Light Field Video Compression

EECE 541 W2018

Group Project Proposal

Jashandeep Kaur	96939160
Sichen Luo	45323136
Nusrat Mehajabin	23647738
Haowei Yu	33955121

Instructor: Dr. Mahsa T. Pourazad

Department of Electrical and Computer Engineering

The University of British Columbia

(Vancouver)

Oct. 29, 2018

Table of Contents

List of Figures 4

List of Tables 4

1. Introduction 5

2. Background Knowledge 6

2.1 Single View Images 7

2.1.1 Digital Images 7

2.2 Multi-View Images 8

3. Light Field Video and Challenges 9

4. Multi-View Video Compression 10

4.1 How Does it Work? 10

4.2 Difference between MV and Light Field 10

5. Existing LF Video Compression Methods 11

5.1 Conventional Approaches 11

5.1.1 Simulcast 11

5.1.2 Transposed Picture Ordering 12

5.1.3 Multi-view Video Coding 12

5.2 Improved Approaches 13

5.2.1 LF-MVC 13

5.2.2 MV-HEVC (Enhanced Spatial Prediction for Compression Efficiency) 14

5.2.3 Distance-Based Reference Selection 14

5.3 Calculation Equation 15

5.3.1 Peak Signal to Noise Ratio (PSNR) 15

5.3.2 Bjøntegaard model 16

6. Random Access Methodology 16

7. Evaluation of Existing Proposed Methods 17

7.1 Performance Metrics 18

8. New Proposed Method 18

8.1 Content Aware Correlation Based Pseudo Sequence 19

8.1.1 Discussion 20

8.2 Pseudo Sequence Facilitating Random Access 21

8.2.1 Discussion 22

9. Conclusion 22

References 24

List of Figures

Figure 1 Multi-View Video Compression Overview 5

Figure 2 Placeholder 6

Figure 3 GOP Structure 9

Figure 4 Multi-View Video Compression Method 10

Figure 5 Light Field Video View Structure 11

Figure 6 Simulcast Encoding 11

Figure 7 Transposed Picture Ordering 12

Figure 8 One-Directional 2D Multi-view Compression 12

Figure 9 LF-MVC Structure 13

Figure 10 LF-MVC Inter-view Reference Structure 13

Figure 11 MV-HEVC Inter-view Reference Structure 14

Figure 12 Distance-Based Reference Selection 15

Figure 13 MV-HEVC Frame Accessibility 16

Figure 14 Random Access for Individual Frames 17

List of Tables

Table 1 Hierarchy of frames 19

1. Introduction

Traditional camera's record only two-dimensional representation of the scene. In contrast, light field technology allows capturing richer visual information from the world. The distribution of light in free space is best described by the light field. The light field is a function which describes the amount of light flowing in every direction in every point of space. The Light Field is like a magic window in another field. The benefit of the light field is that it represents the scene in the same way it captures. There are various devices for recording light fields such as handheld cameras, an array of cameras, preoptic camera, etc. The set of light rays travelling in every direction in every point of 3D space is described by a multidimensional function. This function is represented by L (x,y,z,Θ,phi, γ,t). For this function to generate, there is a need to measure the light rays from every possible angle, location, at every time t and at every wavelength [1]. Light field videos are very large in size and there is data redundancy, so there is a need to compress these videos. The plenoptic camera does not record light field as regular photographs instead it captures as a lenslet images. The idea behind this is to insert an array of microlenses in front of image sensors.

Multi-View video compression is used to compress the light field videos. This method is most efficient. The overview of Multi-View video compression is shown in following figure 1:

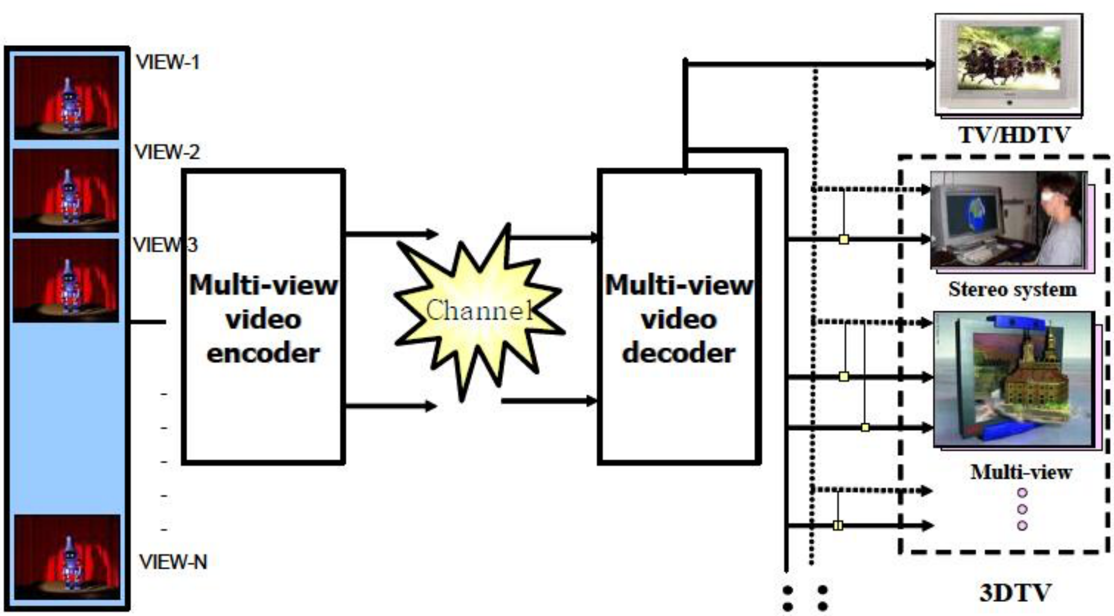

Figure 1 Multi-View Video Compression Overview

The multi-view encoder takes multiple views as input and generates a bit stream which is then passed to video decoder and decodes the input signal. Multi-View video coding exploits the temporal and spatial redundancies contained in the light field video sequences. MVC is an extension of H.264/MPEG-4 AVC standard [2].

