Enhancing Light Field Video Compression Efficiency via View Selection and Synthesis Techniques

Tala Bazzaza, Maissan Bazazeh, Morteza Rashidinia, Weige Qian, Hamid Reza Tohidypour, and Panos Nasiopoulos
Dept. of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
e-mail: {tbazzaza, mbazazeh, mr1997, wgqian, htohidyp, panos}@ece.ubc.ca

***Abstract*— Light Field (LF) imaging represents a transformative technology evolving to replicate human-like visual data and emulate our visual environment. Departing from traditional single-viewpoint cameras, light field cameras capture scenes from multiple perspectives, preserving realistic parallax and capturing the direction of light rays. Despite its potential, the substantial increase in data capture underscores the need for efficient compression techniques. This paper investigates the possibility of enhancing LF data compression by strategically omitting certain views during transmission, and then recreating them at the receiver's end through a specialized synthesis method tailored for light fields. The performance evaluation reveals a delicate balance between bandwidth efficiency and image reconstruction quality in the LF compression and transmission. We believe that a more efficient view synthesis approach that capitalizes on all directional light field views, holds the promise of enhancing LF compression performance.**

Keywords*— *Light field video compression, prediction structures, view synthesis.

I. INTRODUCTION

Light Field (LF) imaging, also referred to as plenoptic imaging, stands as a transformative technology in constant evolution, aiming to replicate human-like visual data and faithfully emulate our visual environment [1]. Diverging from traditional cameras, which capture scenes from a single viewpoint, light field cameras capture light from multiple perspectives, preserving realistic vertical and horizontal parallax. This not only records light intensity but also captures the direction of light rays. The richness of data obtained facilitates post-scene adjustments such as depth of field, focal point, and resolution. Moreover, the inclusion of depth and distance data enhances capabilities in segmentation and object detection. Light field technology finds diverse applications in cinematography, augmented/virtual reality, and medical contexts.

Despite its potential, the significant increase in data capture underscores the crucial need for efficient compression techniques. Traditional compression standards are inadequate for handling the unique characteristics of light field data. Therefore, the development of an effective encoding method is pivotal for managing this vast amount of data, enabling the technology to thrive and unlocking new market opportunities.

State-of-the-art LF compression methods focus on organizing keyframes (I and P frames) and leveraging horizontal and vertical similarities within the hierarchical bi-directional (B) frames. Khoury et al. [2] innovatively positioned the I-frame at the center and expanded the structure by placing the P-frames at the furthest cells horizontally and vertically, achieving a 38% bitrate reduction compared to LF-MVC. Mehajabin's et al. [3] approach utilizes a Structural Similarity Index Measure (SSIM) based selection strategy, determining correlations among views being predicted and their references, and accordingly choosing different types of frames. This method, an extension of Multiview-HEVC, demonstrates a 17% improvement in compression over [2], establishing it as the current state-of-the-art for LF video compression.

In this paper, we explore the potential for achieving greater compression efficiency with light field data by selectively omitting certain views during transmission and subsequently synthesizing them at the receiver end using a synthesis method tailored for LF [4].

The rest of this paper is organized as follows. In Section II, we describe our proposed approach. Section III presents the

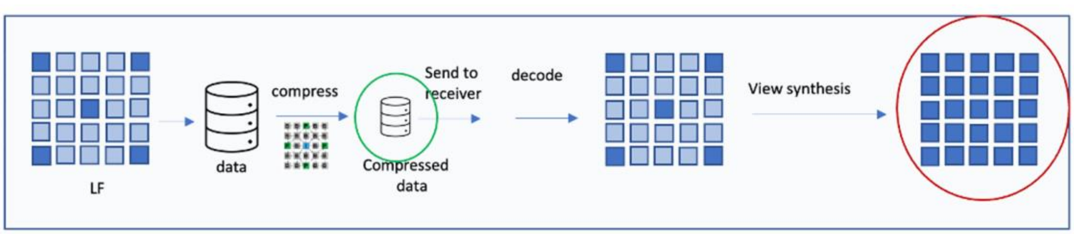
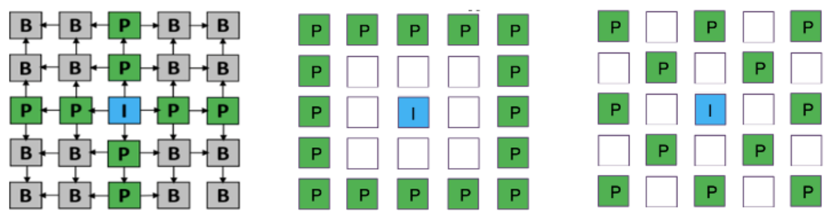


Figure 1. Overview of our proposed workflow.



