An Efficient Compression Strategy for Light Field Video Data Management

Maissan Bazazeh, Hamid Reza Tohidypour, Junbin Zhang, Panos Nasiopoulos, Zizhou Yu, Yanjun Meng, Gengran Li, Avneet Singh and Abrar Wafa
Electrical & Computer Engineering, University of British Columbia
Vancouver, BC, Canada
email: {mbazazeh, htohidyp, zjbthomas, panosn, zizhouyu, yjmeng, ligenran, asing171, awafa}@ece.ubc.ca

***Abstract*—Light field (LF) technology, facilitated by multi-lens cameras, captures comprehensive light arrays from all directions, revolutionizing visual experiences. Despite its immersive potential, the high data volume poses challenges in practical applications due to bandwidth and storage requirements. Traditional compression standards are ill-suited for processing LF data, necessitating the development of efficient encoding techniques. This constraint underscores the need for innovative approaches to manage the extensive data volume, driving technological progress and creating new market prospects. This paper investigates the possibility of dropping half of the views at the transmitting end and synthesizing them using a 4D LF view synthesis approach to reduce the required bandwidth while maintaining visual quality. Performance evaluations across diverse LF video content showed that our suggested approach outperforms prior methods, yielding significant bandwidth savings without compromising visual quality. This paper provides insights into future directions for LF compression research.**

Keywords- Light field; compression; view synthesis.

I. INTRODUCTION

Light field (LF) imaging, also referred to as plenoptic imaging, stands at the forefront of technological innovation, continually advancing to provide a visual experience akin to human perception [1]. Unlike traditional cameras, which capture scenes from a single viewpoint, LF cameras capture light from multiple perspectives, preserving realistic vertical and horizontal parallax [2]. This not only enables the recording of light intensity but also the direction of light rays, resulting in rich data that allows for adjustments in post-processing, such as depth of field, focal point, or resolution. Moreover, the inclusion of depth and distance information facilitates tasks like segmentation and object detection. LF technology finds applications across various domains, including cinematography, augmented/virtual reality, and medical fields.

However, the substantial increase in captured data highlights the critical need for efficient compression techniques. Conventional compression standards prove inadequate for handling LF data. Consequently, the development of effective encoding methods becomes pivotal in managing this vast volume of data, ultimately enabling the technology's advancement and unlocking new market opportunities.

Current state-of-the-art LF compression methods revolve around the organization of keyframes (I and P frames) and the exploitation of horizontal and vertical similarities within the hierarchical bi-directional (B) frames. Khoury et al. introduced a novel approach, positioning the I-frame at the center and expanding the structure by placing the P-frames at the furthest cells horizontally and vertically [3]. This technique achieved a remarkable 38% reduction in bitrate compared to LF-MVC [4].

In another significant advancement, Mehajabin et al. proposed an approach that utilizes a Structural Similarity Index Measure (SSIM) based keyframe selection strategy [5]. This strategy assesses the correlation among the views being predicted and their references to choose the appropriate frame types. This method, an extension of Multiview-High Efficiency Video Coding (MV-HEVC), demonstrated a 17% improvement in compression efficiency over [3], establishing it as the state-of-the-art for LF video compression at the time of this article's composition. Despite these notable advancements in compression efficiency, practical applications stand to benefit significantly from further reductions in data size. Such reductions are crucial for enabling future developments and justifying the transition from existing technologies to LF-based ones. A newer approach, outlined in [6], investigates the possibility of enhancing compression efficiency with LF data by selectively excluding certain views during transmission and then reconstructing them at the receiver's end. However, the evaluation of its performance is limited to a specific type of LF video content, while revealing that transmitting all views remains superior primarily because of the inefficiency of the view synthesis scheme.

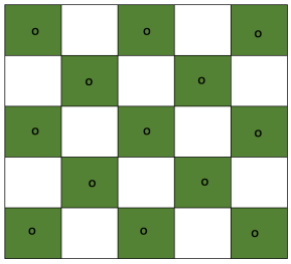

Figure 1. Raster skip structure of LF views. Views shown in green are the only ones transmitted.