In this research, we will measure existing Light field video compression techniques and will determine a unique method. The objective measure which we will use for the evaluation of video compression algorithms is PSNR. First, we will provide background knowledge on different types of images. After that, we will look at light field video challenges and existing compression methods. Finally, we will propose a new method for video compression.

2. Background Knowledge

The most effective way for humans to perceive the world is through vision. And human vision is assisted by the light rays from our environment. These light rays carry abundant information about our 3D world. Capturing and processing this rich information without the human visual system is an age-old pursuit. In this section, we discuss the chronology of photography from pinhole camera images to 4D light field images. Fig. 2 is an exhibition of this chronology.

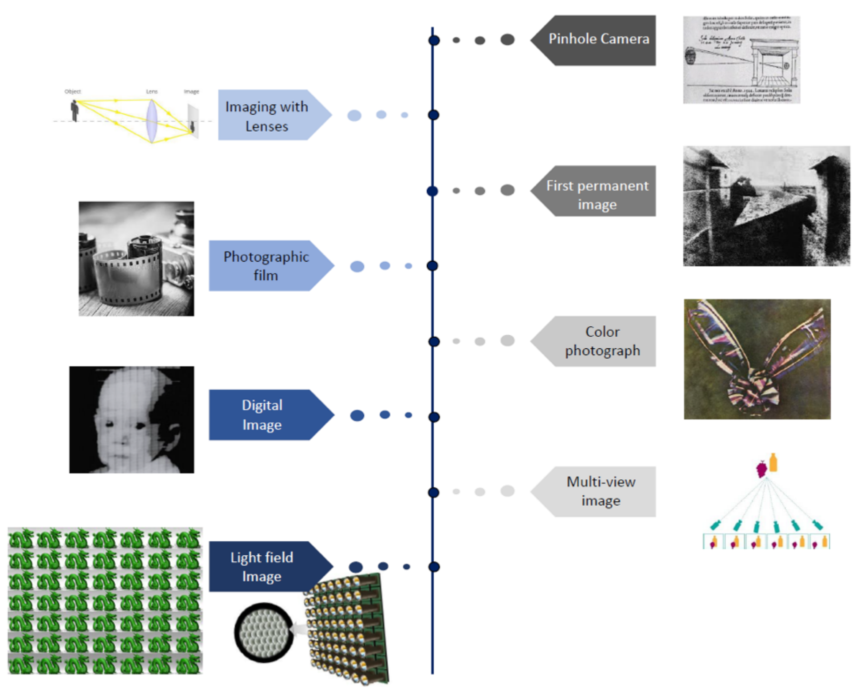

Figure 2 chronological diagram of photography

2.1 Single View Images

The very first images were captured using a dark chamber with a pinhole [3] in it. Light rays would pass through the pinhole and focus on a wall or translucent plate at the back of the dark chamber. This simple yet effective model is the baseline of all modern imaging process. Nevertheless, this model is not ideal, as to create a pinhole which will let only one ray of light focus on one point on the imaging plane is impractical [4].

Therefore, to induce focus of the light rays lenses were introduced [5]. Different lens architectures and combinations of lenses were used to capture images and compensate for different aberrations (e.g. aberration, distortion, chromatic aberration etc. [6]) However, these photographs were not permanent. In the year 1816, the first permanent photograph was captured in a pewter plate coated in bitumen [7]. But the process was time-consuming and injurious to the health of the photographer. Later in 1880, the film was developed diminishing the need for chemicals and metallic plates.

The next major leap forward in the history of image capture is marked by the development of color plates in 1907 [8]. The first color plates were called Autochrome and used a screen that filtered the light using dyed potato starch. Once the plate was developed, colors could be restored. The commercially viable color photography used dye-coupled colors in a three-layer system [9] and is still the way we make color photographs using film.

2.1.1 Digital Images

The first recorded attempt at digital image formation was in 1975 with a camera weighing 4 kilograms and it took 20 seconds to capture a low resolution black and white image [10]. The storage medium used was cassette tapes. However, things escalated fast from there and by 1988 the first true digital camera Fuji DS-1P [11] was developed. By 1990 digital cameras were commercially available and could connect to computers for image download. And since then, numerous advantages have been made in digital image arena, such as image resolution, color, storage, transmission etc. [12].

2.2 Multi-View Images

Single view or single perspective cameras can only capture so much information about the 3D world. By gathering light ray information on a deeper dimension, we can have a much broader understanding of the scene.

The most trivial way to do this is to put more than one camera at a scene. Capturing a scene using 2 cameras will give us two points of perspectives on the scene and we will have 2 facet information [13]. Deriving from this statement putting multiple cameras will give us not only multiple views of the same scene but also provide us with angular information of the scene. This information was missing from the single view images where we only had the spatial information. Being able to extract angular information along with spatial information opens the door to numerous applications.

One variant of multi-view imaging is to have an array of cameras taking photos of the scene simultaneously. This setup will capture the vertical angular information of the scene. A 2-dimensional arrangement of cameras will gather both vertical and horizontal angular information. However, having several cameras pointing to a scene is intrusive and not practical in most cases (e.g. medical imaging). A new kind of camera was designed to mimic the effects of such 2D camera arrangement. These are the light field cameras. These cameras are able to capture light field images and videos [14].

The basic architecture of LF cameras includes the main lens which focuses the incoming light rays to the lenslets standing between the imaging plane and the main lens. Each of these lenslets will create a slightly deviated image (image from a different viewpoint) of the scene on the imaging plane [15].

LF imaging facilitates richer content capture, visualization, and manipulation of the data [16]. LF bring about many different areas of research, e.g., 3D television, biometric recognition, and medical imaging [17]. Among the advantages of employing an LF imaging system is the enabling of new degrees of freedom in terms of content production and manipulation, thus supporting functionalities not straightforwardly available in conventional imaging systems, namely, post-production refocusing, changing depth-of-field, and changing the viewing perspective.

The lucrative promises of LF are not without drawbacks and challenges. In the next section, we discuss two of the major challenges of light field video.

