in [13], we will have a common understanding on the different light field data as follows, light field still images are $4DLF$, while all cameras synchronized light field videos are denoted as $4.5DLF$ and sub-framed light field videos are referred to as $5DLF$.

Figure 2: HaToy dataset [13]

The HaToy scene as shown in Figure 2 incorporates several static and moving components of variable sizes and complex geometry. All the objects in the scene are made visible in all the cameras and they fully capture the static and moving parts of the scene. The dataset includes several spatio-temporal capturing patterns as shown in Figure 3 in addition to the uniform synchronized capturing. These unique sub-framing patterns are derived using two-dimensional bit reversal permutation. In Figure 3, from the highlighted regions, it can be seen that neighboring cameras have different phases and the phases are equidistantly distributed within the layout.

With respect to the $5DLF$ representation, we have $t = 0{:}\,1{:}\,N$, whereby, $0..N-1$ belongs to one full frame and hence the spacing is $1/(N*40)$ seconds (or $25ms/N$). α and β are the camera indices from 0..7, but for intuitive understanding we have the camera numbering, from top left (0) to bottom right (63). Before predicting the HaToy Sub-Aperture Images (SAIs), the most interesting objects to consider are the fast spinning ones like the CD drive and the spin top. From Figure 4, we can observe that only parts of these objects' texture are visible on each camera. For sub-framing by a factor of 4 the center cameras $\#27, 28, 35, 36$ respectively stem from four different sub-frames $\#3, 1, 2, 0$

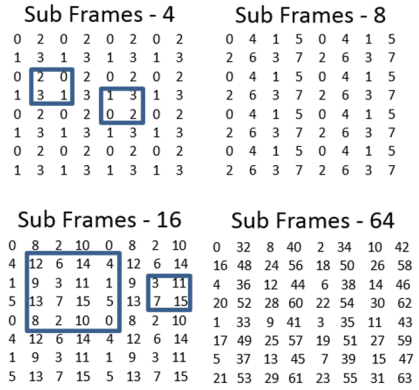

Figure 3: Bit reversal sub-framing

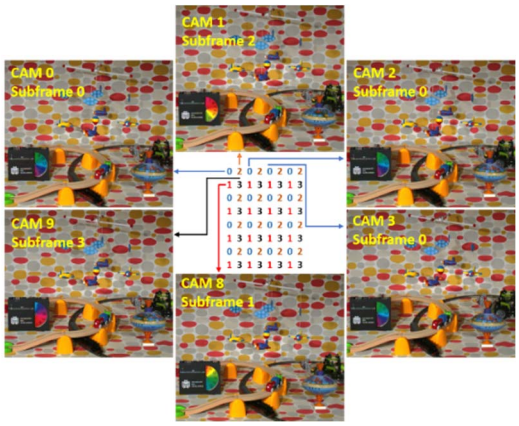

Figure 4: Understanding 5D HaToy dataset

and hence have a high temporal resolution for moving and a high angular resolution for static parts of the scene, while cameras $\#18, 20, 34, 36$ all stem from the same sub-frame $\#0$ and hence are angular neighbors for moving parts of the scene. The temporal behavior is significant and has to be considered while predicting the sub-frames. We have already introduced an efficient prediction technique for $4DLF$ prediction in [10] which can also straight forwardly be applied on $4.5DLF$. While, for the newly available $5D$ sub-framing, the prediction method needs to be redefined.

3. PREDICTIVE CODING AND EVALUATIONS

The proposed idea is an automatic generation of per SAI customized coding structure for HEVC predictive coding. The approach includes calculating the distance for each SAI with its neighbors both temporally and spatially. The absolute difference between the sub-frame numbering is taken as the temporal distance, while the Euclidean metric is used for

deriving the spatial distance. From the different sub-framing patterns as shown in Figure 3 it can be noticed that the spatially adjacent neighbors are already spread temporally equidistant and also from examining the HaToy scene, it is evident that the spatial neighbors have higher correlation as the motion in the scene is limited. Therefore, considering only the spatial distance as represented in Figure 5 for frame 20 or a cost function with a higher weight for the spatial distance would yield similar results.