In this paper, we expand upon the approach of prioritizing the transmission of a reduced number of LF views. However, in contrast, we employ an enhanced view synthesis scheme tailored to leverage the 4D nature of LF videos, while our study incorporates a diverse range of LF video content,

varying in resolution and complexity, and our evaluations include both subjective tests and objective metrics. More specifically, we focus on compressing and transmitting alternate views of the LF stream, as shown in Figure 1. The compression efficiency of this configuration, referred as raster skip in the rest of this paper, is compared against the case of compressing and transmitting all the available LF views. Our findings indicate that this method is very promising, surpassing the performance of the original approach presented in [6].

The rest of paper is structured as follows: Section II outlines our Proposed Methodology. Section III presents the Evaluation and Experimental findings. Section IV provides the Conclusion of our study and future research directions.

II. PROPOSED METHODOLOGY

Our objective is to compare the bandwidth requirements for transmitting LF video content using two different approaches: the conventional method, which compresses all views, and the proposed method, which omits half of the views and synthesizes them at the receiver end. These two schemes are illustrated in Figure 2. Figure 2(a) depicts the proposed LF delivery approach, which compresses every other view, while the intermediate views are excluded during transmission. Subsequently, our suggested 4D LF view synthesis approach, based on deep learning, is employed to generate the missing views. On the other hand, Figure 2(b) represents the traditional approach, where every LF view is compressed and transmitted to the receiver.

A. LF Compression: All Views and Raster Skip Structures

In this paper, the LF structure that includes all the views serves as the reference method for comparison with the raster skip structure that uses only every other view (see Figure 1). In the latter scenario, the amount of data that needs to be compressed can be significantly reduced, although the redundancies between adjacent views are reduced which may somewhat increase the motion vectors and residuals involved in compression. In our implementation, we choose to use the state-of-the-art LF compression method proposed by Mehajabin et al. [5]. This compression method uses a SSIM based keyframe 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.

B. 4D LF View Synthesis Method

The view synthesis approach we utilize in our implementation is centered on a 4D LF deep learning network that is based on the concept of Generative Adversarial Network (GAN) architecture, which is retrained using the spatial and angular data within LF video content [7]. The proposed network takes Multi-Perspective Light Field (MP-LF) images as input and employs a GAN framework to generate novel virtual LF views (see Figure 3).

In this approach, the generator extracts spatial and angular features from the MP-LF image, which are then fed into a residual sub-network to produce spatial and angular feature maps. Subsequently, up-sampling is employed along the angular dimension using sub-pixel convolution techniques on the MP-LF feature maps. Finally, the pixels are reorganized to obtain the Synthesized Angular Light Field (SA-LF) images containing the newly synthesized LF views.

On the discriminator side, a patch discriminator learns the distribution of image patches by examining multiple local patches and determining their authenticity. In essence, the Generator is trained to create intermediary LF images in such a way that the Discriminator cannot distinguish between the patch distribution of the generated SA-LF images and that of the Ground Truth (GT) SA-LF images that were omitted.