(a) (b) (c)

Figure 2. LF video structures, (a) all original views, (b) peripheral, (c) raster skip.

performance evaluation of our method and discusses the results. Finally, Section IV concludes our paper.

II. Our Proposed Method

Our objective is to investigate the potential for achieving greater compression efficiency with light field data by omitting certain views during transmission and then synthesizing them at the receiver end using a synthesis method tailored for LF. To this end, our first task is to compress all the views of the original video sequence as well as two other view structures, which have some of the views omitted. Then, we employ a view synthesis approach designed for LF to reconstruct the missing views at the receiver end. Figure 1 shows an overview of the proposed workflow. The following subsections describe in detail our end-to-end approach.

A. Compression

In the first phase of our approach, we consider compressing three different view arrangements, namely the original video sequence that includes all 25 views and two other arrangements shown in Figure 2, where some of the views have been intentionally omitted. We name these two arrangements (structures) peripheral (Figure 2b) and raster skip (Figure 2c).

In our implementation, we choose to use the state-of-the-art LF compression method proposed by Mehajabin et al., also known as Universal Pseudocode Structure (UPS) [3]. This compression method uses a SSIM based selection strategy to determine the correlation among the views being predicted and their references and choose accordingly the different type of frames. Based on that, it utilizes a hierarchical B-frame prediction structure, which leverages both horizontal and vertical correlation to encode the different views/frames (see Figure 2a). An important advantage of this approach, beyond its exceptional compression efficiency [3], is that it is scalable to different view arrangement as well as view numbers and has both encoding and decoding parallelisms. This is very important for our approach, as we can consider any arrangement of horizontal and vertical views based on which ones we want to transmit.

B. View Synthesis

The missing LF views which were dropped at the transmitting end, are synthesized using the Wafa's et al.'s method [4], a state-of-the-art view synthesis approach that is based on a GAN learning-based model that is trained using the spatial and angular information of light field video content.

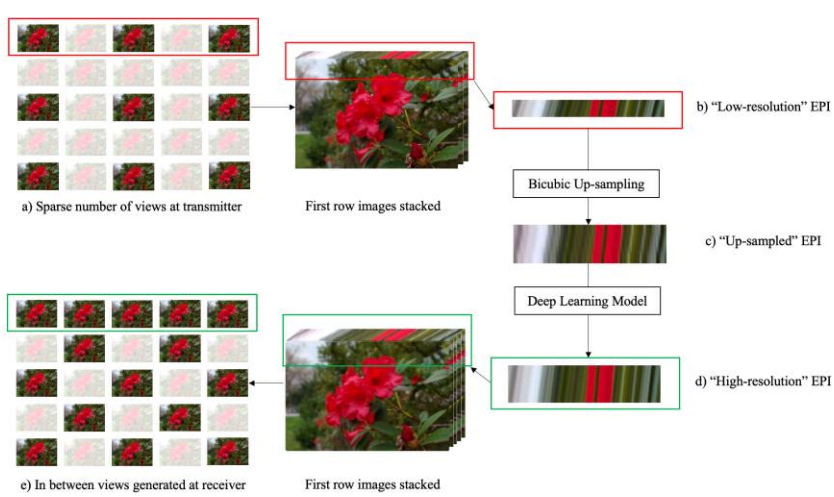


Figure 3. Framework of the LF view sythesis approach used in our implementation.

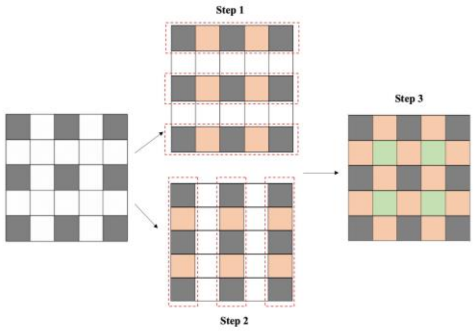


Figure 4. The three steps for synthesizing in between views.