3. Light Field Video and Challenges

A light field video captures the scene from different viewpoints so it provides the powerful capability for understanding the visual data information. Light field provides deep information about the objects in the scene.

On the other hand, a light field video contains more redundancy as well as a large volume of data due to the fact that it captures the scene from different viewpoints. The big challenges in light field videos is how to store and display such large amount of data, so efficient compression techniques are necessary for this and as well as to reduce the transmission costs. The second challenge in light field video is how to analyze and understand the visual information [18]. The simple approach for light field video compression is to encode each view sequence separately. The disadvantage of this scheme is that for each view sequence, it considers only the inter-frame correlation between the frames not the strong correlation [19].

In this scheme first frame of each view is encoded as I-frame and the remaining frames as B-frames. There are three main frames in video coding standards:

I-frames are intra-coded frames which don't require to decode any other frames to obtain information.
B-frames are bi-directionally predicted frames which are decoded using the information from both the previous and forward frames.
P-frames make use of previous I-frames or P-frames to encode.

In video encoding, the GOP structure specifies the positioning of I-frames, P-frames and B-frames [20]. Figure 3 displays the decoding structure of I, P, and B frames.

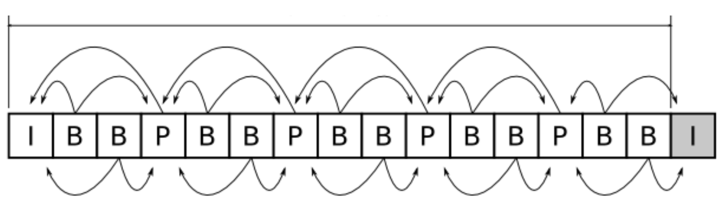

Figure 3 GOP Structure

The multi-view random access method is used to compress the video. In this method, the complexity is measured according to the number of frames it takes to reach any randomly chosen frame. The main problem in this method is that it is complex to access the desired frame [21].

4. Multi-View Video Compression

4.1 How Does it Work?

To understand light field video compression, it is essential to first understand how multi-view video compression operates. A multi-view video has a 2-dimensional array of images: a dimension in the spatial direction to represent the layers in a frame, and a dimension in the temporal direction for the number of frames in the video. To compress this, a pattern of I, P, and B frames to maximize both dimensions is used. Figure 4.1 shows a possible compression method used for a multi-view video. The sample used in the figure has a layer size of five and a frame size of eight. The method uses layer 1, frame 1 as the I-frame. This I-frame is then used to predict the P-frame in the spatial direction (layer 3, frame 1) and the B-frames in the temporal direction (layer 1, frame 3 and layer 1, frame 5). The P-frame is then used in the same structure to predict further P-frames and B-frames, and this pattern repeats until all images have been predicted.

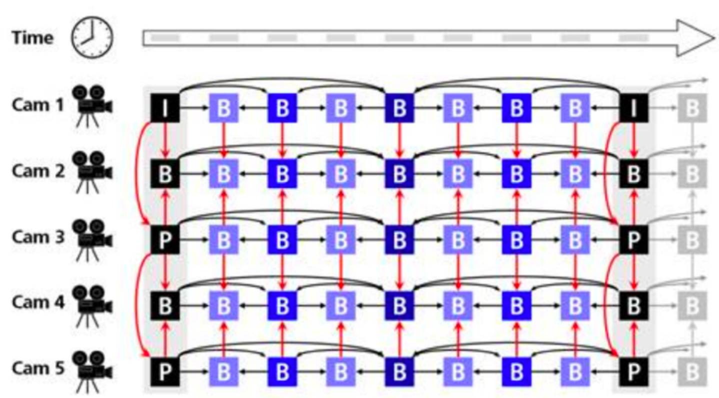

Figure 4 Multi-View Video Compression Method

4.2 Difference between MV and Light Field

However, unlike multi-view videos, light field videos capture images in a two-dimensional array of lenslets. Therefore, light field videos consist of a three-dimensional array of images. It contains one dimension in the temporal dimension to represent the frames but has two dimensions in the spatial direction for representing the layers. Figure 5 shows a representation of the light field video structure. This additional dimension in the spatial direction provides more possibilities for frame prediction since each frame now has an additional degree of freedom in possibilities. As a result, new methods have to be created to maximize compression for light field videos.

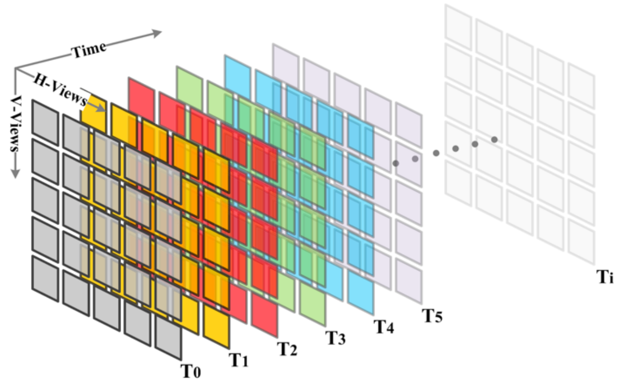

Figure 5 Light Field Video View Structure

5. Existing LF Video Compression Methods

5.1 Conventional Approaches

5.1.1 Simulcast

Simulcast encodes each view sequence separately. Every single view can be encoded using the hierarchical-B coding structure. (First frame is I frame; the rest are B frame). The result of such method for a N-view LF video is an N independent encoded video stream (Figure 6). Although this is a straightforward approach, it does not take advantage of the strong inter-view correlation to achieve a more efficient compression.