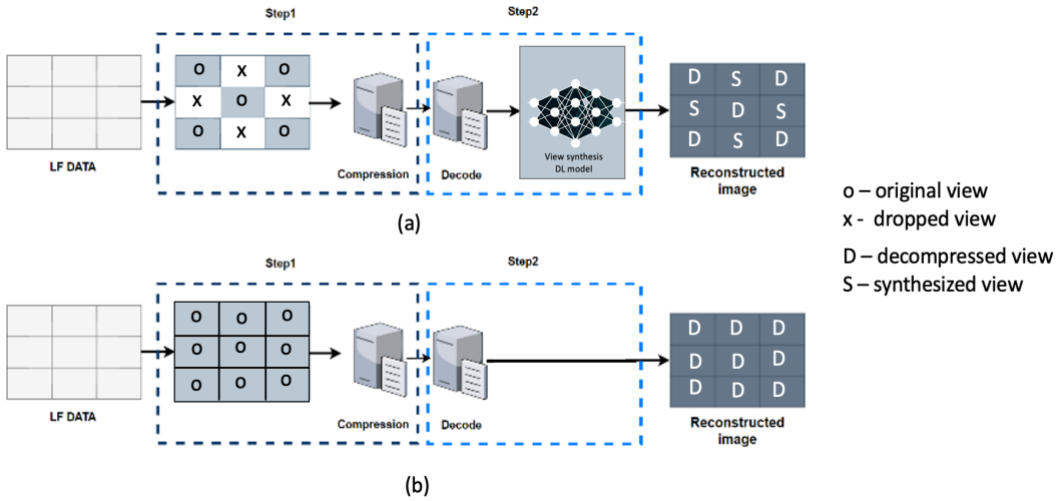

Figure 2. The two schemes to be compared: (a) every other LF view is dropped and only half of the views are compressed; the missing views are synthesized at the receiver end. (b) All LF views are compressed and transmitted.

III. EVALUATION AND RESULTS

To comprehensively evaluate the proposed approach, we use the four different LF video sequences shown in Table I. The Chess and Boxer videos were captured by the Raytrix R8 LF camera, have 5x5 views, resolution 1920x1080 and a total of 300 frames each [8]. The Toys and Car LF videos, on the other hand, were captured by a Lytro Illum LF camera, consist of 15x15 views for a total of 100 frames each and have resolution of 480x320 and 512x352 pixels, respectively [9]. Figure 4 shows a sample frame for each of the above LF videos.

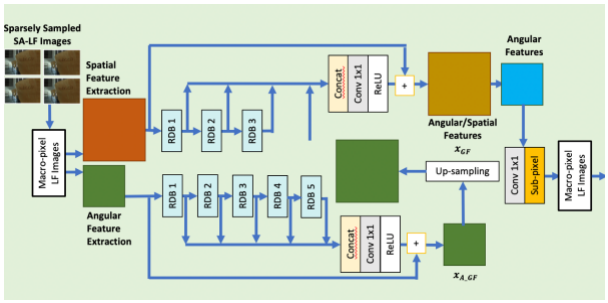

Figure 3. An overview of the 4D LF view synthesis network used in our proposed LF delivery approach.

We examined a wide range of compression quality levels by setting the Quantization Parameter (QP) values to 25, 30, 35 and 40, ensuring that all practical levels of visual quality and bitrates are covered. Figure 5 shows the average bitrate against visual quality for the suggested raster skip and all views structures in terms of Peak Signal-to-Noise Ratio (PSNR).

The Bjøntegaard Delta rate (BD-rate) using the piecewise cubic fitting from the PSNR and bitrate results calculated using the ITU-R BT.500-14 approach yields 36.87% savings [10]. As expected, the raster skip structure results in reduced bitrate requirements for all the different compression levels. This is also an indication that the LF compression scheme we chose for our study is very effective even for the case of raster skip where the adjacent views are further apart. For visual purposes, Figure 6 shows one view of the Boxer sequence: (a) shows the view that is decompressed for the scenario that we transmit all the views and (b) shows the same view but synthesized for the raster skip structure. As it can be seen the synthesized approach achieved similar visual quality as the traditional all-views approach.

A. *Subjective tests*

We performed subjective tests to evaluate the perceptual quality of the raster skip and all-views structures. In the concept of LF, views are displaced from each other, resembling a 3D effect. Thus, for each LF video, we created 3D stereo videos by pairing two adjacent views and showcased them on a 3D TV, prompting our participants to

Figure 4. Examples of datasets.

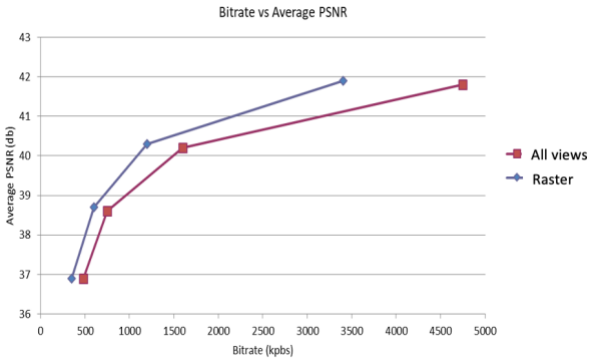

Figure 5. Average compression performance evaluation of bitrate against PSNR for the suggested raster skip and the all views structures.

