FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Pedro de Medeiros Gomes

Pseudo-Sequence Light Field Image Representation and Encoding for Improved Random Access and Scalability

Dissertation supervised by Professor Doctor Luís Alberto da Silva Cruz and submitted to the Electrical and Computer Engineering Department of the Faculty of Science and Technology of the University of Coimbra, in partial fulfillment of the requirements for the Degree of Master in Electrical and Computer Engineering, branch of Computation.

Supervisor:
Prof. Dr. Luís Alberto da Silva Cruz

Coimbra, September 2019

Acknowledgments

I would like to thank my supervisor, Doutor Luís Alberto da Silva Cruz, for his guidance and advice throughout the development of this dissertation. I would also like to thank all the support I got from family and friends over this period. Finally, I would like to thank Telecommunications Institute for the all the resources made available, without which this dissertation could not have been completed.

Abstract

Light field images allow to capture richer visual information from our world. However, these images have large size which increase significantly storage, transmission, and processing costs. Therefore, new strategies to compress light field images are relevant and timely. This dissertation devises a new method to efficiently compress light field images. The compression method was designed to maximize the compression efficiency but also able to provide random access and scalability.

The proposed compressing method is based on pseudo-sequence coding using the Multi-View extension of HEVC . The method first decomposes the light field image into sub-apertures, which are subsequently used by a novel scan order, denoted Quadratic Spiral, to form multiple pseudo-sequences to be encoded by an MV-HEVC encoder. In the encoder, the pseudo-sequences are encoded according to a coding structure. The proposed coding structure was specifically designed to improve the random access performance, while still exploiting the temporal and inter-view correlations. Furthermore, using the layered structure of the MV-HEVC, a scalable representation of the light field can be decoded. The scalable representation is composed of periodic patterns of sub-apertures, the patterns are generated by the way the scan order arranges the sub-apertures in pseudo-sequences.

In order to evaluate the developed compression method, the test conditions defined by JPEG Pleno Group were followed. The experimental results demonstrate that the Quadratic Spiral scan order provides improved random access performance. It was concluded that the proposed method is an efficient compression method for light fields, as it achieves an average PSNR gain of 7.5 dB over reference JPEG, which is in accordance with the best current state-of-art methods.

An optimization based on dynamically adjusting the quantization parameter (QP) individually for each frame was performed. Heuristic methods were used to find the best quantization parameter (QP) values of each frame. It was demonstrated that the QP optimization improves significantly the proposed compression method, as bitrate savings up to 22%, were obtained.

Resumo

Imagens *Light Field* contem uma representação visual mais rica do mundo. No entanto, o tamanho destas imagens acarreta custos significativos de armazenamento, transmissão e processamento. Por isso é actualmente necessário desenvolver novos métodos de compressão para imagens *Light Field*. Nesta dissertação propõe-se um novo método para comprimir eficazmente imagens *Light Field*. O novo método de compressão foi desenhado não só para comprimir eficientemente, mas também para melhorar o acesso aleatório e permitir escalabilidade.

O método de compressão proposto é baseado em codificação de pseudo-sequências usando a extensão Multi-View do HEVC. O método primeiro decompõe a imagem *Light Field* em múltiplas vistas. Estas vistas são de seguida reordenadas, por ordem uma ordem de digitalização desenvolvida nesta dissertação ao qual chamamos *Quadratic Sprial*, onde várias pseudo-sequências são enviadas para um codificador de vídeo MV-HEVC. No codificador as pseudo-sequências são codificadas de acordo com uma estrutura de codificação. A estrutura de codificação foi desenhada especificamente para melhorar o acesso aleatório, sem deixar de explorar a correlação entre as vistas. O método de compressão desenvolvido permite também utilizar a estrutura em camadas do MV-HEVC, para descodificar imagens *Light Field* de forma escalável. No descodificador a imagem é representada de forma escalável, onde cada representação é composta por padrões periódicos de vistas. Estes padrões são gerados pela forma como a ordem de digitalização organiza as vistas em pseudo-sequências.

Para avaliar o método de compressão desenvolvido, foram seguidas as condições de teste definidas pelo grupo JPEG Pleno. Os resultados experimentais obtidos demonstram que a ordem de digitalização *Quadratic Spiral* melhora o acesso aleatório. Foi também concluindo que o método desenvolvido comprime eficientemente imagens *Light Field* conseguindo alcançar um ganho médio PSNR de 7,5 dB comparado com JPEG de referência, o que está de acordo com melhores métodos de compressão actuais.

Para finalizar, o método de compressão foi melhorado ajustando dinamicamente o parâmetro de quantização (QP) individualmente para cada trama. Para encontrar os melhores valores de QP para cada trama foram utilizados métodos heurísticos. Esta optimização do QP melhorou significativamente o método de compressão desenvolvido, permitindo comprimir uma imagem *Light Field* com menos 22% bits comparado com o método sem melhoramento.

List of Figures

List of Tables

Contents

Glossary

LF	Light Field
EI	Elemental Image
JPEG	Joint Photographic Experts Group
HDCA	High Density Camera Array
DCT	Discrete Cosine Transform
HEVC	High Efficiency Video Coding
JCT-VC	Joint Collaborative Team on Video Coding
MV	Motion Vector
CTU	Coding Tree Unit
CTB	Coding Tree Block
CU	Coding Unit
PU	Prediction Unit
TB	Transform Block
Qstep	Quantization Step
QP	Quantization Parameter
CABAC	Context-Based Adaptive Binary Arithmetic Coding
GOP	Group of Pictures
POC	Picture Order Count
MV-HEVC	Multi View extension of High Efficiency Video Coding
QSpiral	Quadratic Spiral
Q-Serpentine	Quadratic Serpentine
QuadLines	Quadrants Lines
PSpiral	Pure Spiral
HDCI	Heidelberg Collaboratory for Image Procesing
PSNR	Peak signal-to-noise ratio
CD	Compact Disc
I-Picture	A picture coded using intra prediction
P-Picture	A picture coded from uni-directional prediction.

1

Introduction

The leading topic of research of this M.Sc. dissertation is the efficient compression of light field images with a special focus on random access and scalability. This chapter introduces the topic of light field images and the motivation behind the proposed work, followed by a summary breakdown of its main objectives and contributions.

1.1 Motivation

A new form of photography light field (LF) images describes the world with considerably more information than conventional 2D digital images. An LF image contains not only the information about the intensity of light in a scene but also the direction of the light rays in space. Therefore an LF image is an image that contains both spatial and angular information of a scene. Light Field images are becoming increasingly popular in computer vision and computer graphics. This is because the higher dimensional representation of the visual data offers powerful capabilities for scene understanding, and substantially improves the performance of traditional computer vision functionalities such as depth sensing, post-capture refocusing, segmentation, video stabilization, material classification, and many more.

However, while offering unprecedented opportunities for advanced image applications, the high dimension of LF images also brings up new challenges. The large amount of data present in the LF image produces a significant demand in storage capacity and increased transmission costs. These problems call for the development of efficient LF compression schemes. Another challenge that must be considered is that the full LF data are not always necessary, some applications often only need a partial light field or selected views to function properly. Therefore good random access performance and a scalable representation of light field that allows fast transmission and decoding from a base layer are highly desired, particularly in low resource systems.

A light field compression method is presented in this dissertation. The proposed method will decompose the LF image into multiple sub-apertures images. A scan order then rearranges the sub-apertures into multiple pseudo-sequences to be encoded by a multi-view video encoder. In the encoder, a coding structure is used to exploit the correlation between sub-apertures using the combined temporal and inter-layer prediction, provided by multi-view coding. The proposed scan

order and coding structure are designed for improved random access performance, as well as to provide a scalable representation of the light field in periodic patterns of sub-aperture images.

An additional investigation was carried out to study the effects of dynamically adjusting the quantization parameter for each frame. The objective of this investigation is to select the best quantization parameter value for each picture that optimizes the efficiency of the proposed compression method.

1.2 Objectives

The main objective of this dissertation is to develop a compression method with the following features:

- High compression efficiency.
- Good random access performance.
- Scalability

A secondary objective is to improve the efficiency of the compression method by adjusting the quantization parameter of each frame.

1.3 Main Contributions

The contributions of the work developed in this dissertation are the following ones: A novel scan order to rearrange sub-apertures in a pseudo-sequence that exploits the correlation between neighbor sub-apertures. A coding structure to efficiently assign a prediction structure within the multiple pseudo-sequences that provides improved random access as well as a scalable representation of light fields; An optimization study in the individual choice of the quantization parameter for each frame.

In summary, the major contributions of the proposed compression method are:

- A novel scan order (Quadratic Spiral).
- A coding structure optimized for random access and scalability
- Optimization based on the selective choice of the quantization parameter.

Part of the results obtained in this dissertation were used for the elaboration on a paper [7] in 8-*th* European Workshop on Visual Information Processing 2019.

1.4 Document Structure

This dissertation is organized as follows: Chapter 1 is an introduction to the developed work, contains the main objectives and contributions. Chapters 2 and 3 constitute a state-of-the-art and literature review. Chapter 2 describes the concepts in light field imagining, including the theory, possible representations, acquisition methods, applications, and challenges. Chapter 3 includes a description of the High Efficient Video Coding standard and his Multi-view extension, presents a compression strategy of LF using video encoders as well as a discussion of previous compression methods based on this strategy. The proposed compression method is presented in Chapter 4. Chapter 4 discuss the possible scan orders and coding structures as well as a justification for design choices for improved performance and scalability. The experimental setup is described in Chapter 5. Chapter 6 presents and analyses the results obtained with the proposed compression method and reference methods. In chapter 7 our compression method is improved based on quantization parameter optimization, the optimization scheme and results and presented. Chapter 8 concludes the dissertation. Extra results and scripts to replicate the developed work were made available in the annexes.

2

Light Field Images

In this chapter, a brief overview of light field imaging is presented. This chapter describes the concepts in light field imaging, including the theory, possible representations, acquisition methods, applications, and challenges.

2.1 Light Field Theory

We perceive the world around us by capturing light rays with our eyes. These light rays carry the information of our surrounding 3D environment. The plenoptic functions [8] describe the light rays passing through every point in space, from every possible direction, for every wavelength at any point in time. Any observable space can be fully represented by the seven-dimensional plenoptic function.

$$\textit{Plenoptic Function} = f(x,y,z,\theta,\phi,\lambda,t)$$

(x,y,z) = 3D space coordinates
(θ,ϕ) = direction and orientation of light rays
(λ,t) = wavelength and time.

Light field capturing is about sampling the 7D plenoptic function. However, handling such high dimensional data can be difficult. For this reason, the 7D plenoptic function has been simplified into a 4D representation for practical use denoted 4D light field in [9]. The simplification is made by assuming a static scene and that the wavelength λ is independently recorded in different color channels (RGB). Those assumptions allow us to remove the time t and wavelength λ dimensions, reducing the plenoptic function from seven to five dimensions. However, this 5D representation still contains some redundancy. Levoy and Hanrahan [9] and Gorttler et al [10], by assuming that the light rays were measured in free space, further simplified the function to four dimensions. In free space, the light rays propagate in a fully transparent space and will follow a linear trajectory without deviations or loss of intensity, therefore radiance does not change along a line. Based on this assumption LF can be parameterized in terms of space and direction.

There are several ways to parameterize a 4D light field, the most common is to parameterize

the light rays by the coordinates of their intersections with two parallel planes placed at arbitrary positions as shown in Figure 2.1. Defining the first plane by the coordinates (u,v) and the second plane by the coordinates (s,t), a linear light ray that intersects the first plane at the point u,v and the second plane at s,t can be represent by the 4D function $L(u,v,s,t)$. In summary, a light field image is an image that contains both spatial and angular information of a scene.

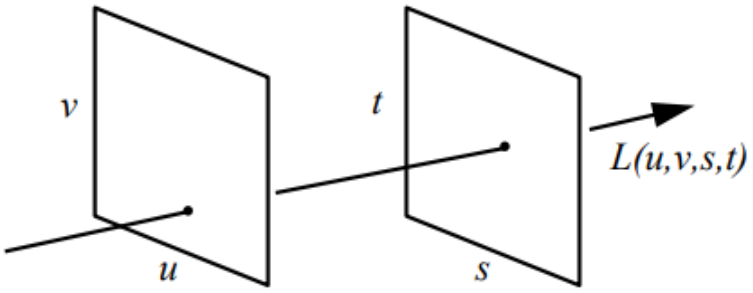


Figure 2.1: Parametrization of the plenoptic function in 4D.

2.2 Light Field Visualization

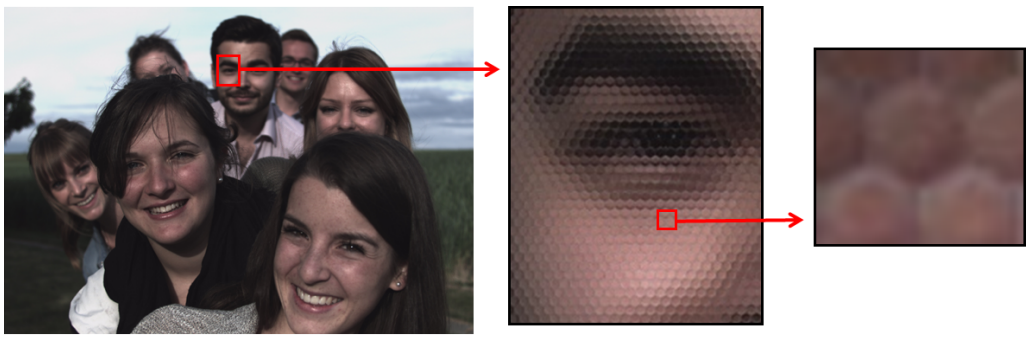
In order to visualize a 4D light field multiples 2D representations were created. In this section, the possible representations are discussed.

In the $L(u,v,s,t)$ light field model, if we consider the *st* plane as a set of cameras focused on the *uv* plane, each camera captures the light rays leaving the *uv* plane and striking a point on the *st* plane. Thinking this way a light field can be seen as capturing an array of viewpoints. Therefore LF can be represented as a 2D array of images. We call these images as sub-apertures images and each one depicts the scene from a slightly different perspective. An example of a sub-aperture representation of a light field is depicted in figure 2.2.

Figure 2.2: Sub-aperture representation of a Light Field.

Another way to visualize a light field is the lenslet format. This is the native format of plenoptic cameras. Figure 2.3 shows an example of a lenslet image from a plenoptic 1.0 camera without color correction. The raw lenslet image is composed of macropixels and each macropixel is composed of multiple pixels. All the pixels in one macropixel contain the angular information for one point

in the scene. If we extracted a pixel from a specific position form each macropixel the result would be an image correspondent to a specific sub-aperture (view-point). The number of pixels in one macropixel defines the angular resolution (the number of different viewpoints) and the number of macropixels defines the spatial resolution. In the next section, the creation of lenslet images is explained in more detail when describing plenoptic cameras.