Figure 5: Proposed prediction structure

Once the distances are obtained, they are sorted for the last 15 SAIs as HEVC standard supports the Decoded Picture Buffer (MaxDpbSize) to a maximum of 16 SAIs, which then can include 15 previously coded SAIs + current SAI. Then, the reference list $L0$ is generated per sub-frame. In the case of sub-framing 4, five previous frames are chosen for reference with the consideration of four spatial sub-frame and one temporal sub-frame neighbor and for sub-framing 8, nine reference frames are chosen respectively as highlighted in Figure 5. The complete coding structure for a Group of Pictures (GOP) of 64 SAIs is then integrated to the configuration file. The streams are predictively coded with the HEVC[3] reference implementation.

The results for some of the sub-framing patterns are discussed. Table 1 illustrates the evaluation of sub-framing by a factor 4 and 8 for different Quantization Parameter (QP) values. We have selected intermediate QP values from the available range $[0 - 51]$ to analyze finer to coarser levels of quantization. Mean YUV-PSNR is calculated for the proposed technique against row-wise default HEVC predictive coding.

Subframes 4	YUV – PSNR [dB]		
	Original	Proposed	Difference
QP 40	33.0531	34.5148	1.4617
QP 36	35.2884	36.7361	1.4477
QP 32	37.4772	38.8394	1.3622
QP 28	40.3468	41.2355	0.8887
QP 24	42.9847	43.5799	0.5952
QP 20	45.1979	45.4335	0.2356
QP 16	46.9441	47.1174	0.1733

Subframes 8	YUV – PSNR [dB]		
	Original	Proposed	Difference
QP 40	33.0516	34.5340	1.4824
QP 36	35.2835	36.7413	1.4578
QP 32	37.5421	38.8545	1.3124
QP 28	40.3313	41.2214	0.8901
QP 24	42.9805	43.5708	0.5903
QP 20	45.2098	45.4316	0.2218
QP 16	46.9453	47.1224	0.1771

Table 1: YUV - PSNR for different sub-frames

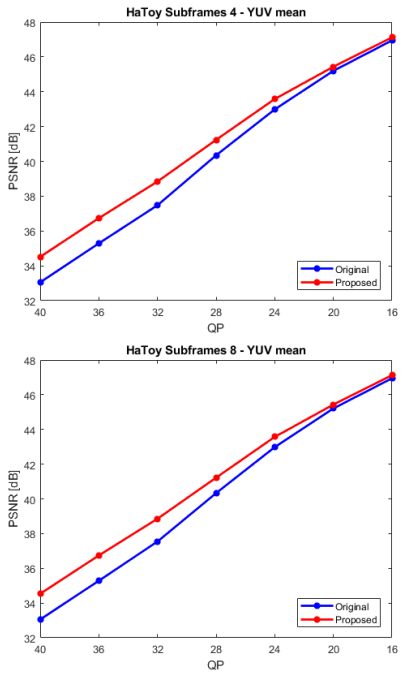

Figure 6: PSNR vs QP curves

For all the compression ratios the proposed coding structure with customized reference list has outperformed traditional line wise sequence prediction. For fairness in the comparison, we have coded our images in the camera order numbering but using the best candidates for prediction. A gain of more than

[3] https://hevc.hhi.fraunhofer.de/

1.4 dB PSNR is obtained for higher QP values. As mentioned in the literature work on Data Compression [14], with just an increase of 0.5 dB PSNR, the magnitude of improvement is already visible to the human eye.

HaToy	BD – PSNR (dB)	BD – BR (%)
Subframes 4	2.2917	-44.7668
Subframes 8	2.1792	-43.0875
Average	2.2354	-43.9272

Table 2: BD-PSNR and BD-BR results

The Bjontegaard metric [15] calculation in Table 2, shows the delta peak signal-to-noise ratio (BD-PSNR) and the delta bitrate (BD-BR) results. The improvement of encoding performance are presented in terms of datarate savings and at the equivalent datarate, the quality improvement. So, from interpreting the BD-BR and BD-PSNR results, it can be observed that our proposed approach, for equivalent quality, provides 43.93% average bitrate savings and for equivalent bandwidth, the average quality improvement is 2.24 dB compared to default encoding.