TABLE I. LF VIDEOS

LF Video	Camera	Resolution	Views	Total frames
Chess-Pieces	Raytrix R8	1920x1080	5x5	300
Boxer-IrishMan-Gladiator	Raytrix R8	1920x1080	5x5	300
Toys	Lytro Illum	480x320	15x15	100
Car	Lytro Illum	512x352	15x15	100

evaluate their quality. For the subjective test, we employed the Degradation Category Rating (DCR) test method, which involved scoring each of the two compared methods in relation to the reference [11]. Our test consisted of 32 Basic Test Cells (BTCs), half of them were for the raster skip and the rest for the traditional approach. During the subjective test, for each BTC, the BTC number was shown on mid-grey background for two seconds, followed by the reference video for 10 seconds. Following that, the phrase "Impaired video" was displayed for 2 seconds on a mid-gray background.

Figure 6. Visual comparison of the same LF view (a) decompressed using the transmission of all views, (b) synthesized view for the Raster skip structure.

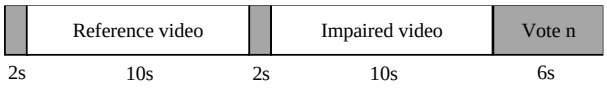

Figure 7. The degradation category rating (DCR) test method used for our subjective tests.

Afterward, the impaired video, which was generated by one of the two approaches, was shown for a duration of 10 seconds. Then, the subjects were given 6 seconds to score. If the video length was shorter than 10 seconds, the video was looped to ensure a duration of 10 seconds. Figure 7 shows the DCR test method.

To create synthesized stereo videos, we randomly selected two synthesized views of the raster skip positioned away from the border. Each of these synthesized views was paired with the previous horizontally-viewed frame from the traditional approach, resulting in the generation of a stereo video. One view was randomly selected from those located at the left border, and the next view from the traditional sequence was chosen to create a stereo video. Similarly, we randomly chose one view from those located at the right border, and the previous view from the traditional sequence was used to produce another stereo video. The same view arrangement from the original (uncompressed) views was used to make the corresponding reference stereo videos. Additionally, for the traditional approach, the same view arrangement from the traditional views was employed to produce the stereo videos.

Ten male and eight female subjects with ages ranging from twenty-four to thirty-two took part in our subjective evaluation. Prior to the test, all subjects were screened for color blindness and visual acuity using the Ishihara and Snellen charts, respectively. A training session allowed all subjects to become familiar with the test procedure. During the subjective test, the BTCs were shown to the subjects in a random order. For each BTC, the subjects were asked to score between 0 to 10 according to ITU-R BT. 500 [12]. Table II shows the meaning of the 11 grades of the numerical scale. After completing all the subjective tests, we applied the outlier detection method outlined in ITU-T P.913 to identify any outliers. One outlier was identified and subsequently excluded from the corresponding results.

Our raster skip approach achieved 13.91% savings in bitrate, as demonstrated by the calculated BD-rate for the same average Mean Opinion Score (MOS). These results are presented in Figure 8, which shows the average bitrate against visual quality for both the suggested raster skip and all views

TABLE II. MEANING OF THE 11 GRADES NUMERICAL SCALE

Score	Impairment item	
10	Imperceptible	
9	Slightly perceptible	somewhere
8		everywhere
7	Perceptible	somewhere
6		everywhere
5	Clearly perceptible	somewhere
4		everywhere
3	Annoying	somewhere
2		everywhere
1	Severely annoying	somewhere
0		everywhere

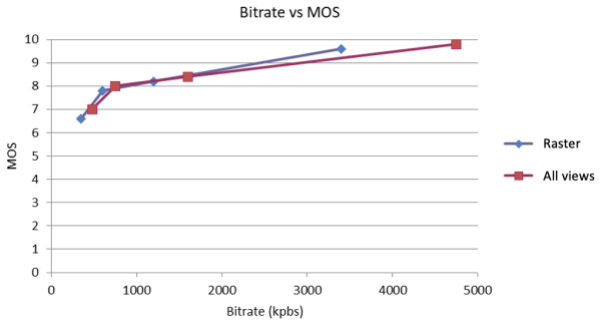

Figure 8. Average compression performance evaluation of bitrate against MOS for the suggested raster skip and the all views structures.

structures. This confirms that our raster skip method is very effective, even in cases where the adjacent views are further apart. This means the discrepancy in perceptual quality is nearly negligible, suggesting that synthesizing the omitted views within the raster skip structure instead of transmitting all of them does not result in any significant visual difference.