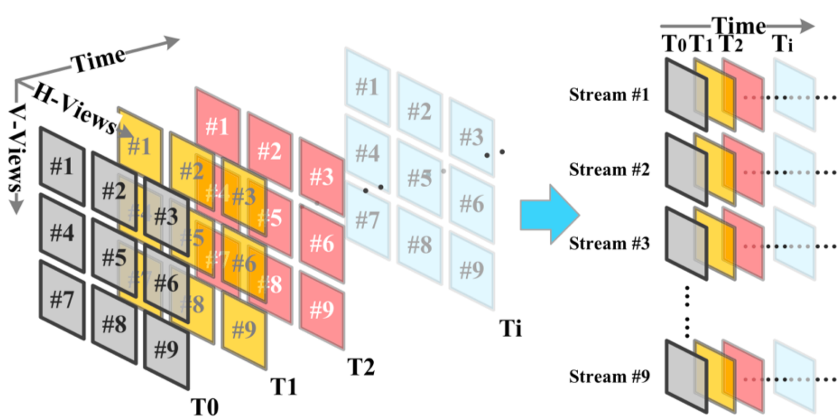

Figure 6 Simulcast Encoding

5.1.2 Transposed Picture Ordering

The transposed picture ordering takes advantage of the inter-view correlation by traversing all the views on the temporal plane. It turns the entire file into a 2D video sequence (Figure 7). However, this method does not take advantage of the temporal correlation.

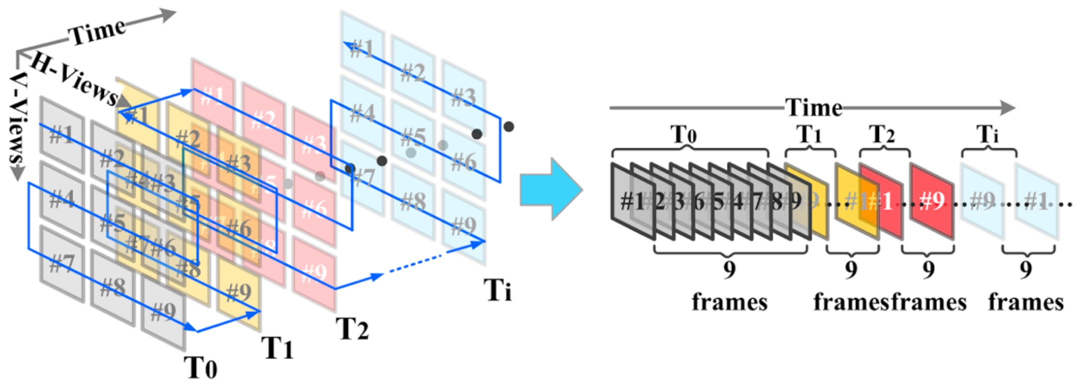

Figure 7 Transposed Picture Ordering

5.1.3 Multi-view Video Coding

In order to combine the advantage of inter-view and inter-frame correlations, one direct method is converting the 2D Multi-view video sequence into a 1D Multi-view video sequence by traversing the 2D Multi-view matrix with different topologies (Figure 8) such as Zigzag, row-by-row, perpendicular, and diagonal. The compression performance can be dramatically improved by this approach. The drawback of this method, however, is that the vertical correlation disappears after the array->vector conversion, and the converted view sequence might cause inefficient prediction.

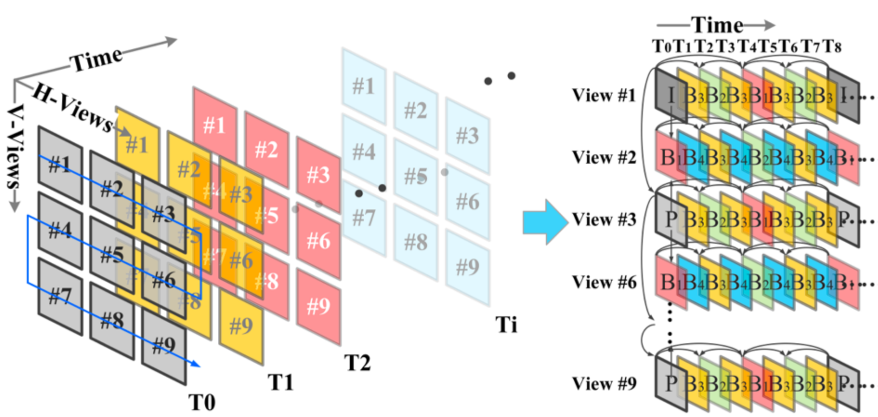

Figure 8 One-Directional 2D Multi-view Compression

5.2 Improved Approaches

5.2.1 LF-MVC

To improve the conventional approaches of LF video coding, Wang et. al introduce a new method - LF-MVC. The method shares the same inter-frame prediction scheme as the standard MVC. The difference is that LF-MVC incorporates a two-directional parallel prediction structure (Figure 9). Unlike the row-by-row prediction flow (Figure 10 Left), LF-MVC makes use of both horizontal and vertical correlations when frames are referenced (Figure 10 Right). The prediction flow diverges from the top left corner. The prediction flow forms closed-loops inside each squared sub-structure. This approach breaks the limitation from MVC that vertical correlation is unused, achieving a 26% bit-rate reduction for a 5x5 view LF video, and 34% for a 10x10 view LF video.