This view synthesis approach is based on a deep learning model that is trained using the Epipolar Plane Images (EPI) of light field content. The overall framework is shown in Figure 3. The transparent views are the ones dropped at the transmitting end, leaving only the other views to be transmitted or stored. The EPI of rows and columns are fed to a modified Deep Recursive Residual Network (DRRN), which produces the synthesized views at the receiver end.

As seen in Figure 3 (a), some of the views are omitted at the transmitting end. The frames of the first row are then stacked to obtain the "low-resolution" EPI (Figure 3 (b)). Next, the "up-sampled" EPI is obtained by using bicubic interpolation, as seen in Figure 3 (c), to obtain an EPI of the same size as the original number of views. The upscaled EPI is used by the trained model to attain the "full size" EPI, shown in Figure 3 (d). This EPI is finally used to generate the in between views at the receiver end (Figure 3 (e)).

In general, the reconstruction process of synthesizing the in between views for the 4D light field, involves the three steps that are shown in Figure 4. In the first step, a horizontal 2D EPI is generated for each row of the inputs, as explained in Figure 3. This step is repeated for every row of views in the scene. In the second step, the vertical columns of views are used to generate the up-sampled vertical 2D EPIs and then these are used to reconstruct the intermediate views in the LF columns. In the third step, the synthesized views from either the horizontal or the vertical EPIs are utilized to generate the remaining views, represented by green boxes in Figure 4 (step 3).

III. Performance Evaluation and Discussion

We evaluated our approach on publicly available microlens based light field videos captured by the Raytrix LF camera [5]. The video sequences are 30fps with 2K resolution and duration of 10 seconds. We examined various compression quality levels by setting QP values to 25, 30, 35 and 40.

Figure 5 provides a comparative analysis of bitrate versus average Peak Signal-to-Noise Ratio (PSNR) for light field video compression structures Raster skip, Peripheral, and the Universal Prediction Structure applied to all the original views, which is the standard for comparison in this context. Note that here we only consider the bitrate and PSNR of the views transmitted, meaning that the raster skip and peripheral

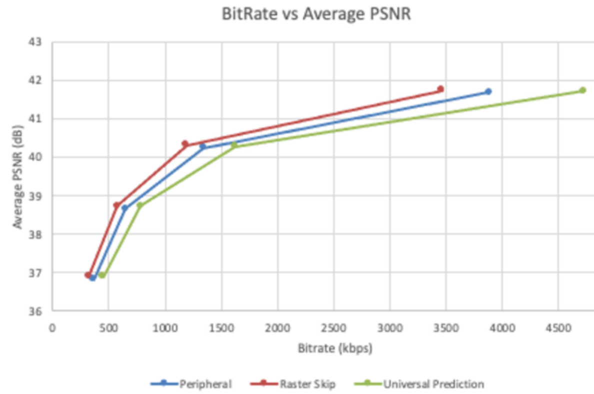


Figure 5. Performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, only for the transmitted views.

have fewer views. We observe that although all the methods seem to achieve similar compression results, there are still distinct differences in their performance.

The raster skip strategy outperforms the others, including the peripheral method, which is slightly more aligned with compressing all the available views (UPS). The superiority of raster skip can be rationalized by the fact that it omits the most views, therefore reducing the amount of data that needs to be compressed. It also means that Mehajabin's compression approach is very efficient, performing extremely well for the case that views are dropped, not really being affected from the fact that the remaining views are farther apart than the case that all views are available for compression (UPS).

Regarding view synthesis, we observe that when we consider a 3x3 window within the raster skip and peripheral structures, both share the same four corners as reference points (see Figure 6). It is worth noting that the raster skip structure has fewer views that need to be synthesized compared to the peripheral, which includes two adjacent views.

For a visual evaluation of the view synthesis approach, Figure 7a shows the original view and 7b represents the synthesized view for the raster skip structure. As it can be seen, the synthesized view looks almost identical to the original, careful observation of the background text and numbers indicates some visual artifacts such as reduction in text boldness.

Figure 8 depicts the performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, including the transmitted and synthesized views for the raster skip and peripheral structures.