IV. CONCLUSION

This paper concentrates on tackling the challenge of managing the substantial data volume associated with LF content, and offers insights into a prospective avenue for future LF compression research. LF technology offers a paradigm shift in visual experiences, but despite its potential, the inherent challenge lies in the substantial data volume, necessitating efficient compression techniques for practical applications. Our study focused on exploring the feasibility of reducing bandwidth requirements by selectively dropping views at the transmission end and synthesizing them using a 4D LF view synthesis approach.

Through comprehensive evaluations across diverse LF video content, we demonstrated the efficiency of our proposed approach in achieving 13.91% bandwidth savings over existing methods without compromising visual quality.

Our research paves the way for more practical LF applications in various domains and sheds light on the future directions of LF compression research, emphasizing the need for continued innovation to address the evolving demands of this technology. As LF imaging continues to advance, unlocking new avenues for immersive visual experiences, the optimization of compression methods remains crucial for widespread adoption and integration into real-life applications.

In essence, our findings underscore the importance of efficient encoding methods in managing the vast data volume of LF imaging, driving technological advancements, and unlocking new market opportunities in the ever-evolving landscape of visual technology.

REFERENCES

[1] S. Zhou et al. “Review of light field technologies,” Visual Computing for Industry, Biomedicine, and Art, vol. 4, no. 29, pp. 1-13, Dec. 2021.

[2] W. Zhang, W. Ke, D. Yang, H. Sheng, and Z. Xiong, “Light field super-resolution using complementary-view feature attention,” Computational Visual Media, vol. 9, pp. 843-858, July 2023.

[3] J. Khoury, M. T. Pourazad and P. Nasiopoulos, "A New Prediction Structure for Efficient MV-HEVC based Light Field Video Compression," 2019 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, pp. 588-591, 2019.

[4] G. Wang, W. Xiang, M. Pickering, and C. W. Chen, “Light Field Multi-View Video Coding with Two-Directional Parallel Inter-View Prediction,” IEEE Trans. Image Process., vol. 25, no. 11, pp. 5104– 5117, 2016.

[5] N. Mehajabin, M. Pourazad, and P. Nasiopoulos, ”SSIM Assisted Pseudo-sequence-based Prediction Structure for Light Field Video Compression,” 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-2, 2020.

[6] T. Bazzaza et al. ”SSIM Assisted Pseudo-sequence-based Prediction Structure for Light Field Video Compression,” 2024 IARIA The Thirteenth International Conference on Intelligent Systems and Applications (INTELLI), Athens, Greece, pp. 43-46, March 2024.

[7] A. Wafa and P. Nasiopoulos, “Light Field GAN-based View Synthesis using full 4D information,” in CVMP '22: Proceedings of the 19th ACM SIGGRAPH European Conference on Visual Media Production, pp. 1-7, Dec. 2022.

[8] L. Guillo, X. Jiang, G. Lafruit, and C. Guillemot, “Light field video dataset captured by a R8 Raytrix camera (with disparity maps),” ISO /IEC JTC1/SC29/WG11 MPEG2018/m42468, ISO/IEC JTC1/SC29/WG1 JPEG2018/m79046, international organisation for standardization, ISO/IEC JTC1/SC29/WG1 & WG11, San Diego, CA, US, April 2018.

[9] B. Wang, Q. Peng, E. Wang, K. Han and W. Xiang, "Region-of-Interest Compression and View Synthesis for Light Field Video Streaming," in IEEE Access, vol. 7, pp. 41183-41192, 2019.

[10] K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström, G. J. Sullivan, A. Tourapis, “Working practices using objective metrics for evaluation of video coding efficiency experiments,” ITU-T Tech. Paper HSTP-VID-WPOM and ISO/IEC TR 23008-8, July 2020.

[11] Recommendation ITU-T P.910 (2008), Subjective video quality assessment methods for multimedia applications.

[12] Recommendation ITU-R BT.500-14 (2019), Methodologies for the subjective assessment of the quality of television images.