(a) (b) (c)

Figure 2.3: Lenslet Format (a) Full lenslet image; (b) Portion of lenslet image; (c) Macropixel

A light field can also have a epipolar representation. This format can be seen as a 2D slice of the 4D light field. A epipolar image can be obtained by collecting light field samples using a fixed spatial and angular coordinate. Contrary to a sub-aperture image the epipolar image contains both spatial and angular dimensions therefore depth information can be easily extracted. For this reason, epipolar images are frequently used in computer vision [11] to infer scene geometry.

(a) Horizontal slice **(b)** Vertical slice

Figure 2.4: Epipolar images

2.3 Light Field Acquisition

A conventional 2D camera sensor can only capture the information from two dimensions of a scene at a single moment. To capture a light field we need to capture multiple images along the angular dimensions. Currently, the most common ways to capture a light field are: plenoptic

cameras; multiple cameras; moving cameras;

Plenoptic Cameras: Plenoptic cameras use a micro-lens array between the objective and the photosensor, the micro-lens separate the light rays striking each micro-lens into a small image on the photosensor pixels, this allows to capture a 4D light field using a single 2D sensor. Two optical designs were developed for plenoptic cameras, depicted in Figure 2.5. The ”Plenoptic 1.0” design [12] also called unfocused camera has the micro-lens array at a focal distance of the objective, in which the scene is focused on the micro-lens, and the ”Plenoptic 2.0” [13] also called focused camera, in which the image plane of the objective is the object plane of micro-lens arrays, this way a focused image appears on the photosensor.

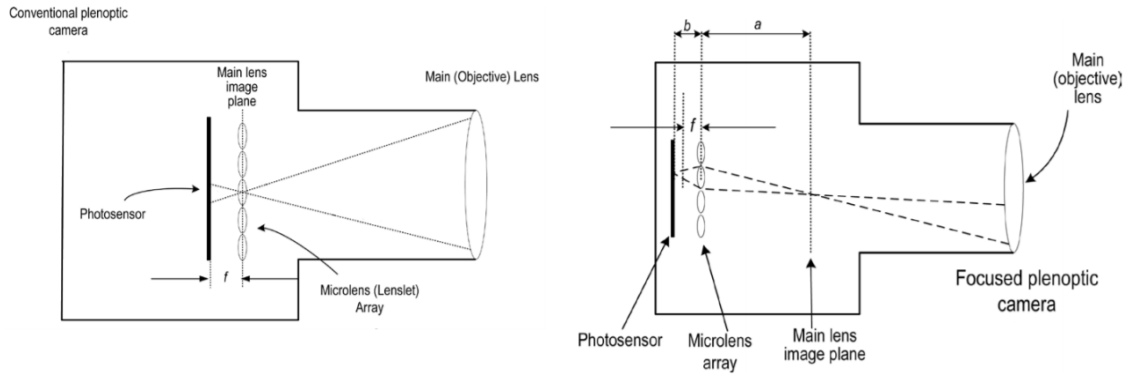


Figure 2.5: Scheme of plenoptic 1.0 (left) and plenoptic 2.0 (right), figure adapted from [1].

Both cameras produce different captures of the same scene as shown in figure 2.6. The raw data captured by plenoptic cameras is a lenslet image composed of micro-images also called Elemental Images (EI). Each EI matches the set of pixies behind each micro-lens.

In the ”Plenoptic 1.0 ” the light emitted by one point in the 3D scene is spread over several pixel sensors of the raw lenslet data. A micro-image corresponds to the macropixel in the lenslet format from the previous section. Every pixel in a micro-image corresponds to a different view-point,
In the ”Plenoptic 2.0” version each micro-image captures a part of the scene. A certain point from a micro-image is visible across multiple micro-image from slightly different perspectives. The final result is multiple micro-images with overlapping areas [2].

Multiple Cameras: Also know as High-Density Camera Arrays (HDCA), an LF HDCA is composed of multiple cameras mounted at regular intervals all aimed at the same point. This setup can capture a light field. Figure 2.7a shows the HDCA build by Stanford University.

Moving Cameras: By mounting a digital camera on a movable gantry one can acquire light fields, by capturing multiples images of the same scene from different positions with the same camera (Figure 2.7b).

(a) plenoptic 1.0.

(b) plenoptic 2.0.

Figure 2.6: Raw image (lenslet image) capture by different plenoptic cameras, image adapted from [2].

(a) The Stanford cameras array setup [14].

(b) Gantry used on The Digital Michelangelo Project [15].

Figure 2.7: HDCA images.

2.4 Light Field Applications.

Technological advancements in the last decades enabled plenoptic cameras to become more portable and inexpensive. This lead to a growth in the popularity of light field images. Nowadays light field cameras are commercially available, for example, Lyrto [16] and Raytrix [17] have introduced plenoptic cameras in the market and recently they also began being integrated into smartphones [18].

The widespread use of light field camera lead to the appearance of new applications and are today present in state-of-art technology from medicine [19], microscopy [20] to autonomous vehicles [21] among others. Those applications exploit the large volume of data from both the spatial and angular information of a scene to support a range of post-capture functionalities such as:

- Refocus
- Depth Estimation
- Different view points visualization.
- Focus Stacking
- Super-Resolution
- 3D Reconstruction

2.5 Light Field Challenges

However, the functionalities provided by light field images come with a cost. The high volume of LF images produces a significant demand in terms of storage capacity as well as communications costs. Additionally, the high complexity of LF data requires extra computational power for processing. For these reasons, efficient compression schemes are highly required.

Since an LF image exhibits repetitive patterns and a large amount of redundant information, in theory, one would expect LF images to be easy to compress. However, in reality, due to the high dimension of the signal, it can be hard to take benefit from the redundancy. Early solutions based on conventional image encoders such as JPEG achieved poor results [10]. The reason behind this is that conventional image encoders were developed for natural two-dimensional images. In natural images, the discrete cosine transform (DCT) coefficients of the pixels follow a Laplace distribution, so image coding standards typically rely on the DCT for compression. In contrast, for lenslet images, the pixels coefficients distribution does not follow the typical distribution of natural images, which leads to low compression efficiency by conventional image encoders. Solutions based on predictive coding using video encoders have thus been considered to further improve compression efficiency [22].

Another challenge that must be considered is that the full light field data is not always needed. Some applications may only require a partial light field to function properly. For example, only a single row of sub-apertures is necessary to form a epipolar image and extract depth information of one dimension. For that reason, easy access to selected sub-apertures can be highly advantageous, and schemes with good random access performance are desired. Additionally transmitting a full light field image will most likely cause transmission latency and increased network costs. Therefore a scalable representation of light field that performs quick transmission and decoding from a base layer is desired, especially in low resource scenarios.

Due to these challenges, as mention in the Objectives section (sec.1.2), the objectives in this dissertation are to 1) efficiently compress LF images 2) have good random access performance 3) provide a scalable representation of the LF.

3

High Efficiency Video Coding

High Efficiency Video Coding (HEVC), also known as H.265, is a video compression standard defined by Joint Collaborative Team on Video Coding (JCT-VC), a collaboration between the ISO/IEC MPEG and ITU-T VCEG [23]. The HEVC video coding layer design is based on conventional block-based motion compensated hybrid video coding concepts, which consist of inter/intra-picture prediction and 2D transform coding. This chapter provides a general overview of the HEVC standard, a presentation of the Multi-View extension for HEVC and previous works on light field compression using video encoders.

3.1 General Architecture

An HEVC video encoder is designed to find redundant and visually irrelevant information in the video that can be discarded, reducing the file size. The way it works is by comparing different parts of a frame with other parts within the same frame or from other frames and find redundant areas. These redundant areas are then replaced with a short description instead of the original pixels. The information with low visual impact that can be removed without significant loss of the scene perception can also be discarded according to the level of compression.

A typical video HEVC encoder would start by splitting every picture (frames) of the video sequence, into block shaped regions. The first picture (or the first picture at each clean random access point) is coded using intra-picture prediction. Intra-picture prediction is when the prediction of the blocks in the picture is based only on the information in that picture. For all other pictures, inter-picture prediction is used. The encoding process for inter-picture prediction consists of searching and choosing the prediction information from other previously encoded pictures. The residual signal of the intra- or inter-picture prediction, which is the difference between the original block and its prediction, is transformed by a linear spatial transform. The transform coefficients are then scaled, quantized, entropy coded, and transmitted together with the prediction information. After the prediction methods are finished and the picture goes through the loop filters, the final picture representation is stored in the decoded picture buffer. Pictures stored in the decoded picture buffer can be used for the prediction of other pictures. A typical HEVC architecture is depicted in figure 3.1. In the following subsections, the several steps to encode a video using HEVC are explained

in more detail.

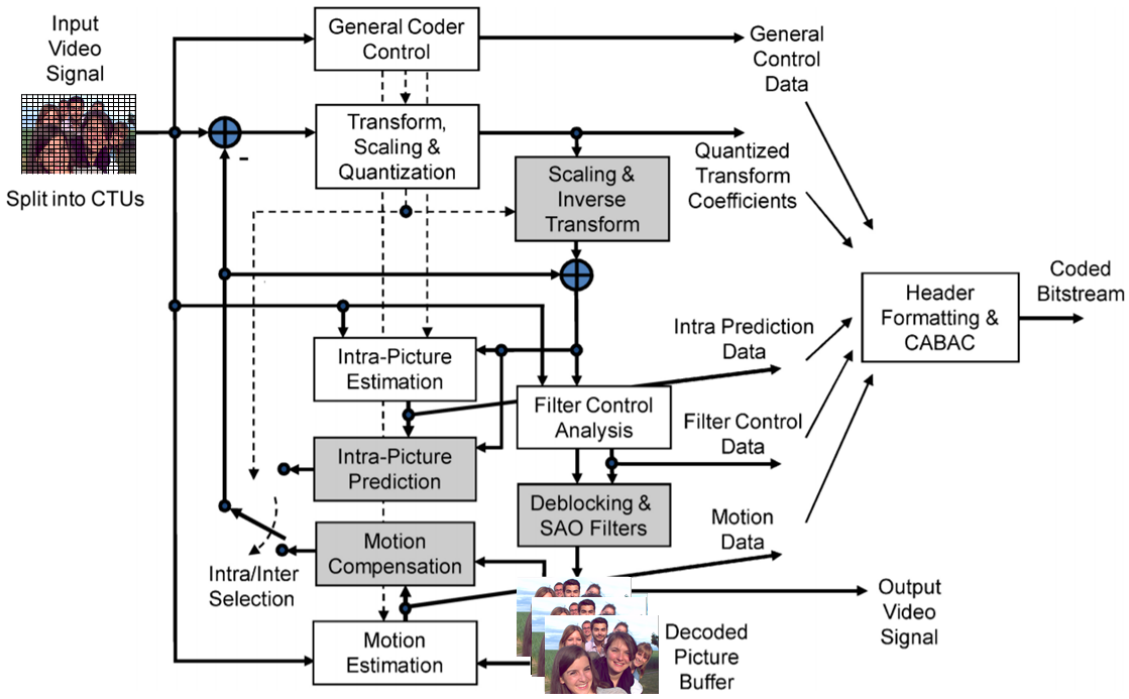


Figure 3.1: Typical HEVC Architecture, figure adapted from [3].

3.1.1 Sampled Representation of Pictures

HEVC supports multiple color encoding representations [24], but as a rule the YCbCr color space with 4:2:0 sampling is typically used. The Y component is denoted luma and represents brightness. The two components Cb and Cr are the chroma components and represent the extent to which the color deviates from gray toward blue and red, respectively. The YCbCr space is preferentially used because the human visual system is more sensitive to variations of luma than chroma, this allows the YCbCr space to be sampled to 4:2:0 without the user perceiving any difference. Using YCbCr 4:2:0 each chroma component has one-fourth of the number of samples of the luma component (half the number of samples in both the horizontal and vertical dimensions), and each component is typically represented with 8 bits.

3.1.2 Division of Frames in Coding Tree Units

While the previous video encoder standards each picture was divided in fixed macroblock units consisting of 16x16 luma samples and, in the usual case of 4:2:0 color sampling, two corresponding 8×8 blocks of chroma samples, the HEVC standard, on the other hand, uses Coding Tree Units (CTU) with are composed of Coding Tree Blocks (CTB) for each luma and chroma component. The size of the CTBs is defined by the encoder and can be larger than a traditional macroblock,

allowing 16x16, 32x32 and 64x64 partition, with the larger sizes typically enabling better compression. CTBs can then be divided into smaller coding units (CU), by also using a tree structure, eventually resulting in four smaller regions. Such quadtree splitting process can then be iterated until the coding block reaches the minimum allowed size defined by the encoder configuration [25]. Figure 3.2 shows an example of a subdivision of a picture in a CTU.

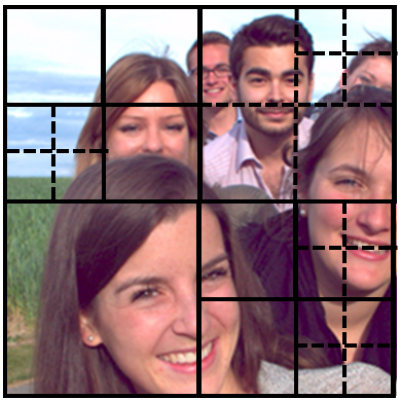


Figure 3.2: Subdivision of a picture in a CTU.

3.1.3 Intra-Prediction

Intra prediction is used to remove correlation within local regions of a picture. The basic assumption for intra prediction is that the texture of a picture region is similar to the texture in the local neighborhood and can thus be predicted from there. Therefore in intra picture prediction, the information of adjacent CTBs from the same picture is used for spatial prediction of the current prediction unit (PU). The values of the available neighboring samples are combined to form a prediction signal [3]. There are a total of 35 intra picture prediction modes available in HEVC, corresponding to 33 different directional modes, a DC and a planar mode. Figure 3.3 shows an example of four intra prediction cases in different directional modes. DC mode encoding simply uses a single value matching the mean value of boundary samples for the prediction. Finally, the planar mode assumes an amplitude surface with a horizontal and vertical slope derived from the boundaries.