Additional evaluations on using all the pictures from the DPB as references, only doubles the processing power with very negligible gain in quality and compression ratio. This again validates that selecting the most correlated adjacent temporal and spatial neighbors for prediction will suffice. From the curves illustrated in Figure 6, the effect is that for coarser quantization the differences are more or less quantized towards zero and hence large QPs directly reflect the quality of prediction. By resorting and giving the right reference candidates we directly influence the prediction.

4. CONCLUSION

In this research paper, we have proposed a low-complexity predictive coding approach for 5D light fields by automatic generation of per frame customized coding structure exploiting both spatial and temporal neighbors. A PSNR gain of more than 1.4 dB is achieved with HEVC codec. In our future work, we will implement ideas to enlarge the GOP to include SAIs from consecutive frames and exploit both intra and inter frame coding for light field videos.

ACKNOWLEDGEMENT

This work bases on research conducted in the SAUCE project[4] funded by the EU Horizon 2020 Research and Innovation Programme under Grant Agreement No 78070.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH*, 1996, pp. 31–42.

[2] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Tech. Rep. CTSR 2005-02, Stanford University*, 2005.

[3] T. Herfet, T. Lange and H. P. Hariharan, "Enabling Multiview- and Light Field-Video for Veridical Visual Experiences," in *4th IEEE International Conference on Computer and Communications*, Chengdu, China, 2018.

[4] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr and P. Debevec, "A System for Acquiring Compressing, and Rendering Panoramic Light Field Stills for Virtual Reality," in *SIGGRAPH Asia 2018 Technical Papers ACM*, 2018.

[5] T. Ebrahimi, S. Foessel, F. Pereira and P. Schelkens, " JPEG Pleno: Toward an Efficient Representation of Visual Reality," in *IEEE Multimedia*, Oct-Dec 2016.

[6] M. Magnor and B. Girod, "Data compression for light-field rendering," in *Circuits and Systems for Video Technology, IEEE Transactions* , 2000, vol. 10, no. 3, pp. 338–343.

[7] U. Fecker and A. Kaup, "H.264/AVC-compatible coding of dynamic light fields using transposed picture ordering," in *13th European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005.

[8] A. Vieira, H. Duarte, C. Perra, L. Tavora and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," in *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2015.

[9] D. Lui, L. Wang, L. Li, Z. Xiong, F. Wu and W. Zeng, "Pseudo-sequence-based light field image compression," in *IEEE International Conference on Multimedia and Expo*, Seattle, USA, July,2016.

[10] H. P. Hariharan, T. Lange and T. Herfet, "Low complexity light field compression based on pseudo-temporal circular sequencing," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, Italy, June 2017.

[11] N. Bakir, W. Hamidouche, O. Déforges, K. Samrouth and M. Khalil, "Light Field Image Compression Based on Convolutional Neural Networks and Linear Approximation," in *25th IEEE International Conference on Image Processing (ICIP)*, 2018.

[12] C. Jia, X. Zhang, S. Wang, S. Wang and S. Ma, "Light Field Image Compression Using Generative Adversarial Network-Based View Synthesis," in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2019.

[13] T. Herfet, T. Lange and C. Kelvin, "5D Light Field Video Capture," in *The 16th ACM SIGGRAPH European Conference on Visual Media Production*, BFI Southbank, London, UK, Dec, 2019.

[14] D. Solomon, Data Compression: The Complete Reference, Springer, 1998.

[15] G. Bjøntegaard, "Calculation of Avarage PSNR Difference Between RD-Curves," in *ITU-T, VCEG-M33*, Austin, TX, USA, Apr, 2001.

[4] https://www.sauceproject.eu/