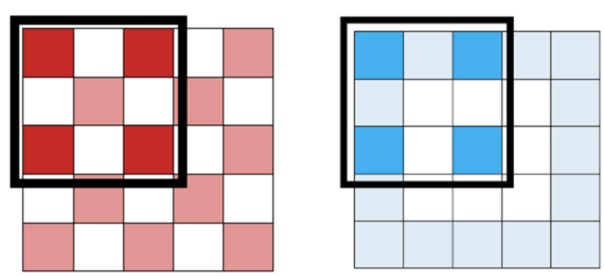


Figure 6. Window for the raster skip and peripheral structures.

(a)

(b)

Figure 7. (a) Original, (b) Synthesized (raster skip).

We observe that although at low bitrates the raster skip performs better, at medium and high bitrates compressing and transmitting all the views shows better efficiency. We could conclude that despite the advancements shown in our synthesized outputs, there remains a gap when compared to the results obtained from configurations utilizing all available views. Our synthesized version, while impressive, understandably falls short of this benchmark due to the fact that it does not take simultaneously take advantage of horizontal, diagonal and vertical views. Our future work focuses on improving the existing view synthesis approach to address this shortcoming and we feel very optimistic that this improvement will yield the desirable bandwidth savings.

IV. CONCLUSIONS

In this paper, we explore the potential for improving compression efficiency of light field video content by selectively omitting certain views during transmission and subsequently synthesizing them at the receiver end. The main reason for our quest is that although there have been notable advancements in compression algorithms for light field video content, there is still a need for additional enhancements in compression. These improvements are essential to fulfill the bandwidth requirements for the practical application of light field technology.

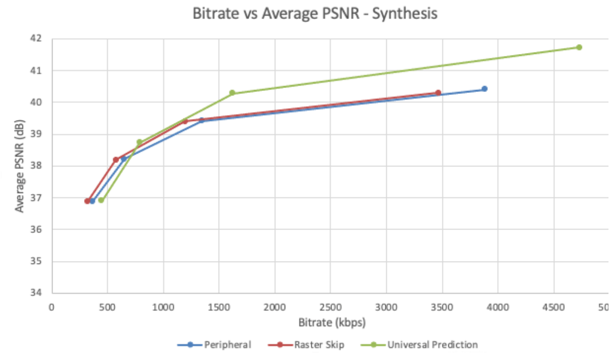


Figure 8. Performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, including the transmitted and synthesized views for the raster skip and peripheral structures.

Our findings highlight the delicate balance between bandwidth efficiency and reconstruction quality in light field compression and transmission. While our current attempts did not yield the desired outcomes, we believe that a more efficient view synthesis approach, capitalizing on all directional light field views, holds the promise of enhancing compression performance. This will be the primary focus of our future work in this domain.

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC – PG 11R12450), and TELUS (PG 11R10321). This research was enabled in part by support provided by the Digital Research Alliance of Canada.

REFERENCES

[1] I. Ihrke, J. Restrepo, and L. Mignard-Debise, Principles of Light Field Imaging: Briefly revisiting 25 years of research. In IEEE Signal Processing Magazine, vol. 33, no. 5, pp. 59 – 69, September 2016.

[2] J. Khoury, N. Mehajabin, M. T. Pourazad, P. Nasiopoulos, and V. C. M. Leung, An efficient three-dimensional prediction structure for coding light field video content using the MV-HEVC standard. In International Journal of Multimedia Intelligence and Security, vol. 4, no. 1, pp. 47-64, 2022.

[3] N. Mehajabin, M. T. Pourazad and P. Nasiopoulos, An Efficient Pseudo-Sequence-Based Light Field Video Coding Utilizing View Similarities for Prediction Structure. In IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 4, pp. 2356-2370, April 2022, doi: 10.1109/TCSVT.2021.3092282.

[4] A. Wafa and P. Nasiopoulos, Light Field GAN-based View Synthesis using full 4D information. ACM SIGGRAPH European Conference on Visual Media Production (CVMP), London, UK, December 2022.

[5] L. Guillo, X. Jiang, G. Lafruit, and C. Guillemot, Light field video dataset captured by a R8 Raytrix camera (with disparity maps), International Organisation for Standardisation iso/iec jtc1/sc29/wg1 & wg11, phd diss., 2018.

 ISBN: 978-1-68558-132-9