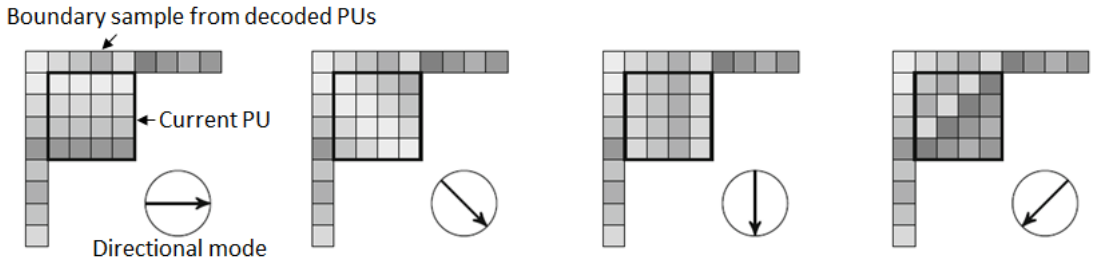


Figure 3.3: Examples of four directional intra prediction modes of the current block from neighboring samples, figure adapted from [3].

3.1.4 Inter-Prediction

Usually in a video sequence only small differences of the scene content can be observed from frame to frame, and these differences are mostly due to motion. This is the basic assumption behind inter-prediction, that a significant part of the content of the pictures in a video sequence consists of moving objects. The goal is to use this motion for prediction. Pictures are predicted from previously encoded frames (uni-directional) or from previous and future frames (bi-directional), as shown in figure 3.4. The bidirectional prediction uses frames from the future, this requires the video frames to be coded and stored out of order.

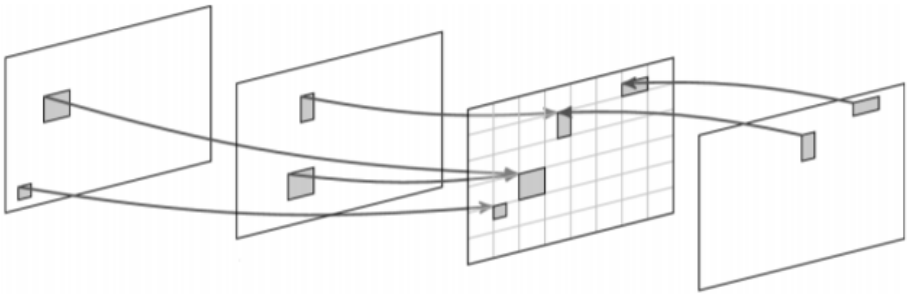


Figure 3.4: Example of uni and bi-directional inter prediction, figure adapted from [3].

The encoder searches for blocks similar to the one it is encoding on a previous/future encoded frame denoted as a reference frame. Such search procedure is known as motion estimation, resulting in the identification of a motion vector, which points to the position of the best prediction block in the reference frame. However, since the identified block will most likely not be an exact match of the encoding block, the resulting difference (residue) has to be encoded and transmitted to the decoding end, so that it can be read by the decoder, the residue is also known as prediction error. This way the scene can be efficiently represented by motion vectors and a prediction error signal. Where the motion vectors indicate how picture regions should be moved from the reference picture to the current picture to form the prediction and prediction error signal contains the part of the scene content that could not be described by the applied motion model [3][23].

3.1.5 Transform and Quantization

The residual signal contains the part of the original signal which could not be predicted by the selected prediction methods previously described. This signal is further compressed by the HEVC using transform coding and then quantization, as depicted in figure 3.5. To perform transform coding each CTB is partitioned in multiple transform blocks (TB) of size 4x4, 8x8, 16x16 or 32x32. Two-dimensional transforms are computed by applying 1-D transforms in the horizontal and vertical directions. The elements of the transform matrix were derived from integer basis functions based on the DCT [23]. This results in a representation of the correlated parts of the residual signal in the residual block by a potentially small number of transform coefficients. The transform coefficients are then quantized, before being coded into the bitstream.

Figure 3.5: Example of the process of transform and quantization, figure adapted from [4].

The quantization process is a compression technique that maps signal amplitudes to a predefined set of representative values. This is done by dividing the resulting block of coefficients by a quantization matrix and a subsequent rounding of each element. An element of the quantization matrix is denoted quantization step (QStep) and indicates the step size is going to be for that corresponding element position in the DCT matrix (our block after transform coding) [4]. The actual formula for quantization is displayed below.

$$Quantized\ Value(i,j) = round\Big(\frac{DCT(i,j)}{QStep(i,j)}\Big)$$

The quantization matrix is designed to provide more resolution to more perceivable frequency components over less perceivable components (usually lower frequencies over high frequencies) in addition to transforming many components to 0, which can be encoded with greater efficiency. Therefore the coefficients that are most significant to the compressed rendition of the image are encoded with small step size, while less important coefficients are encoded with larger step sizes. The quantization step is controlled by a quantization parameter (QP), the range of values the QP can assume is defined from 0 to 51. The QP and Qstep have a logarithmic relationship as represent in figure 3.6. An increase of 1 in QP means an increase of the quantization step size by approximately 12%. The Qstep value regulates how much spatial detail is saved. When Qstep is very small, almost all that detail is retained. As QP is increased, some of that detail is aggregated so that the bitrate is lower but it comes at the price of some increase in distortion and loss of quality.

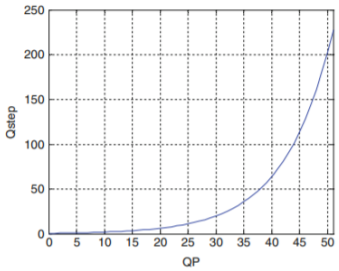


Figure 3.6: Relationship between quantization step size (Qstep) and quantization parameter (QP), image from [3].

3.1.6 Entropy Coding

In this module all the syntax elements which represent the quantized transform coefficients, motion vectors, intra-prediction directions are coded into a bitstream output. The input data is first converted to binary symbols, i.e. into 0 and 1. This is done to reduce complexity and allow for probability modeling for more frequently used bits of any symbol. Then, an entropy coding technique is applied to compress the data and originate the output coded bitstream. HEVC specifies only one entropy coding method, the Context-Based Adaptive Binary Arithmetic Coding (CABAC) [24].

3.1.7 Temporal Prediction and Coding Structures

The coding structures define the prediction scheme and the organization of the pictures in a coded video sequence in terms of coding order and output order. A coding structure is composed of a set of pictures denoted Group pf Pictures (GOP). Each picture is uniquely identified by the picture order count (POC). The POC represents the order in which the pictures have been generated and in which should be output for display. However, the pictures can be encoded out of the temporal order in which have been recorded. This is the decoding order and specifies the order in which pictures are reconstructed at the decoder, and thereby defines which pictures may be used for reference. Naturally only the previously decoded picture can be used as a reference for prediction.

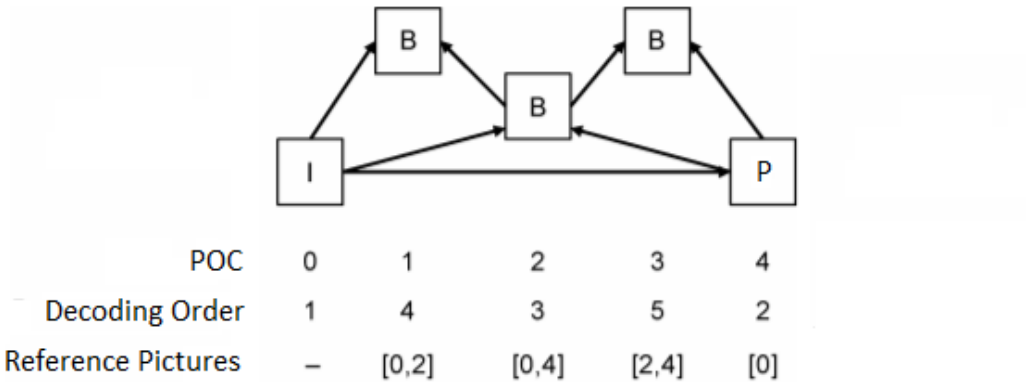


Figure 3.7: Coding structure example.

The prediction scheme defines which pictures should be taken as a reference during the encoding, it determines the dependencies between all the pictures within a GOP. The full video sequence is usually represented by a periodic repetition of the coding structure. Figure 3.7 shows an example of a coding structure. In the proposed method chapter a unique coding structure will be defined.

3.2 Multi-view Extension

The Multi-View extension of HEVC (MV-HEVC) [26] allows to efficiently encode multiple multiple video sequences simultaneously. This extension was designed for systems with multiple

cameras recording the same scene at the same time, but from different locations. Where each video sequence would correspond to a different viewpoint of the scene. The next logical step would be then to encode the different video sequences in the same encoder in order to exploit the high correlation between them. In this document, the video sequences used as input for an MV-HEVC encoder are denoted video views.

In an MV-HEVC encoder, each video view is encoded in a different layer. It is possible while encoding frames form different video views, on the same temporal order, to use shared information across layers [27]. This is a frame from a certain video view can be predicted not only from temporally related frames from the same video view, but also from frames of different video views. An example coding structure for multiple video views is shown in figure 3.8

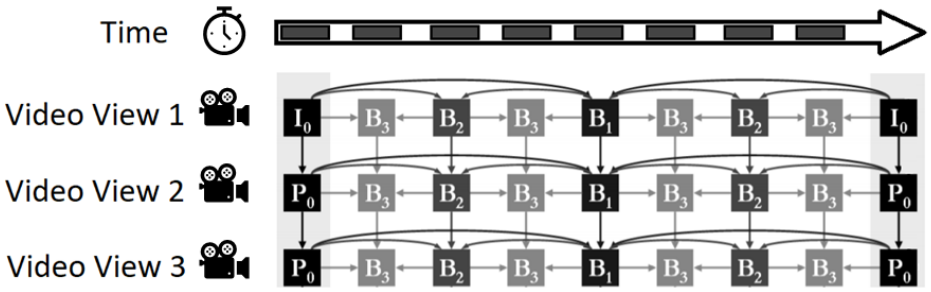


Figure 3.8: Multi-layer coding structure example.

The layered structure of the MV-HEVC supports a scalable bit-stream that allows access to selected video views with minimum decoding effort. The level of access depends on the coding structure, for example in the coding structure from figure 3.8, the first video view can be individually decoded. On the other hand, to decode the third video view the first and second video views need to be decoded beforehand.

3.3 Pseudo-Sequence Light Field Compression

As mentioned in the previous chapter efficient compression is a critical aspect for the practical use of light field images. Since early solutions based on conventional image encoders achieved poor results, solutions based on predictive coding using video encoders such as HEVC have then been considered to further improve compression efficiency. These techniques were classified as pseudo-sequence based coding [28]. In this section, the state-of-art of pseudo-sequence light field compression techniques is presented.

The pseudo-sequence coding strategy is to interpret the light field image as a video sequence. First, the 4D light field image is represented as multiples sub-apertures images. Next, the sub-apertures are rearranged as a pseudo-sequence, where each sub-aperture is interpreted as a video frame. The pseudo-sequence is then sent to a video encoder such as HEVC, to be compressed. Each sub-aperture represented the scene form a slightly different viewpoint, thus consequently have high redundancy in booth spatial and angular dimensions [10]. Since a video encoder works

by exploiting redundancy between frames, a video encoder can be used to efficiently exploit the redundancy between sub-apertures.

Several studies have aimed to find the best scan order to rearrange the sub-apertures in pseudo-sequences. Dai et al. [5] used a zig-zag and a spiral scan order to generate a single pseudo-sequence, as depicted in figure 3.9. They concluded that the spiral scan order outperforms the zig-zag scan order because of a smoother rearrangement of the sub-apertures. Raster, serpentine, zigzag and circular scan orders are compared in [29].

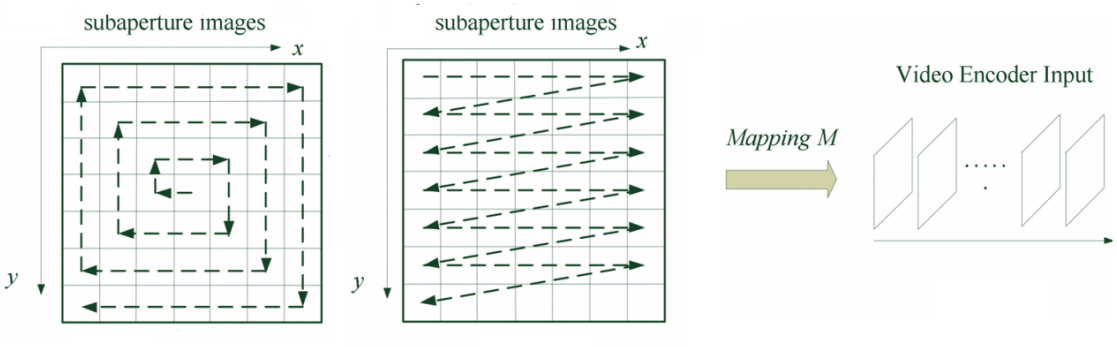


Figure 3.9: Spiral and zig-zag scan orders, figure adapted from [5]

Liu et al. [30] proposed a hierarchical coding structure, where the central sub-aperture is encoded as Intra, the remaining sub-apertures chose a sub-aperture at each if the top, bottom, left and right directions from the already encoded. In this way, each sub-apertures have at most four references. The scheme compared conventional image encoder JPEG [31] achieved a gain in image quality of 4.5 dB.

Li et al. [32] proposed a using a 2D hierarchical coding structure. The sub-apertures are divided into four quadrants and encoded a quadrant after another to reduce the buffer size as much as possible. Inside each quadrant, all the sub-apertures are encoded hierarchically in both horizontal and vertical directions. The scheme achieves an average gain of 4.631 dB over JPEG.

Perra and Assunção [33] propose the partitioning of the raw light field image (lenslet format) into tiles of equal size and then these tiles are ordered using a serpentine scan order as pseudo-sequence to be encoded by HEVC. This approach obtained lower compression efficiency compared to the previously described methods with a gain of 4 dB compared to the JPEG compression scheme

While the previous works mentioned here only consider a single layer pseudo sequence, more recent compression methods have adapted the multi-view extension for light field compression. Ahmad et al [34] interpreted each row of sub-apertures as a pseudo sequence. The video views are then compressed by MV-HEVC using a two dimensional weighted prediction. This method obtained an average PSNR gain of 7.7 dB over JPEG.

4

Proposed Compression Method

In this chapter, a light field compression method based on the multi-view extension of the HEVC standard is proposed. The proposed method consists of a novel scan order combined with a coding structure to improve compression efficiency and random access. This chapter includes a presentation on possible scan orders and coding structures as well as a justification for the design choices for improved performance and scalability.

4.1 General Description of the Method

The proposed method first decomposes the light field image into multiple sub-aperture images. These sub-apertures are then organized in 2-D structures as defined in section 2.2. A novel scan order is used to generate multiples pseudo-sequences to be encoded as video-views by MV-HEVC encoder as depicted in Fig. 4.1. Since neighboring sub-aperture images differ by a slightly perspective shift when the scan order forms the pseudo-sequences the main goal is to use the combined temporal and inter-layer prediction, provided by multi-view coding, to better exploit the correlation between them.