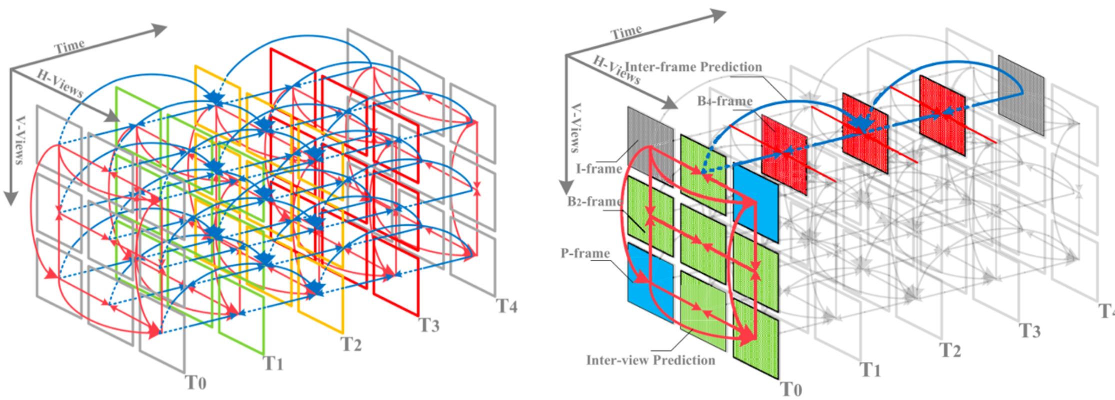

Figure 9 LF-MVC Structure

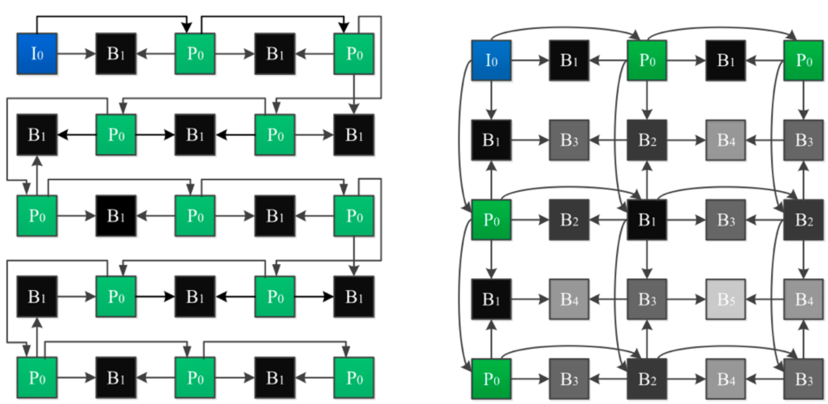

Figure 10 LF-MVC Inter-view Reference Structure

5.2.2 MV-HEVC (Enhanced Spatial Prediction for Compression Efficiency)

Khoury et al. discover that in light field video content there is a high similarity between the views around the center than the views around the corner. The authors also suggest that the diagonal distance from the top left to the bottom right corner is much larger than the distance from the center to the corners. Therefore, they propose a new two-dimensional inter-view prediction scheme using the Multi-view extension of HEVC (MV-HEVC). Instead of using the top-left frame in LF-MVC, MV-HEVC uses the central frame as the I-frame (Figure 11) to take advantage of the high similarity in that area. P frames located at the middle of each edge uses the central I-frame as the only reference, contributing to the B frames in the rest of the views. All four corner views are encoded as B frames. As a result, this approach allows the P frames to be distributed symmetrically at the end of the central axes and maximizes the number of outgoing edges from the P frame (5 directions). With a similar temporal prediction structure, this new inter-view prediction reduces the Bjøntegaard Delta Rate (BD-rate, see 5.3.2) by 38.18% (with a 5x5 view video) in comparison to the LF-MVC approach from 5.2.1.

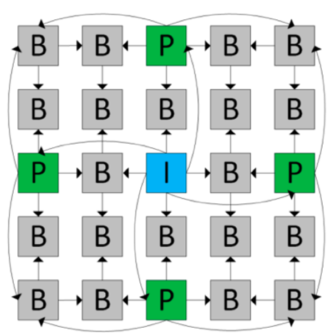

Figure 11 MV-HEVC Inter-view Reference Structure

5.2.3 Distance-Based Reference Selection

In the previous examples, the reference frames are selected statically (top left corner or center). Li et. al propose a hierarchical two-dimensional coding structure that can select better reference frames based on the distances between views. The solution constructs a two-dimensional hierarchical structure that subdivides the views into four individual quadrants to make better use of intercorrelations among all views (Figure 12 Left). To save buffer size, in each quadrant (Figures 12 Right), the views are encoded row-by-row, with a depth-first sequence: 0, 6, 3, 5, 4, 2, 1. Then, frames are divided into four categories based on their reference frequency within the quadrant. Whichever frame has higher frequency of being referenced will be stored in the buffer. (Red: most frequently referenced frame to guarantee all frames have a relatively near reference frame; Green: Second most frequently referenced, will be referenced by frames reside in the current row; Yellow: Only be referenced by the frame next to them; Black: non-reference frame.) For each frame, two lists - l0 and l1 are used to store the information of the forward and backward reference frames. The size of the list can be adjusted depending on how many reference frames are needed. The lists implicitly express the arrow

Their solution achieves an average reduction of both Y-BDrate and YUV-BDrate -6.5%. In their paper, the proposed method is applied to LF image compression.

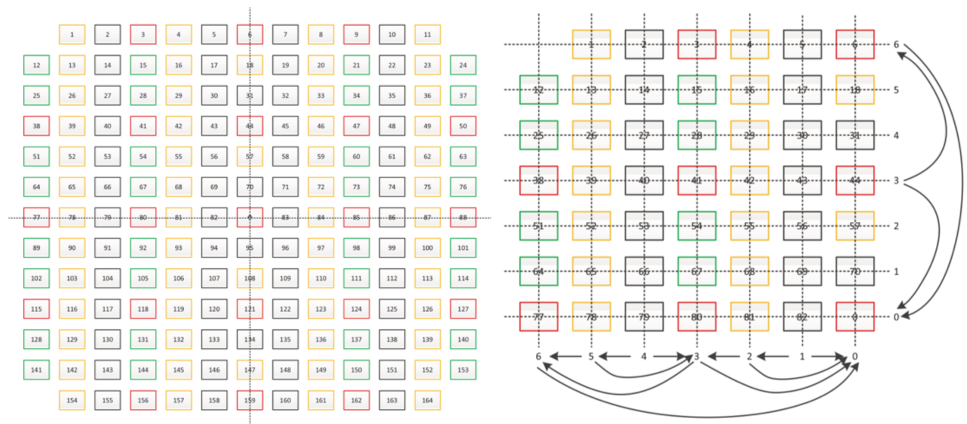

Figure 12 Distance-Based Reference Selection

5.3 Calculation Equation

5.3.1 Peak Signal to Noise Ratio (PSNR)

Peak Signal to Noise Ratio can be used to evaluate the objective quality of the compression, where MSE the Mean Square Error between the original videos and the compressed videos. L represents the maximum pixel value of the image. In most cases, including out project, pixels are represented with 8 bits per sample. Therefore, the value of $L = 2^8 - 1 = 255$.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$$