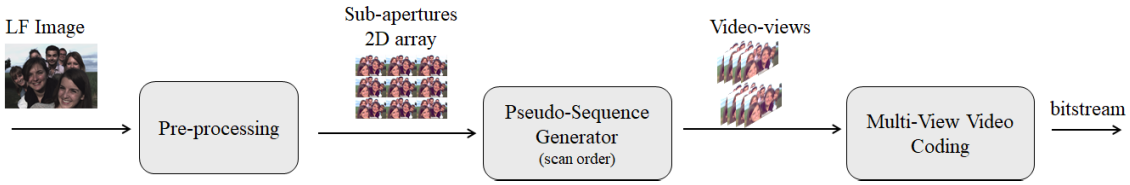


Figure 4.1: Proposed method scheme

A two-dimensional coding structure is also proposed to efficiently assign a prediction structure within the multiple video-views. The prediction structure defines how sub-apertures will be encoded as frames by MV-HEVC. The prediction structure is designed in order to provide good random access and to allow a scalable representation of light fields based on the layered structure of MV-HEVC.

4.1.1 Design goals

While designing a scan order and a coding structure, some factors should be kept in mind for improved performance. Due to the way a video encoder works the similarity between a sub-aperture image and its references is crucial to obtain compression efficiency. Since it can be observed that adjacent sub-apertures (both horizontally and vertically) exhibit higher similarity with each other [32], the closer is the sub-aperture image to the reference the higher is the compression performance. Therefore the scan order and coding structure objective is to ensure prediction from adjacent sub-apertures as much as possible. Additionally, the sub-apertures near the center have higher similarity to other sub-apertures compared with the sub-apertures near the border [35], this is the closest to the corner the more unique is the sub-aperture. For this reason, the more central sub-apertures can better predict the other sub-apertures and should be preferred as reference pictures.

Aspects to improve random access capabilities are also taken into account. Given that to decode a randomly chosen picture form a video sequence all the pictures taken as references during the encoding process will also have to be decoded, the random access cost increases with the number and size of the pictures used as references. Thus to provide a good random access performance, the coding structure should try to keep the number of reference pictures at a minimum while still exploiting the temporal and inter-layer correlations.

The scan orders and coding structures presented in the next sections were designed based on these considerations.

4.2 Proposed Scan Orders

The novel scan order introduced in this dissertation is denoted the Quadratic Spiral scan order. Two other scan orders the Serpentine scan order and Lines scan order are also defined, to be used as comparison.

Serpentine Scan order: Defined by JPEG Pleno Group [36] as an anchor to compare results, a serpentine path is used to form a pseudo-sequence, as illustrated in Fig. 4.2. Starting from the top left sub-aperture image, the sequence follows a horizontal path until the end of the row where reverses the direction for the next row. This is a single-view pseudo-sequence, therefore it does not take advantage of the multi-view extension of HEVC.

Lines Scan order: A straight forward scan order rearranges the sub-apertures the most obvious way into multi-layer pseudo-sequences. The Lines scan order generates four pseudo-sequences of straight lines (Fig. 4.3). Each pseudo sequence follows a straight horizontal path capturing all sub-aperture images in the row, the path repeats itself each four rows.

In the figures, each sub-aperture is identified by a Roman numeral corresponding to the layer the sub-aperture belongs to, and a number indicating its temporal position in the pseudo-sequence.

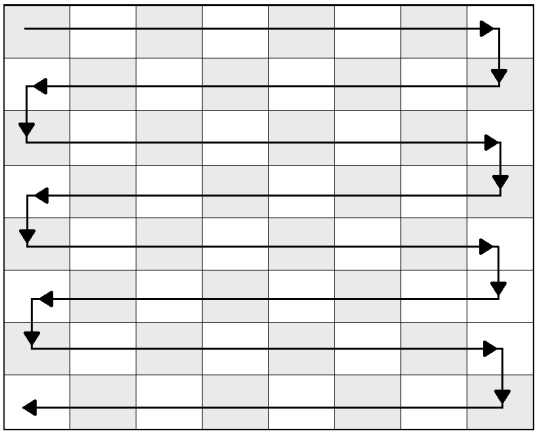


Figure 4.2: Serpentine Scan order

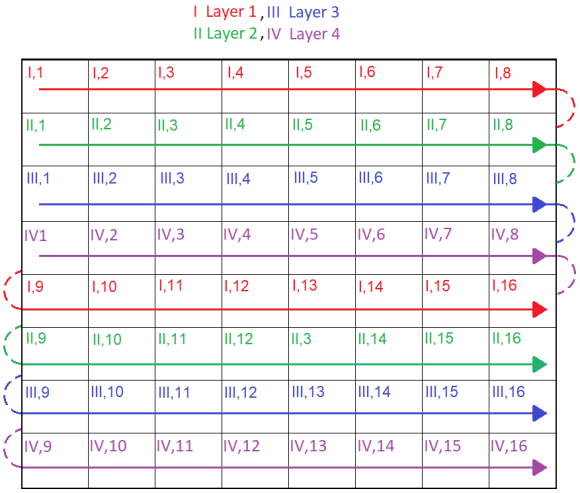


Figure 4.3: Lines Scan Order

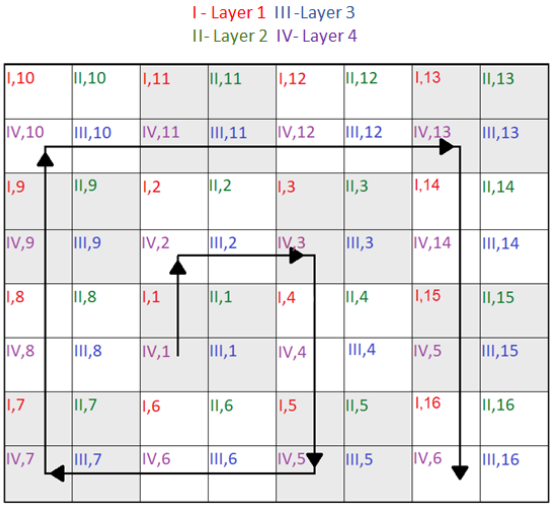


Figure 4.4: Quadratic Spiral Scan Order

Quadratic Spiral Scan order (QSpiral): Illustrated in Fig. 4.4, the sub-apertures images are organized in blocks of 2x2 images, so that each image inside a block will belong to a different pseudo-sequence.We form four pseudo-sequences, each one starting within the central 2x2 block, following a spiral path (block by block), till all sub-apertures images have been assigned to a pseudo-sequence.

A disadvantage of the QSpiral is because it forms a fixed number of four pseudo-sequences, it can only be used with an even number of rows and columns of sub-aperture images. Additionally, the QSprial requires the number of rows to be equal to the number of columns. This, however, is not a common problem since it is standard in light field photography to capture light field with an equal number of rows and columns of sub-apertures.

4.2.1 Other Scan Orders:

In this section, we briefly present a set of scans orders that were considered during our study but failed to achieve the same compression efficiency as the scan orders in the previous section, and were consequently discarded. Nevertheless these scan orders are worth consideration and can serve as inspiration for future works. The experimental results for these scan orders are presented in Annex A.

Quadratic Serpentine Scan Order (Q-Serpentine): Illustrated in figure 4.5. Similar to the QSpiral we also divide the sub-apertures into 2x2 blocks of sub-apertures, where each sub-aperture within the block belongs to a different pseudo-sequence. However, instead of starting with the central block, the four pseudo-sequences will begin in the left superior block and follow a serpentine path till all the sub-apertures belong to a pseudo-sequence. The random access and scalability properties of the QSpiral order described in sections 4.4 and 4.5 will also be applied to the QSerpentine.

Quadrants Lines Scan Order (QuadLines): Depicted in figure 4.6. The sub-apertures images are divided in four quadrants. Four pseudo-sequences will be formed. The first pseudo-sequence will begin in the superior left quadrant with the first sub-aperture in the quadrant, the second frame in pseudo sequence will be the corresponding sub-aperture in the up-right quadrant, the third in down-right quadrant and the forth in the down-left quadrant, the path then lowers one sub-aperture and repeats this pattern. The other pseudo-sequences follow the same path, starting one sub-aperture to the right. The final result is all the sub-apertures in a column belong to the same pseudo-sequence, but the temporal order follows a quadratic path.

Pure Spiral Scan Order (PSpiral): Represented in figure 4.7. This is a simple scan order. Each pseudo sequence starts in one of the central sub-apertures and follows a spiral path. This scan order had its merits with a low count of images (for example 8x8 sub-apertures) but it was not a good choice for bigger sizes. Since the distance between sub-apertures form different pseudo-sequences increase with each curve of the spiral path, it was difficult to find an efficient prediction structure.

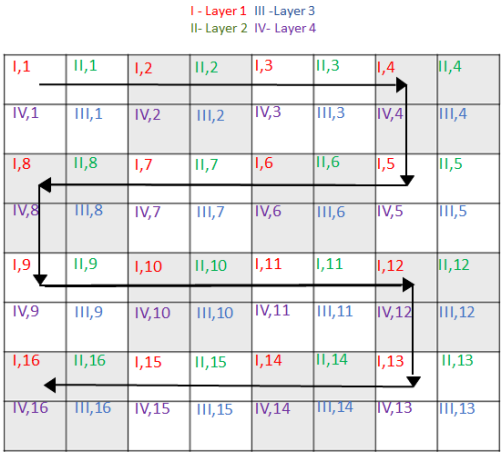


Figure 4.5: Quadratic Serpentine Scan Order

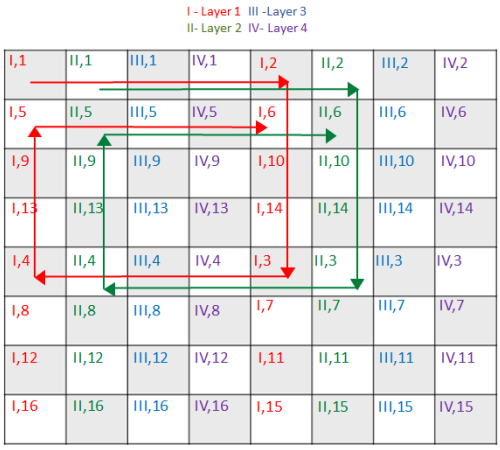


Figure 4.6: Quadrants Lines Scan Order

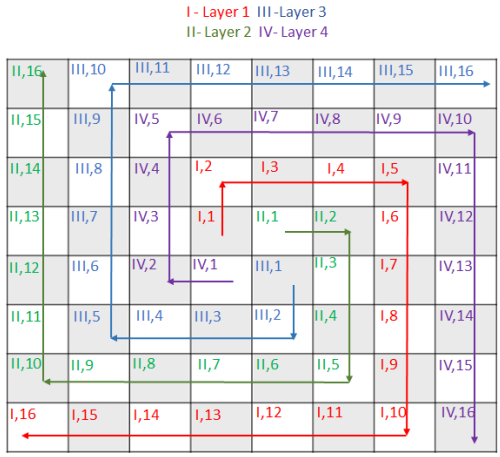


Figure 4.7: Pure Spiral Scan order

4.3 Proposed Multi-View Coding Structure

The coding structure defines how sub-apertures will be encoded as frames by MV-HEVC and the scan order defines the temporal order in which the sub-apertures are rearranged into frames. For this reason, the coding structures and the scan order should be designed together, with the purpose of taking advantage of each other. Based on this concept our proposed coding structure was tailor-made for the QSpiral scan order not only to improve random access performance and compression efficiency but also to capitalize on the multiples video-views and the layer structure to allow a scalable representation of the light field.

In our study of possible prediction configurations, the design choices of section 4.1.1 were followed, with the spatial distance between sub-apertures being the most important factor when choosing the sub-apertures to used as references. The B-Hierarchical configuration [37] was found to provide a good balance between both coding efficiency and random access performance, for that reason it was chosen to be the basic structure at single-layer coding. The configuration is used in the experimental test with a Group of Pictures (GOP) of 8 and an intra period of 16 pictures. Two variants of the same proposed two-dimensional multi-view prediction scheme were developed:

Variant 1: Depicted in Fig. 4.8. All the layers follow the same B-Hierarchical intra-layer prediction scheme. The first layer is completely independent of the other layers, no inter-layer reference is used. In the remaining layers, a picture belonging to the first layer is used as a reference, the sub-aperture image used as a reference is the picture with the same temporal position as the sub-aperture image to be encoded. The second, third and fourth layers can be decoded after decoding only the first layer, this allows to take full advantage of the layered structure of the MV-HEVC.

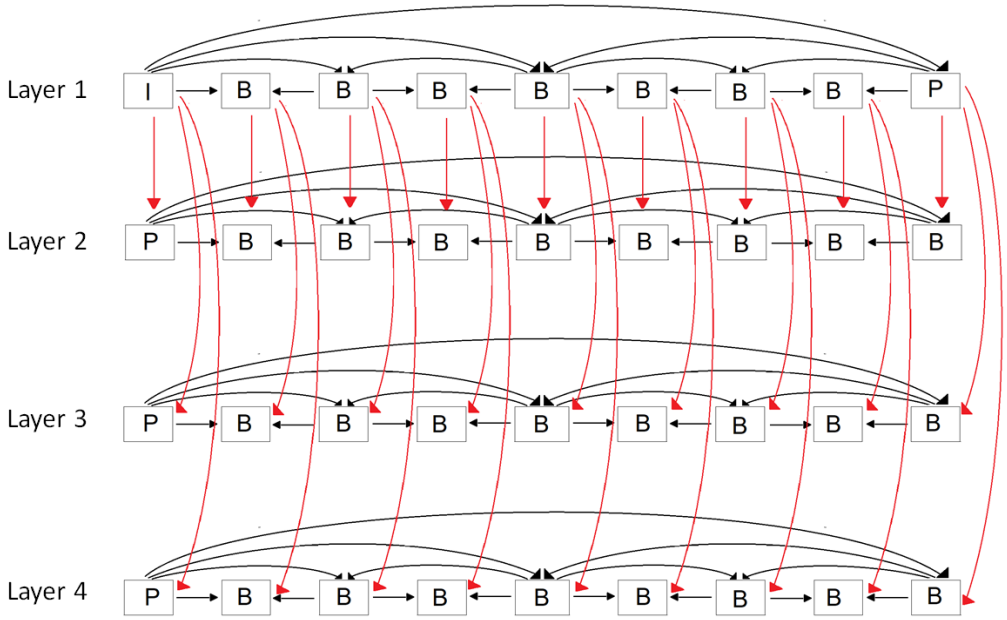


Figure 4.8: *Variant 1*

Variant 2: Depicted in Figure 4.9. This variant differs from the *Variant 1* only in the third layer. In the third layer, there is no intra-layer prediction, all the reference pictures belong to the second and fourth layers. The pictures used as reference are the pictures with the same temporal position as the sub-aperture image to be encoded.

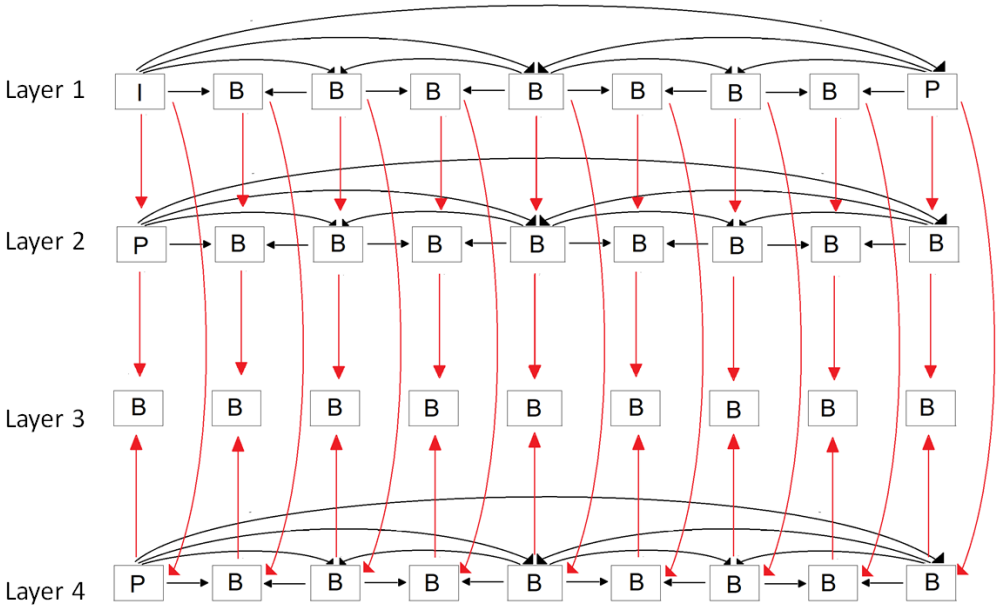


Figure 4.9: *Variant 2*

The reasoning behind the *Variant 2* configuration is in the pseudo-sequences formed using the QSpiral scan order the sub-apertures form the second and fourth layers are spatially closer to sub-apertures form the third layer compared to other sub-apertures within the same layer, this effect is illustrated in figure 4.10. The *Variant 2* allows for pictures in the third layer to be predicted only from pictures inside the 2x2 block. Since using the *Variant 2* the sub-apertures are used as reference spatially closer, the compression efficiency will increase. However, to decode a picture from the third layer the pictures form the first, second and fourth layers will also need to be decoded. As a consequence, the gains in compression efficiency come at the cost of worse random access performance.

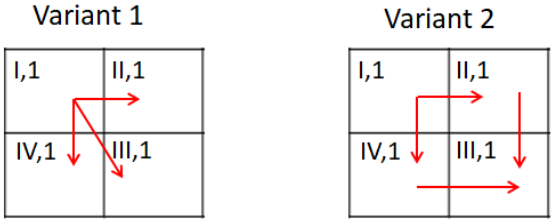
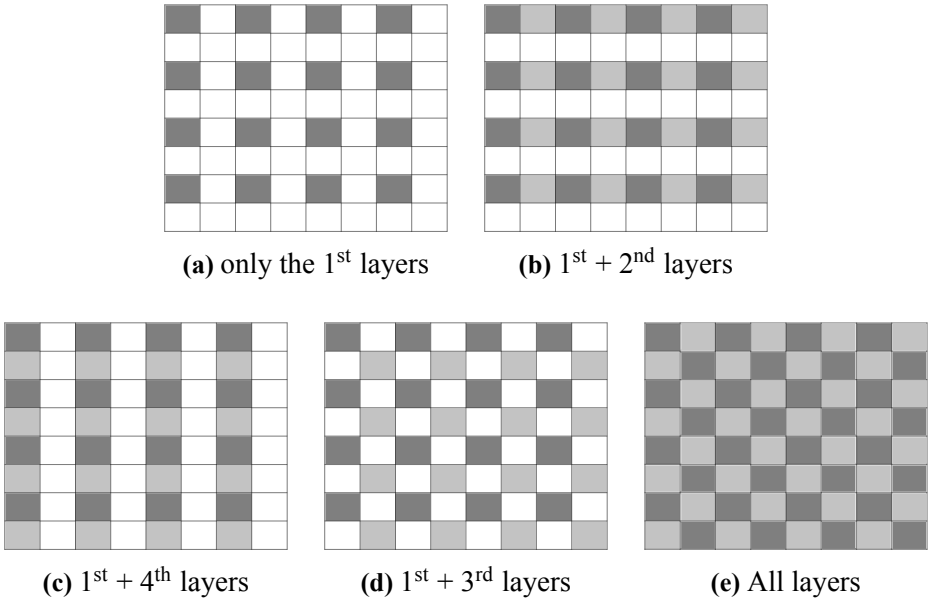


Figure 4.10: Inter-layer prediction of both variants on 2x2 block from the QSpiral

4.4 Scalability Design

In the proposed method the scalability is introduced because pseudo-sequences are encoded by the MV-HEVC in different layers. As described in section 3.2 the layered structure of the MV-HEVC supports a scalable bit-stream depending on the coding structure. Therefore while designing the coding structure it was chosen to make the first layer totally independent of other layers, and the second, third, and fourth layers only dependent on the first layer (*Variant 1*). This coding structure allows to decode the first layer without having to decode any additional information of the other layers and to decode the remaining layers using only the first layer.

The QSpiral was designed to take advantage of this scalable functionality provided by the multi-layer coding structure. Each individual pseudo-sequence generated by the QSpiral scan order contains a set of sub-apertures evenly distributed in a two dimensional representation. Since MV-HEVC allows to scalable decode layers, it is also possible to scalable decode our sets of sub-apertures. Due to the fact that the sub-apertures are evenly distributed, they can represent the light field by periodic patterns of sub-apertures as shown in Fig. 4.11. Fig. 4.11a shows the periodic pattern corresponding to the decoding of the first layer. When the first and the second layer are decoded the periodic pattern is comprised of alternated rows depicted in Fig. 4.11b The periodic pattern of alternated columns shown in Fig. 4.11c is obtained when the first and fourth layer is decoded. Finally, when the first and third layers are decoded the check pattern represented in Fig. 4.11d is obtained. This scalable representation is useful for applications that do not require decoding the full light field image and can take advantage of the patterns.



(a) only the 1st layers **(b)** 1st + 2nd layers

(c) 1st + 4th layers **(d)** 1st + 3rd layers **(e)** All layers

Figure 4.11: Decoding of different layers generated by the Quadratic Spiral scan order

The *Variant 2* however, cannot produce the check pattern in Fig.4.11d since to decode the third layer all the other layers have also to be decoded.

4.5 Random Access Design

The way the QSpiral scan order combines with the coding structure is optimized to improve the random access performance.

In our proposed coding structure the sub-aperture images from the first layer will be the baseline images used as references for the remaining layers. Using the QSpiral scan order each one of these sub-apertures will be used to predict sub-apertures from the other layers in the corresponding 2x2 block. Since the sub-apertures from the first layer are evenly distributed between the rows and column, it allows access to the remaining sub-apertures with reduced decoding effort.

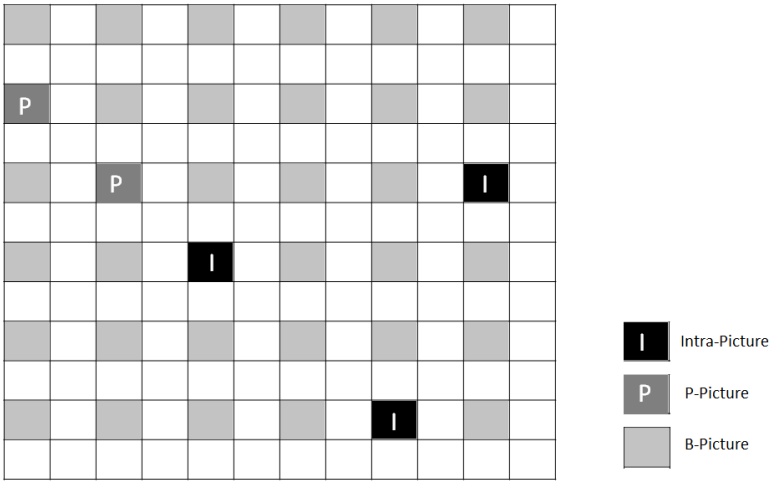


Figure 4.12: Sub-apertures images from the first layer in a 12x12 sub-aperture representation

Figure 4.12 shows an example of 12x12 sub-apertures representation of a light field where the sub-apertures from the first layer are colored in grey tones. One can observe the even distribution of sub-apertures from the first layer done by QSpiral. The figure also shows the distribution of sub-apertures coded as intra and as P-pictures. The sub-apertures coded in intra mode function as a clean random access point and the P-pictures are directly encoded from the intra pictures. This type of pictures facilitate the access to other nearby sub-apertures. Due to the proposed coding structure, the I-pictures and P-pictures are almost evenly spread across the 2D sub-aperture representation, which improves the random access performance.

In section 5.3.4 the results for random access performance are presented and analyzed.

5

Experimental Setup

This chapter describes the common test conditions used in the light field coding experiments. The experimental setup follows the conditions defined by JPEG Pleno Group [36] and ICME Grand Challenge 2016 [6]. Compliance with these conditions allows our developed work to be compared with light field state-of-art compression techniques.

5.1 Test Images

The light field test images used in the experiment were made available by JPEG Pleno Group at https://jpeg.org/plenodb/. The images are provided by multiples research centers and are chosen to be diverse in terms of acquisition/creation technology; scene geometry; spatial resolution; number of sub-apertures.

Lenslet Lytro Illum Camera: This light field images were captured using a Lytro Illium B01 light field camera, the images selected from this set selected for experiments are *Bikes, Danger de Mort, Stone Pillars Outside* and *Friends 1*, depicted in figure 5.1

- Acquisition: Plenoptic 1.0 camera.
- Content: Natural, outdoor.
- Spatial resolution: 625 x434.
- Number of sub-apertures: 15x15.

Synthetic HCI HDCA: Created synthetically and made available by the Heidelberg Collaboratory for Image Processing (HDCI) [38], the images *Greek and Sideboard* were chosen for tests.

- Acquisition: Synthetic HCI HDCA
- Content: Synthetic objects
- Spatial resolution: 512x512.
- Number of sub-apertures: 9x9.

Stanford HDCA: The Computer Graphics department at Stanford University used a High Density Camera Array to capture several light field images for research purposes [14]. The *Legos Knight* light field image (Figure 5.3a) is selected for experiments .

(a) Bikes

(b) Danger de Mort

(c) Stone Pillars

(d) Friends 1

Figure 5.1: Plenoptic Images examples sub-apertures

- Acquisition: HDCA
- Content: Indoors with complex specularities.
- Spatial resolution: 1024x1024.
- Number of sub-apertures: 17x17.

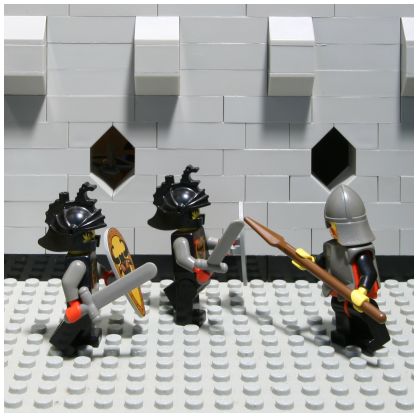
Poznan: This light fields were captured by the Poznan University of Technology [39], currently the JPEG Pleno only includes one light image from the Poznan set named *Laboratory 1* (Figure 5.3b). This light field images is considerably bigger than the previous selected test images, both in resolution and number of view.

- Acquisition: HDCA
- Content: Natural, indoors.
- Spatial resolution: 1936x1288.
- Number of sub-apertures: 31x31.

(a) Greek example sub-aperture

(b) Sideboard example sub-aperture

Figure 5.2: Synthetic light fields



(a) Legos Knights example sub-aperture

(b) Laboratory1 example sub-aperture

Figure 5.3: HDCA light fields

5.2 Test Conditions

In the case of LF images captured by plenoptic cameras, the Matlab Lytro Toolbox [40] is used to demosaicing and devignetting the raw data. The lenslet image (with a resolution of 7728x5368px) is converted to light field 5D data structure (15x15x434x625x4) as shown in Figure 5.5, with 15x15 sub-apertures, each sub-aperture with 434x625 pixels. Each pixel has RGBW components, the W values are used for color correction and discarded before encoding. All the sub-apertures are converted from RGB to YUV 4:2:0. We only use the central 12x12 sub-apertures to avoid using the corner sub-apertures darkened due to the vignetting. Each sub-aperture is also padded to 440x632 by adding columns and lines of zeros to have a resolution compatible with HEVC coding.

The HDCA images were used without the need for special processing or padding. The HDCA images were converted to YUV 4:2:0 8 bits format, and to keep the number of images even, we only use the central 16x16 images from HDCA form Standford, the central 8x8 images from Synthetic HDCI HDCA images and the central 30x30 images form the Poznan University.

The Matlab scripts available in Annex B were used for color conversion from RGB to YUV and to down-sample form 4:4:4 to 4:2:0. The Serpentine pseudo-sequence is encoded with a single layer B-Hierarchical configuration described in section 4.3. We used the HTM.16.0 [26] reference-software compiled to MV-HEVC for multi layers scan order schemes and complied to plain HM for the Serpentine order, the results were achieved by setting QP values to 15, 20, 25, 30. It was used a CU size of 64.

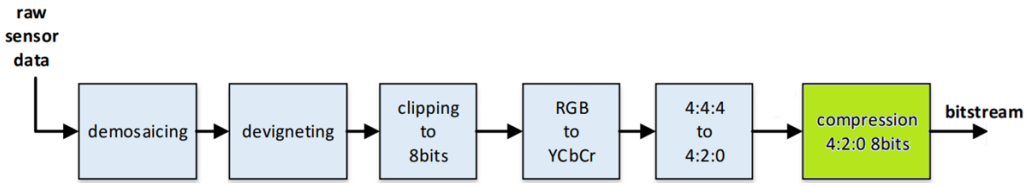


Figure 5.4: Processing pipeline adapted from the ICME Grand Challenge 2016 [6].