$$PSNR = 10\log_{10}\frac{L^2}{MSE}$$

5.3.2 Bjøntegaard model

Proposed by Gisle Bjøntegaard, the Bjøntegaard model is a well-known tool to evaluate the coding efficiency of a video codec with a reference codec over. The Bjøntegaard Delta bit rate (BD-rate) compares the average bitrate savings of one coding method over the other under the same video quality. In the equation below, $R_A(D)$ represents interpolated reference and $R_B(D)$ represents the test bit rate curves. D is the level of quality/distortion ratio [22]

$$.\Delta R(D) = \frac{R_B(D) - R_A(D)}{R_A(D)}$$

6. Random Access Methodology

Random access complexity is defined as the number of frames it takes to reach an arbitrarily chosen frame. In the case of light field, random access is calculated the number of frames in the decoding path required to access a random view in a layer of a random frame. Figure 13 below shows a MV-HEVC frame access method [21]. The base frame would be at view 0 at time 0, which would require zero frames to reach. From here, if the randomly accessed frame was determined to be at view 2 at time 1, to reach this frame we would have to first decode three frames (v = 0 at t = 0, v = 2 at t = 0, and v = 2 at t = 2).

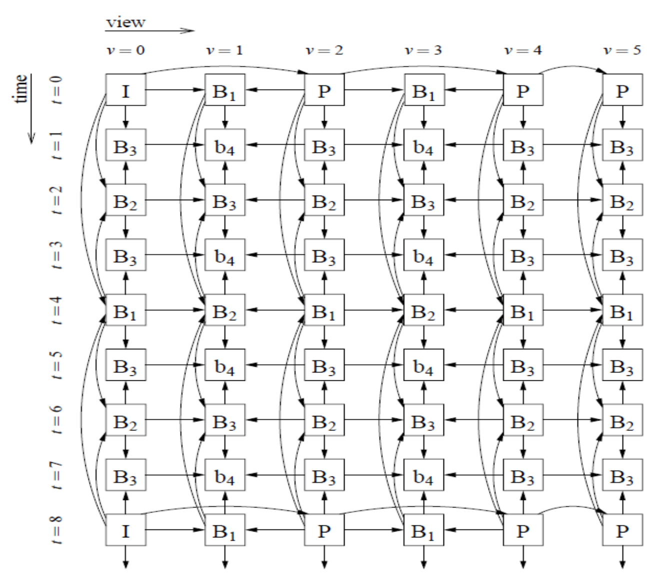

Figure 13 MV-HEVC Frame Accessibility

However, in practicality it is impossible for the decoder to know which frame will be accessed next - the user may decide to skip to a random view at a random location in the video. Thus, random access complexity cannot be calculated based on a specific arbitrary frame and would instead be calculated as the average number of frames it will take to reach any frame in the video. Figure 14 graphs the number of frames decoded to reach all the individual frames of Figure 13 [21]. Thus, the average equation used to calculate the random access complexity for a method is as follows:

$$Random\ Access\ Complexity = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

Where n is the total number of frames (both in view and time) and $f(x_i)$ is the number of frames decoded to reach frame i.

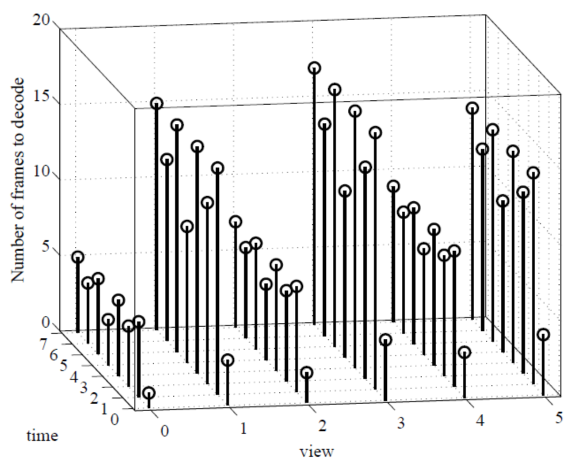

Figure 14 Random Access for Individual Frames

7. Evaluation of Existing Proposed Methods

From Section 3, we know that there is a trade-off between random access efficiency and compression efficiency. For this project, will be evaluating three different light field compression methods and comparing their random access efficiencies as well as compression efficiencies. The methods being compared against one another are 1. LF-MVC method from Section 5.2.1 [20], 2. Joseph's new prediction structure for light field compression [23] and 3. The four-quadrant hierarchical correlation method proposed in [24].

7.1 Performance Metrics

The three methods stated above will be measured based on the same performance metrics. The test data we will be using will be a light field video with 25 lenslet views in a 5x5 array and a total of 48 frames. There will be a total of three test samples, containing a light field video capturing a chess piece, a Lego piece and a train respectively.

Random access complexity of each method will be measured based on the average random access complexity equation in Section 6. The three methods will be compared, and the method with the lowest random access complexity would be determined to be the most efficient method with regards to the accessibility of a random frame in the light field video.

For compression, we will be measuring the results of their bitrate to PSNR curve as described in Section 5.3. We will calculate the result based on four determined quantization values between 25-50, and all methods will be evaluated with the same quantization factor parameters. By analyzing and comparing the three bitrates to PSNR curves on a single graph, we can determine the most efficient compression method by observing the curve which produces the highest quality content for the same bit rate transfer.

After obtaining the results of both random access efficiency and compression efficiency for the methods, it is possible to analyze and conclude on the best overall method. This will be determined by comparing the cost benefits between the random access efficiency and compression efficiency of one method when compared to the other methods. There is no set equation for calculating the cost-benefit ratio as this will depend on the relative difference in scalability between the methods, and consequently will be determined based on analysis of the results.

8. New Proposed Method

In this section, we discuss the possible new methods that might solve the above discussed shortcomings of the existing techniques. Our first proposed method targets exploiting the content captured by the imaging device. One similarity encompassing all three works [20] [23] [24] is the pseudo sequencing of the encoded frames, reflecting the sequence of the lenslet arrangement in the camera. However, we would like to approach the problem in a content aware manner rather than architecture dependent structure. And our second method leverage the intuition that random access will be facilitated by finding a "P" or an "I" frame as soon as possible. And this only happen if "P" and "I" frames are placed in such places where the average distance from other frames are minimal.