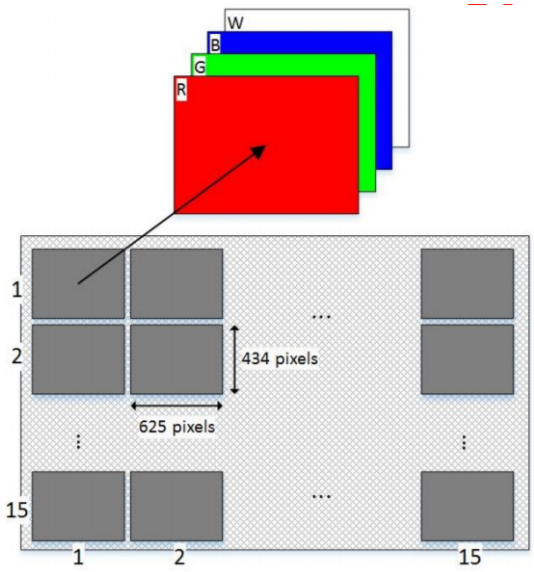


Figure 5.5: Light Field structure of plenoptic image, image form [6]

5.3 Performance Metrics

This section defines the metrics to evaluate the performance of the proposed method and other solutions. We evaluate all the methods in terms of compression rate, image quality and random access performance.

5.3.1 Rate Metrics

To evaluate compression factor the main rate metric used is the number of bits per pixel (*bpp*) defined as:

$$bpp = \frac{Total\_Nr\_Bits}{Total\_Nr\_Pixels}$$

where $Total\_Nr\_Bits$ is the number of bits of the encoded bit-stream and $Nr\_Total\_Nr\_Pixels$ is the number of pixels in the light field image, for example plenoptic images have 12x12 sub-apertures each with a 440*625 resolution , therefore the in this case $Total\_Nr\_Pixels = 12 * 12 * 440 * 632 = 40{,}043{,}520$ pixels. A smaller value of *bpp* corresponds to a higher compression rate.

5.3.2 Quality Metric

When compressing an image, the high impact information is being discarded to save file size, this creates distortion and quality loss the higher the compression rate. To measure the quality difference between the original image *I* and the reconstructed image *R* recovered from the compressed bitstream the Peak Signal Noise Ratio (PSNR) is used. The PSNR is expressed in terms of the logarithmic decibel scale (dB). PSNR_YUV is calculated by the HEVC encoder as an average of the PSNR values of the Y, U, and V components as follows:

$$PSNR_Y(k,l) = 10log_{10}\frac{225^2}{MSE(k,l)}$$

$$MSE(k,l) = \frac{1}{m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - R(i,j)]^2$$

where *k* and *l* identify each individual sub-aperture image. The *m* and *n* are the dimension of each sub-aperture in pixels. $I(i,j)$ and $R(i,j)$, is the value of pixel on position (i,j) of the corresponding channel (Y,U or V).

The $PSNR_{YUV}(k,l)$ value for each individual sub-aperture is computed as:

$$PSNR_{YUV}(k,l) = \frac{6PSNR_Y(k,l) + PSNR_U(k,l) + PSNR_V(k,l)}{8}$$

The average $PSNR_{YUVmean}$ value of the full light field images is computed as a mean $PSNR_{YUV}$ value for each individual sub-aperture as:

$$PSNR_{YUVmean} = \frac{1}{K * L}\sum_{k=0}^{K}\sum_{l=0}^{L}PSNR_{YUV}(k,l)$$

where K and L correspond to amount of sub-apertures in each horizontal and vertical direction, respectively.

5.3.3 Compression Efficiency - Bjøntegaard Metric

There is a trade-off between compression rate and image quality. An encoder is more compression efficient compared to another encoder if, for the same image quality, it can obtain higher compression rates. Therefore the compression efficiency of a method can be successfully represented by a rate/distortion (RD) curve. To compare the performance of different compression methods we will use the Bjøntegaard Delta rate [41]. The Bjontegaard metric allows computing the average distance between two rate/distortion curves. Using this metric the performance of a compression method can be expressed in terms of: Bit-rate savings (BD-Rate) and quality gains (BD-PSNR). The BD-Rate express the average percentage of bits saved using the method compared to a reference method for the same quality. The BD-PSNR express the average gains in quality (PSNR) for the same compression rate. Figure 5.6 show a example Bjøntegaard delta rate metric between two curves from different methods.

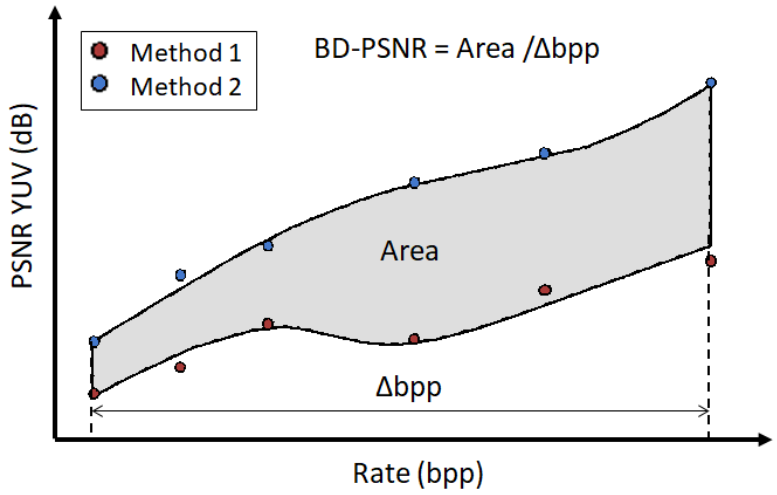


Figure 5.6: Example of the Bjøntegaard delta rate metric (BD-PSNR).

When comparing the performance of the different scan orders Bjontegaard metric will be applied using the Serpentine scan order as reference. In the case of plenoptic images from Lenslet Lytro Camera, the JPEG Pleno group made available anchor images to be used as a comparison. These anchors were generated by a JPEG encoder using the same test conditions previously described (YUV 4:2:0). For plenoptic images, the Bjontegaard metric is additionally applied using the JPEG anchors as reference.

5.3.4 Random Access Metrics

In order to evaluate the random access capability of a compression method the size in bits of all the encoded sub-apertures is determined and added accordingly to the predicted structure to calculate the necessary bits to decode any sub-aperture. This allows defining the random access cost (j) as the ratio between the number required of bits to decode a specific sub-aperture (j) and

the total amount of encoded bits.

$$Random\ Access\ Cost\ (j) = \frac{bits\ required\ to\ decode\ a\ sub-aperture(j)}{Total\ bits\ needed\ to\ decode\ the\ full\ light\ field}$$

To give the general performance of a method the metric random access index is defined as the mean of random access cost.

$$Random\ Access\ Index = \frac{Average\ bits\ required\ to\ decode\ a\ sub-aperture}{Total\ bits\ needed\ to\ decode\ the\ full\ light\ field}$$

Initially the worst-case scenario defined as the ratio between the maximum amount of encoded bits required to decode a sub-aperture and the total amount of encoded bits was also considered to be used as a metric of random access. However, since the behavior of the worst-case scenario is the same as the random access index, the results section only the random access index is used to evaluate the method random access performance.

6

Results and Analysis

This chapter presents and analyzes the experimental results. The results are divided into three sections, the results for compression efficacy, the results of random access and to end the trade-off between compression efficiency and random-access performance.

6.1 Compression Efficiency

The compression efficiency results are shown in Table 6.1 for plenoptic images and in Table 6.2 for HDCA images. The tables present the bitrate (BD-Rate) gains and PSNR gains (BD-PSNR) compared to the Serpentine scan order encoded by a HEVC encoder. A negative value of BD-Rate means the method can compress the LF with % less bitrate compared to the Serpentine scan order. The Rate-Distortion curves of the different scan orders are shown in Fig. 6.2.

Image	Configuration B-Hierarchical	PSequence	BD-Rate (%)	BD-PSNR (dB)
I01 Bikes	Variant 1	Lines	-8.13	0.23
		QSpiral	-2,52	0,06
	Variant 2	Lines	-26.36	0.83
		QSpiral	-15.56	0.42
I02 Danger de Mort	Variant 1	Lines	-16.26	0.55
		QSpiral	0.46	-0.03
	Variant 2	Lines	-26.29	0.9517
		QSpiral	-7.68	0.22
I04 Stone Pillars	Variant 1	Lines	-28.33	0.78
		QSpiral	-12.47	0.25
	Variant 2	Lines	-33.67	0.99
		QSpiral	-19.63	0.45
I10 Friends 1	Variant 1	Lines	-15.41	0.37
		QSpiral	-11.45	0.26
	Variant 2	Lines	-35.09	0.97
		QSpiral	-26.55	0.64
Average	**Variant 1**	**Lines**	**-17.0325**	**0.4825**
		QSpiral	**-7.82**	**0.16**
	Variant 2	**Lines**	**-30.3525**	**0.935425**
		QSpiral	**-17.355**	**0.4325**

Table 6.1: Plenoptic Images

Image	Configuration B-Hierarchical	PSequence	BD-Rate (%)	BD-PSNR (dB)
HDCI Greek	Variant 1	Lines	-28.54	0.9
		QSpiral	-27.88	0.82
	Variant 2	Lines	-35.39	1.16
		QSpiral	-29.3	0.86
HDCI Sideboard	Variant 1	Lines	-23.1	1.03
		QSpiral	-13.53	0.57
	Variant 2	Lines	-34.79	1.57
		QSpiral	-18.99	0.75
HDCA Standford Legos Knight	Variant 1	Lines	-16.32	0.48
		QSpiral	-6.12	0.16
	Variant 2	Lines	-16.60	0.49
		QSpiral	-7.0	0.18
HDCA Pozan	Variant 1	Lines	-17.87	0.30
		QSpiral	-3.77	0.05
	Variant 2	Lines	-28.7	0.54
		QSpiral	-9.71	0.17
Average	**Variant 1**	**Lines**	**-21.4575**	**0.6775**
		QSpiral	**-12.825**	**0.4**
	Variant 2	**Lines**	**-28.87**	**0.94**
		QSpiral	**-16.25**	**0.49**

Table 6.2: HDCA Images

It can be observed that the Lines scan order is the best performing scan order in all tests with the QSpiral not far behind. The QSpiral has an average BD-PNSR difference from the Lines scan order of 0.413 dB in plenoptic images and 0.364 dB in HDCA images. The reason the Lines scan order outperforms the QSpiral is because the sub-apertures are rearranged in the pseudo-sequence

in the most obvious way to explore the high correlation between adjacent sub-aperture images, for each frame of the pseudo-sequence the next and previous frame are adjacent sub-aperture to the one being encoded.

As expected the *Variant 2* achieves consistently higher compression efficiency compared to the *Variant 1*, with an average difference of 0.73 dB in plenoptic images and 0.35 dB in HDCA images. Since the RD-curves using the *Variant 1* have the same behavior as using the *Variant 2*, in figure 6.2 only the RD-curves obtained using the *Variant 2* are presented. An analysis of the obtained rate-distortion curves supports the conclusion that the pseudo-sequence based compression approach is efficient, especially in low bit-rate scenarios [42].

Table 6.3 reports the overall performance of the QSpiral scan order, it shows improvements over 7.8 dB compared with reference JPEG for plenoptic images. These results are in accordance with the best performing current state of the art multi-view pseudo-sequence methods. Figure 6.1 shows the performance of the proposed compression method compared with the method developed by Ahmad et al in [34]. It should be noted that both methods use different prediction schemes, the method from [34] have more pictures coded as Intra, which results in a lower compression efficiency. These results verify that one of our main goals, to obtain high compression efficiency is accomplished.

Image	**Configuration B-Hierarchical**	**BD-Rate (%)**	**BD-PSNR (dB)**
I01 Bike	Variant 1	-82.15	7.3
	Variant 2	-86.98	7.79
I02 Danger de Mort	Variant 1	-82.15	7.3
	Variant 2	-82.15	7.3
I04 Stone Pillars	Variant 1	-83.72	6.77
	Variant 2	-84.99	7.14
I10 Friends	Variant 1	-86.98	7.79
	Variant 2	-92.1	7.75
Average	**Variant 1**	**-83.75**	**7.29**
	Variant 2	**-86.55**	**7.49**

Table 6.3: BD-PSNR/Rate: Comparison of QSpiral method with reference JPEG anchors (Plenoptic images).

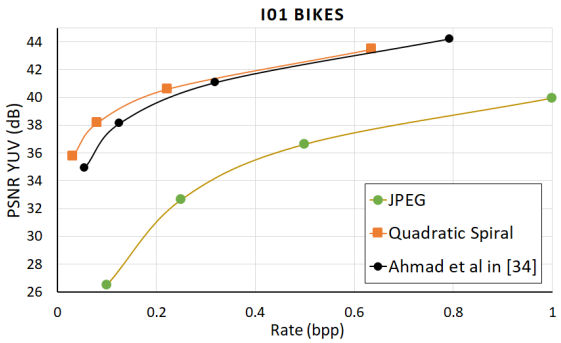


Figure 6.1: Developed method compared with the method from [33] for the *I01 Bikes* Light Field

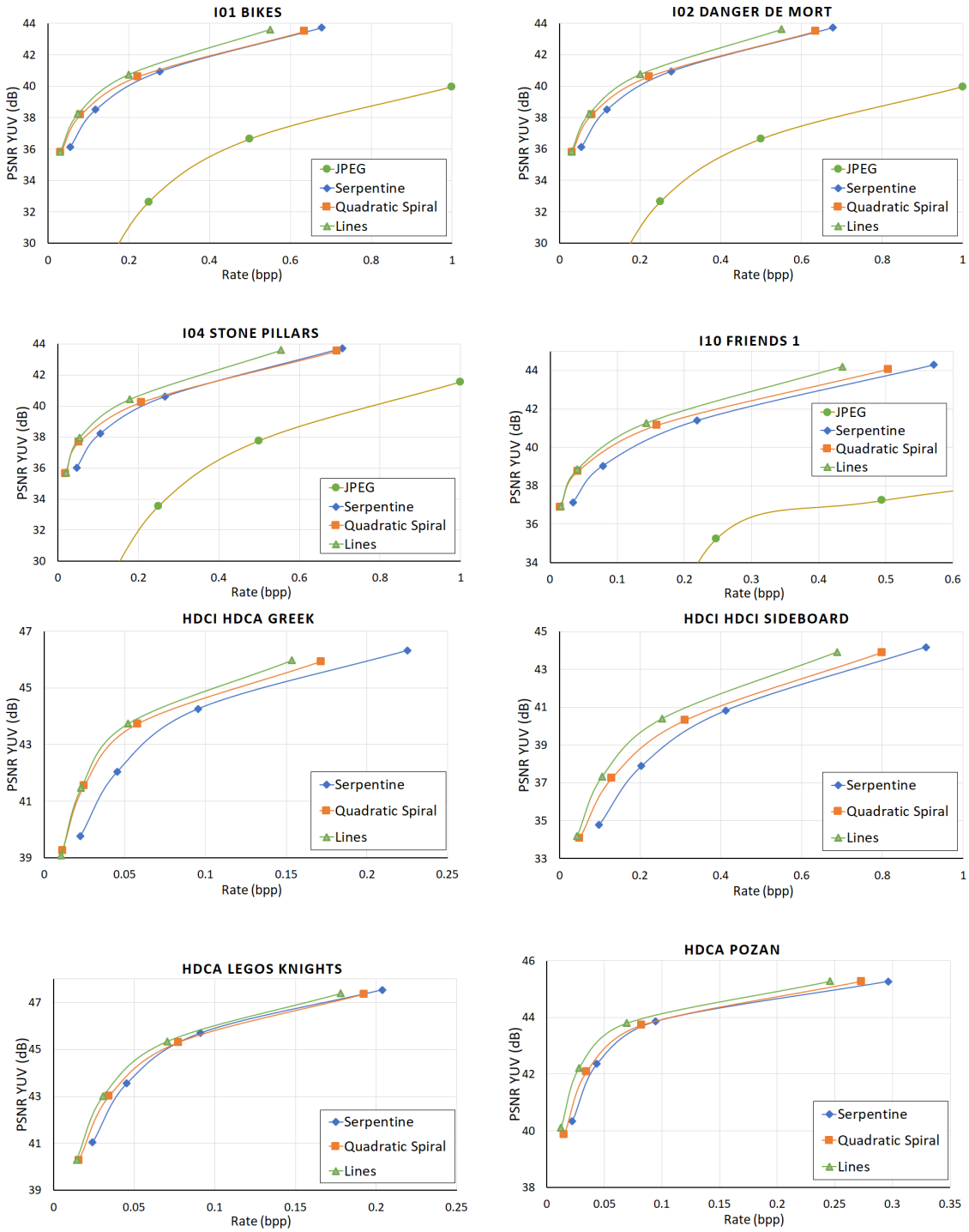


Figure 6.2: Rate Distortion curves (*Variant 2*).

6.2 Random Access Performance

Another objective was to develop a coding scheme with improved random access performance. This section analyses the random access performance of our scan orders and coding structure. Figure 6.3 shows the random access cost of each sub-aperture form the *I01 Bikes* light field. The data form Lines and QSpiral scan orders is referring to pictures from the first layer. The effect of the proposed coding structure is evident in the figure. The pictures with POC 0 and 16 have the lower random access cost because are encoded in intra mode, for the remaining pictures the hierarchical coding structure is visible especially in multi-layer scan orders.

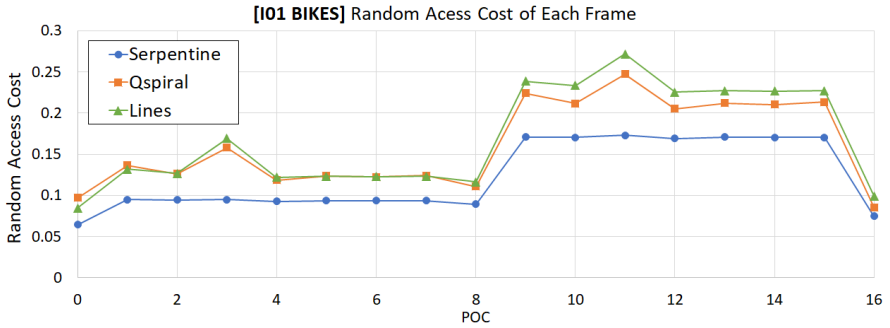


Figure 6.3: The random access cost of each frame from the *I01 Bikes* image (*Variant 1*, QP=30).

Figure 6.4 shows the random access index in relation to the compression rate for each image. It can be observed that as the compression rate increases the random access index also increases by a small margin, this is probably due to the headers effect on the bit-stream. In higher bit-rate scenarios this effect is negligible and the random access index assumes a constant value.

In all the tests the Serpentine scan order has lowest Random Access Index values. This happens because the Serpentine order uses a single-view predictive scheme and consequently uses fewer references pictures during encoding. Additionally, the coding structure has a GOP of 8 pictures but since it is a single-view scheme it will have more repetition of the same GOP and therefore more intra-encoded pictures, lowering the random access index.

Although both the QSpiral and the Lines scan orders use the same multi-view prediction scheme described in section 4.3, the QSpiral order achieves lower random access index values. This confirms that our proposed QSpiral scan order has improved random access performance. This improvement is a direct result from the design choices described in section 4.5.

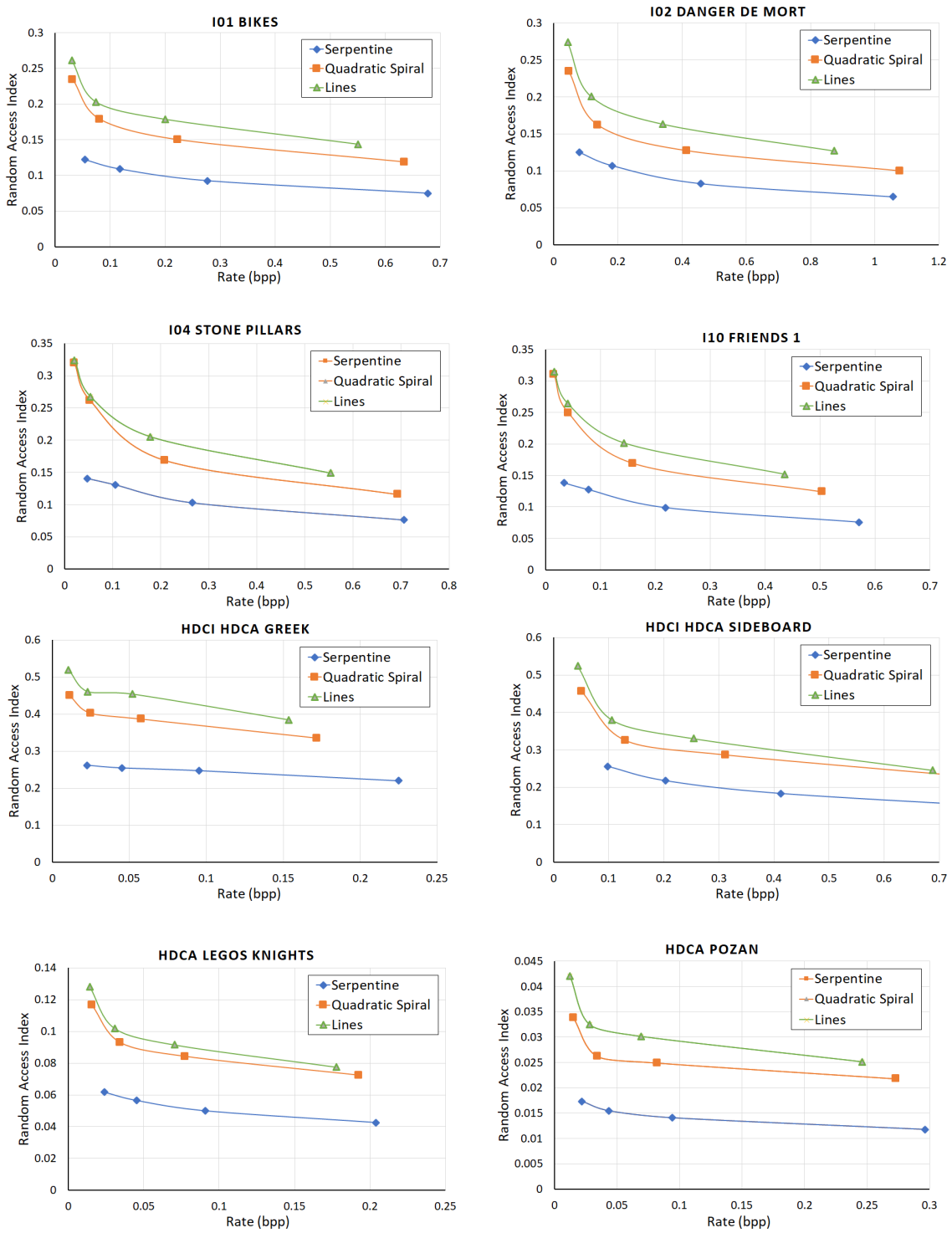


Figure 6.4: Random Access Index to compression rate (*Variant 2*).

6.2.1 Random Access to Central Views

It is also worth pointing, due to the fact the QSpiral scan order starts its path with the central sub-apertures and follows a spiral path to the outer limits the access to the most central sub-apertures is facilitated. We defined the 25% sub-apertures more spatially close to the central sub-aperture as the central sub-apertures, as exemplified in figure 6.5. The experimental results show an average 20% less random access cost to decode the central sub-apertures compared with the average random access cost. The relevance of this property is the central sub-apertures represent the front-view of a scene, and it is common practice in photography to capture the most important information from the front, therefore there is a high probability the central sub-apertures will be required in applications.

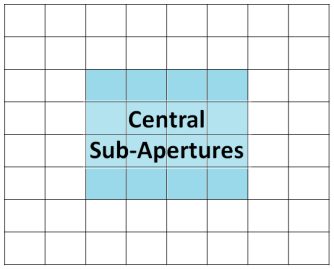


Figure 6.5: Example of central sub-apertures in a 8x8 sub-aperture representation of a light field.

6.3 Analysis of the Trade-off between Compression Efficiency and Random Access

The results from the previous sections show a clear trade-off between random access and compression efficiency. The higher is the compression efficiency of a scheme the worst is the random access performance (i.e higher random access index). It is possible to visualize this trade-off as is done in figure 6.6 by graphing the bitrate gains versus the mean random access index of each scan order.

The results show that for all the tests there is not an optimal solution, i.e a scheme with both the best compression efficiency and random access performance. The decision of the preferred scheme (scan order + coding structure) would vary according to the needs. There are however some cases, most noticeably *I02 Danger de Mort* and *HDCI Sideboard* light fields, where the QSpiral under-performed.

It can be noticed a significant difference between the points obtained using different light field images. This is due to the high variance of the test images (plenoptic, HDCA, synthetic HDCI HDCA). In a general way, the QSpiral and both variants of the coding structure fulfill the purpose that they were designed for, improving the compression efficiency while still having a good random access performance.

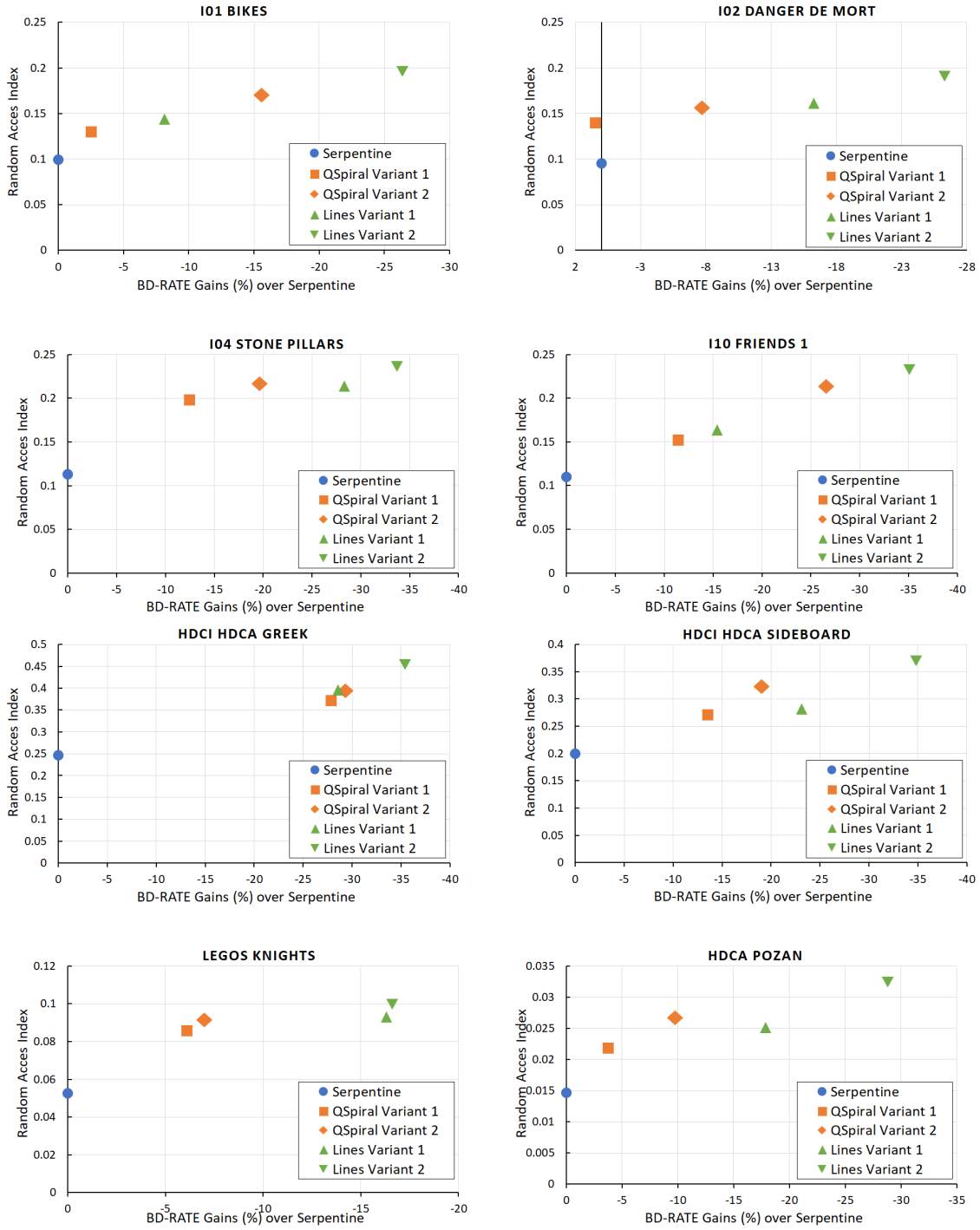


Figure 6.6: Trade-off between compression efficiency and random access.

7

Quantization Parameter Optimization

The experimental results presented in the previous chapter were obtained using a fixed quantization parameter (QP) during the encoding process. However, improved performance can be obtained if the QP is adjusted to each frame of the pseudo-sequence representing the LF image. In this chapter, the QP is assigned to each frame dynamically according to the frame requirements. It is demonstrated that this optimization improves significantly the proposed compression method. However, it should be noted that the proposed optimization does not guarantee the best solution from all the possible solutions. The best QP values for each frame are found within limited space of solutions, that seeks to approximate the most suitable solutions. Therefore it is not an optimization in a strict sense but rather an empirical search for the best performing QP values.