8.1 Content Aware Correlation Based Pseudo Sequence

For our first proposal, we intend to calculate the correlation matrix for all the views with respect to each other and choose the most highly correlated view as our I frame. Next, we choose our P frames based on the correlation matrix. We can either threshold the value of the correlation or set the maximum number of P frames. For the rest of the views we repeat this process. That is, we define a hierarchy of frames and label them according to the hierarchy. The following list is a sample hierarchy:

Type of Frame	Number of Frames
I-P	number of P frames
P-P	x number of frames
I-B	y number of frames
P-B	z number of frames
B-B	alpha number of frames

Table 1 Hierarchy of frames

And repeat this process until we achieve IBP frames for all the views. The flowchart (Figure 15) below illustrates the process. As this process will have the bad reconstruction, we intend to change the thresholds and experimentally (subjective tests) find good positions and references for all the views.

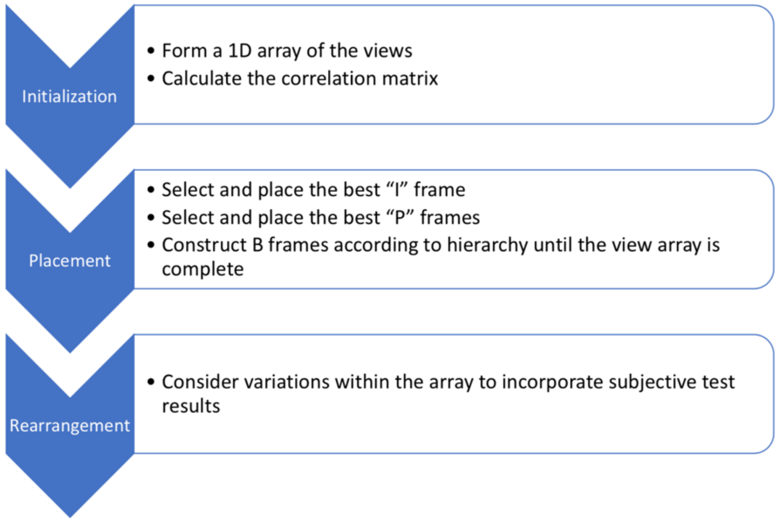

Figure 15 Correlation based pseudo sequence generation.

View 1	View 2	View 3	View 4	... ...	View 25

Figure 16 1D Array of Views

	View 1	View 2	View 3	View 4	View 5	... ...	Score
View 1	-	high	low	low	high		high-low
View 2	high	-	low	high	low		high-low
View 3	high	high	-	high	low		high
View 4	low	low	high	-	low		low
View 5	low	low	high	low	-		low
... ...							

Figure 17 Correlation matrix for each view.

8.1.1 Discussion

All the previous models assumed that the central portion of a light field image has the most similarity with all the other views. Though this assumption has solid foundation in an intuitive manner, we address the problem mathematically. We calculate which frame has the highest similarity using a correlation matrix and advance from there. This model depends on the captured content and the mathematical analysis of their similarity. We need to deliberate on how we will calculate the correlation among the views. We intend to take into account that every view is slightly shifted from the views around it. While placing the "B" frames we would consider their corresponding location in the camera array. Furthermore, while placing the "P" frames we plan to review its would be neighborhood.

8.2 Pseudo Sequence Facilitating Random Access

As the IBP is a hierarchical structure, the depth of the hierarchy correspond to the random access complexity. Reducing the depth of the structure highly compromises the compression efficiency. However, increasing the connectivity among the different levels will find a higher level frame faster. In other words, currently the view array has a vertical and horizontal relationship with its neighbors, introduction of diagonal relations from the higher level frames to lower level frames will ensure faster random access. The following figure illustrates the arrangement:

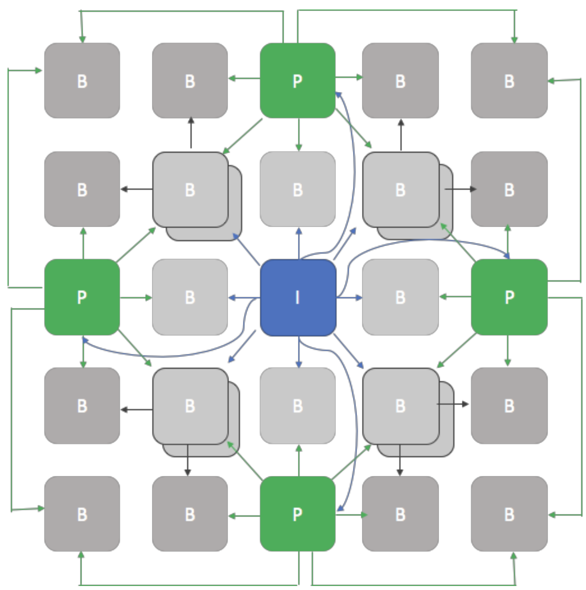

Figure 18 Pseudo sequence using diagonal prediction links

In this case, accessing every frame randomly boils down to decoding at most three frames for a 5x5 view array. Additionally, the amount of reference frames for prediction each frame has to store would be limited to three as well, contributing to compression efficiency. On the other hand, this technique will incur sacrifice in compression efficiency as there will be two "B" frames in some select situation for the same view.

8.2.1 Discussion

Theoretically, this arrangement will increase random access efficiency to a view however, we need to evaluate how much compression sacrifice is required and how much is tolerable in the applications suggested.

Other than the above, one of our major realizations is random access imply view from any point *p(u,v)* of the *uv* plane. And does not necessarily correspond to one of the views from the view array. Refer to the Figure 19 below for illustration. Here, the point *p(u,v)* does not correspond to any view generated by the lenslet array. We can generate this view using interpolation. However, the existing random-access complexity measurement methods do not consider these intermediate views.

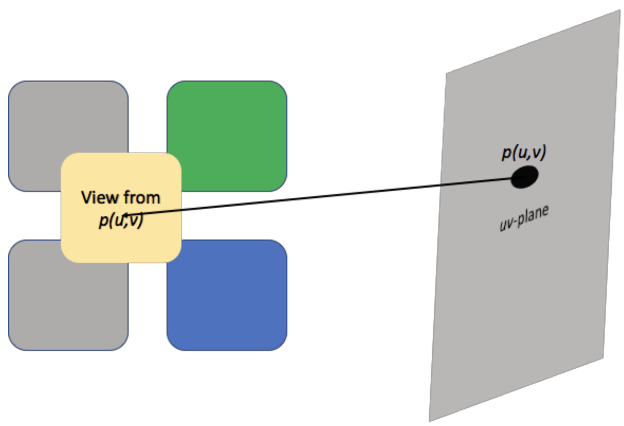

Figure 19 Unique Interpolation View in UV-Plane

9. Conclusion

In this section, we will summarize the proposal report based on what has been discussed and what we plan on completing in the future stage.

We discussed about light video videos and the challenges currently present in data compression and random accessibility. For compression efficiency, we explained methods in multi-view compression and how light field compression advances from its method. Compression efficiency for each method will be calculated based on the bitrate vs. PSNR curve. We also defined random access efficiency as the average number of frames decoded to reach any frame in a light field video. The three existing methods of light field video compression stated in this report are LF-MVC [20], Khoury's method of compression [23] and distance-based reference selection [24]. However, we discovered that all three methods stated assumed that frames beside each other would have the highest correlation, thereby reducing compression efficiency. This may not necessarily be the case, and as a result we proposed to calculate the correlation between all frames to determine the best correlation method for compression. The three stated existing methods did not perform maximization of random access efficiency, and as a result we proposed a method to reduce random access efficiency based on minimizing the distance between frame prediction.

For our next milestone, we will perform two major tasks. For the first part, we will evaluate the compression efficiency and random efficiency for each of the three stated existing methods, then comparing the cost benefits of each method with the others to determine the most efficient method overall. For the second part of our project, we will implement our proposed methods of light field video compression and accessibility. These methods will be calculated for both compression efficiency and random access. Finally, we will compare the results with the three existing methods to improve either current compression efficiency or random access efficiency.

References

[1] U. Fecker and A. Kaup, "H.264/AVC-Compatible Coding of dynamic light fields using transposed picture ordering," in *13th European Signal Processing Conference*, 2005.

[2] K. M. W. Philipp Merkle, "Efficient prediction structure for multiview video coding," *IEEE transactions on circuits and systems for video technology,* vol. 17, no. 11, 2017.

[3] J. M. Franke, "Field-widened pinhole camera," *Applied Optics,* vol. 18, p. 2913, 1979.

[4] S. H. Gage and H. P. Gage, Optic projection : principles, installation and use of the magic lantern, projection microscope, reflecting lantern, moving picture machine, fully illustrated with plates and with over 400 text-figures, Comstock Pub. Co., , 1914.

[5] Fujifilm Global, "Innovation: FUJIX DS-1P: the World's First Digital Camera," [Online]. Available: www.fujifilm.com/innovation/achievements/ds-1p/. [Accessed 29 10 2018].

[6] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai and I. S. Kweon, "Accurate Depth Map Estimation from a Lenslet Light Field Camera," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[7] R. Kingslake, Lenses in Photography: the Practical Guide to Optics for Photographers, Barnes, 1963.

[8] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen and C. Zhang, "Multiview Imaging and 3DTV," *IEEE Signal Processing Magazine,* vol. 24, no. 6, pp. 10-21, 2007.

[9] M. W. Marien, Photography: a Cultural History, Pearson, 2015.

[10] C. A. Bouman, E. L. Miller and I. Pollak, Computational imaging IV, SPIE, 2006.

[11] U. Brandon M. Smith University of Wisconsin-Madison, L. Zhang, H. Jin and A. Agarwala, "Light field video stabilization," in *IEEE 12th International Conference on Computer Vision*, 2009.

[12] J. M. Brennan, "Digital photo kiosk". United States of America Patent US5587740A, 17 08 1995.

[13] S. Spyropoulos, "3D Light Field Camera Technology," Raytrix, 18 1 2018. [Online]. Available: raytrix.de.

[14] A. Tailor, "Introductory Digital Image Processing: a Remote Sensing Perspective," *Image and Vision Computing,* vol. 4, no. 4, p. 229, 1986.

[15] J. Wang, F. Shi, J. Zhang and Y. Liu, "A New Calibration Model of Camera Lens Distortion," *Pattern Recognition,* vol. 41, no. 2, pp. 607-615, 2008.

[16] J. Taylor, "Pictorial Photography in the First World War," *History of Photography,* vol. 6, no. 2, pp. 119-141, 1982.

[17] B. Wilburn, "High-Performance Imaging with Large Camera Arrays," in *ACM SIGGRAPH 2006 Courses on SIGGRAPH '06*, 2006.

[18] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai and Y. Liu, "Light field Image processing: an overview," *IEEE Journal of selected topics in signal processing,* vol. 11, no. 7, 2017.

[19] P. Merkle, K. Muller, A. Smolic and T. Wiegand, "Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC," in *IEEE International Conference on Multimedia and Expo*, 2006.

[20] G. Wang, W. Xiang, M. Pickering and C. W. Chen, "Light field multi-view video coding with two directional parallel inter-view prediction," *IEEE Transaction on Image processing,* vol. 25, no. 11, 2016.

[21] U. Fecker and A. Kaup, "Complexity evaluation of random access to coded multi-view video data," in *15th European Signal Processing Conference (EUSIPCO 2007)*, 2007.

[22] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J.-R. Ohm and G. J. Sullivan, "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,* vol. 26, no. 1, 2016.

[23] J. Khoury, M. T. Pourazad and P. Nasiopoulos, "A New Prediction Structure for Efficient MV-HEVC based Light Field Video Compression," 2018.

[24] L. Li, Z. Li, B. Li, D. Liu and H. Li, "Pseudo sequence based 2-D hierarchical coding structure for light-field image compression," in *Data Compression Conference*, 2017.