7.1 Quantization Step and Rate Control

The QP controls the size of the quantization step (Qstep). As explained in section 3.1.5, the residual signal after being transformed is divided by the Qstep value and rounded to the next integer. The Qstep value defines the amount of detail saved. An increase in the Qstep results in more information discarded and more distortion. However, a higher compression rate is achieved leading to lower required bitrate for the transmission.

In the previous chapters, fixed QP values were used in the experimental tests. However, we envisaged that this is not the best option and, we anticipate that similarly to happens in traditional video encoding LF images could benefit from a QP step size adjustment. To verify our hypothesis the most straight forward approach was to apply one of the most widely used rate control mechanism provided by HEVC [43]. In rate control, a target bitrate is defined beforehand and the encoder adjusts the QP value of each picture to achieve the target bitrate using a R-λ model [43]. Rate control is very useful especially in systems with a predefined bandwidth (ex CDs, streaming services) since unlike fixed QP it allows control of the output bitrate. In this chapter, the pseudo-sequences formed by the QSpiral scan order were encoded using rate control. Figure 7.1 shows the rate control RD-curve compared with using a fixed QP (QSpiral, *Variant 1*), for two LF images. From both images, it is noticeable that fixed QP always outperforms the HEVC control rate mechanism in compression efficiency. However, for the higher bitrates the results of both schemes converge.

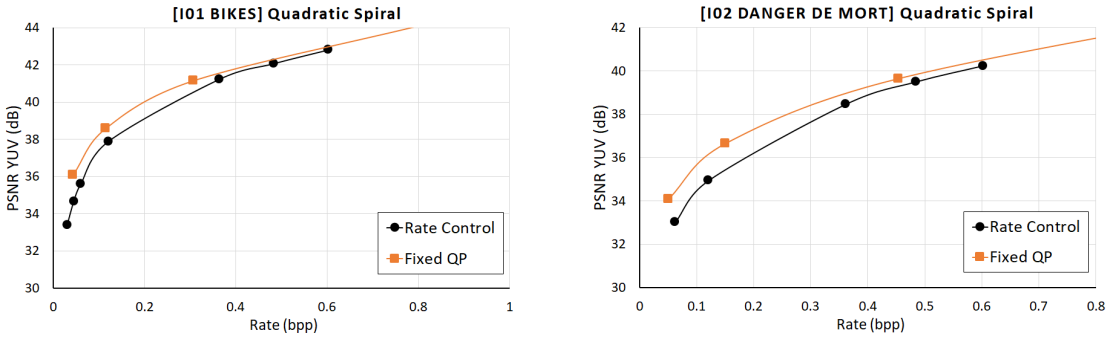


Figure 7.1: Rate control RD-curves compared to fixed QP RD-curves

These tests were applied to several light field images, and the results show that rate control provides consistently poorer compression efficiency compared to using fixed QP. The reason for this is because the rate control mechanism is not well optimized for multi-view coding [44] [45]. Furthermore, rate control is designed to assign a QP value to each picture according to its complexity, however, in pseudo-sequences formed form light field images all the pictures have similar complexity.

From the results was concluded that rate control is not a good solution for our multi-view compression method. As demonstrated here, in some cases higher compression efficiency can be achieved using predefined fixed QP values. For example *Netflix* created a framework exploiting individual fixed QP encoding for each frame to achieve optimal compression efficiency [46]. A similar solution to optimize our compression method is proposed in this chapter.

7.2 Proposed Method - Quantization Step Dimensioning

The objective is to find the best QP values for each picture that improves the efficiency of the proposed compression method. Our reasoning starting point is to recognize that if the reference pictures have good quality, the reconstruction quality will also be good. If the reference pictures have poor quality its reconstruction will have poor quality and due to propagation errors, the subsequent picture reconstruction will also present poor quality. Therefore the key concept behind the proposed QP dimensioning is to assign to pictures that are used more often as references lower QP values (better quality) to improve the compression efficiency.

Setting the reference pictures with lower QP values will increase quality but also their size. Consequently increasing the random access cost of the predicted pictures that use them as references. For this reason, while choosing QP values attention must also be paid to the compression efficiency/random access trade-off.

To find out the best QP values for each picture, a heuristic study was performed. The study consists of a series of encoding tests. For each test the QP of the pictures taken less often as a reference is gradually increased. The results obtained will be compared and analyzed, in order

to identify the difference of QP between pictures that lead to an optimal solution. To assign QP values to the pictures two distinct approaches will be followed:
(i) Assign different QP values to each layer of the MV-HEVC;
(ii) Assign different QP values to each picture within a layer based on a level system.

7.3 QP optimization in Layers

The first approach studies the effect of different QP values in the layered structure of the compression method. In this approach, all the pictures in the same layer are encoded using the same QP value. However, since the second, third and fourth layers take pictures from the first layer as a reference, the first layer should be encoded with a lower QP compared to the other layers.

A heuristic algorithm was designed to find the best difference of QP between layers. The algorithm performs ten tests, each one encoding the LF four times. For each test, the first layer is encoded using the base QP values of QP_b=30,25,20,15. The second, third and fourth layers are all encoded with the same QP value. The QP value of second, third and fourth layers will be $QP_b + i$, where i starts at one and is increased by one each test. This algorithm will encode our pseudo-sequences up to a difference of 10 QP between the first and the other layers. The pseudo-code of QP optimization in layers heuristic algorithm is presented next.

Algorithm 1 QP optimization in layers heuristic algorithm

for *i=1:1:10* **do**
  for $QP_b = 15 : 5 : 30$ **do**
    $QP_{Layer1} = QP_b$
    $QP_{Layer2} = QP_b + i$
    $QP_{Layer3} = QP_b + i$
    $QP_{Layer4} = QP_b + i$
    $Perform\ codification$
    end
    Save results as QP+i
  end

The experimental setup is the same as described in chapter 5. Only the QSpiral scan order (using *Variant 1*) is used since it is the scan order of interest. The same light field test images were used, excluding the HDCI HDCA *Sideboard* and the Poznan *Laboratory 1*. The reason these images were discarded is because the heuristic study contains a high number of experimental tests, and due to the massive size of the LF images the computation time of the tests can be very long. To identify each test interaction the nomenclature QP+N is used, where N is the amount of increased QP in the second, third and fourth layers. For example, QP+6 represents the test where the first

layer is coded with QP=30,25,20,15 and the second, third, and fourth layers with a QP increase of +6, this is with QP=36,31,26,21.

Figure 7.2 depicts the RD-curves from fixed QP, QP+6, and QP+10 tests. It can be easily seen the improvement provided by QP+6 and QP+10 tests over a fixed QP. One can also see that the QP+10 underperforms compared to the QP+6 although it has a bigger increase of QP in layers.

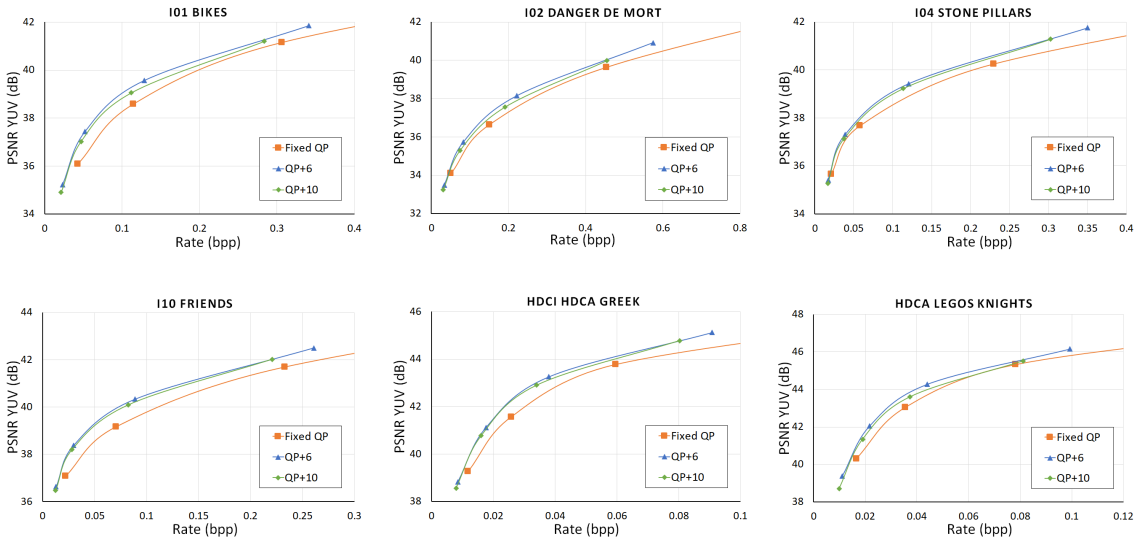


Figure 7.2: Rate Distortion curve of fixed QP, QP+6 and QP+10.

Image	BD-Rate Gains (%) compared to fixed QP (increase of QP in Layers)									
	QP+1	**QP+2**	**QP+3**	**QP+4**	**QP+5**	**QP+6**	**QP+7**	**QP+8**	**QP+9**	**QP+10**
I01 Bike	-12.45	-17.18	-20.43	-22.51	-23.70	**-24.04**	-23.78	-22.85	-21.33	-19.30
I02 Danger de Mort	-5.99	-10.62	-13.93	-16.09	-17.34	**-17.66**	-17.17	-15.84	-14.08	-11.76
I04 Stone Pillars	-7.66	-12.74	-15.81	-17.76	-18.47	**-18.52**	-18.13	-17.38	-16.20	-14.68
I10 Friends	-17.03	-22.51	-26.08	-28.29	-29.54	**-30.10**	-30.01	-24.00	-28.62	-27.42
HDCI HDCA Greek	-7.3381	-12.16	-15.49	-17.84	-19.37	**-20.15**	-20.4	-20.05	-19.25	-18.03
HDCA Legos Knights	-5.84	-10.14	-13.13	-15.15	16.34	**-16.76**	-16.54	-15.36	-13.67	-11.30
Average	-9.38	-14.23	-17.48	-19.61	-15.35	**-21.21**	-21.01	-19.25	-18.86	-17.08

Table 7.1: BD-Rate gains of a increasing the QP in layers compared to using a fixed QP.

The performance of each $QP_b + N$ interaction was calculated using the bjøntegaard metric (BD-Rate) over fixed QP values and the results are shown in Table 7.1. It can be noticed that the QP increase in layers can significantly improve the performance of the compression method. With a single increased of one (QP+1), a significant gain in the bitrate is obtained. The QP+1 scheme can encode an LF on average with -9% less bits compared to fixed QP. It can be observed that the compression efficiency increases as the difference between the QP of the first and other layers increases. However, the compression efficiency only increases up to a certain point, after

an increase in the QP of +6 the compression efficiency begins to decrease. Most notably all the test images obtained the best improvement in the QP+6 test. The QP+6 achieved a average of 21% bitrate savings when compared to fixed QP.

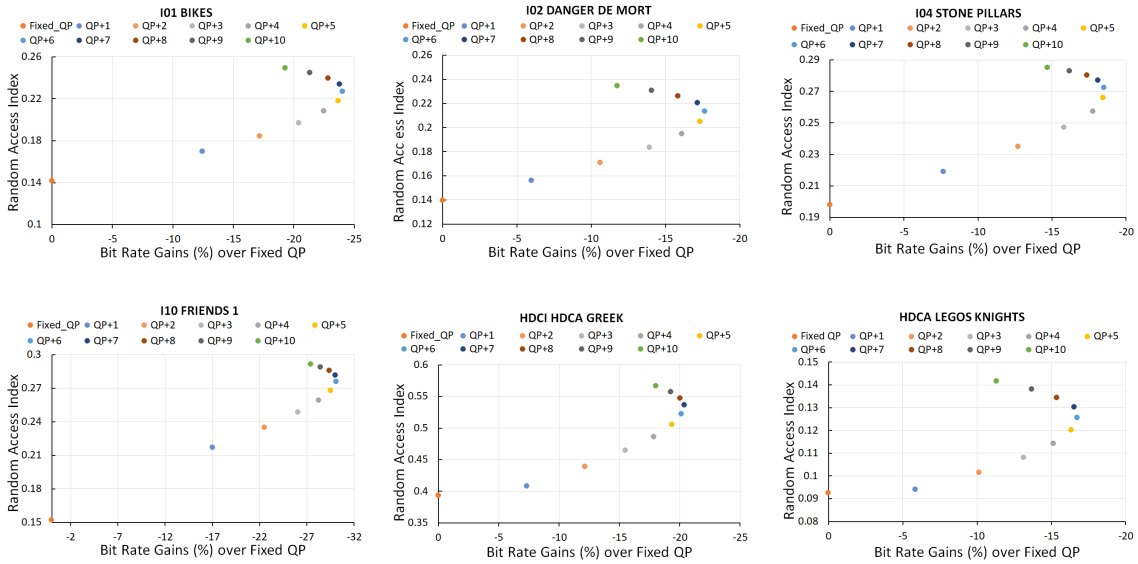


Figure 7.3: Bit-rate savings and random access index of QP optimization in layers.

The effects of the QP difference between layers in the random access performance of the compression method are shown in figure 7.3. The figure presents the bitrate gains versus the random access index of each test. Once again it can be observed the compression efficiency increases as the difference of the QP between the first and the other layers increases. The obtained results form a curve in the graph, where the inflection point matches the most compression efficient test, the QP+6 test. After the QP+6 point, the compression efficiency begins to decrease. As expected the QP difference between layers worsens the random access performance. An increase in the difference of QP between the first and other layers increases the random access index. As anticipated even when the compression efficiency begins to decrease, the random access performance keeps getting worse. This happens because the size of the reference pictures keeps increasing. Therefore all results obtained with a difference greater than +6 QP between layers are not only worse at effectively compressing LF images, but also have worse random access performance.

7.4 QP optimization in Levels

Since inside a layer some pictures will be used as a reference more often than others, the second research approach was to vary the QP values of the pictures within the layer. The theory is the same as before, the pictures used more often as a reference should have the lowest QP values (better quality). A level system was designed to assign the QP values to each picture. The level system will assign each picture within a GOP to a level, and each level will determines the QP value. The level system will be applied to all the layers in an equal manner. The pictures used

more times as a reference will belong to a lower level and the pictures which are never used as a reference will belong to the highest level. The QP will increase by a factor of k each level. A second heuristic algorithm based on the levels concept was designed to optimize the encoding. The algorithm will gradually increased the k factor each test and compare the results. Two versions of a level system were developed one with three levels and one with five levels presented next:

Three Level System (3LS) : A simple system with three levels. The number of times the picture is used as a reference to predict other pictures ($Nr\_References$) is the deciding factor to determinate the level. The more times the pictures is taken as reference the lower is the level assigned. The following formula was designed to calculate the QP of each picture:

$$Level_{3LS} = 2 - \frac{Nr\_References}{2},$$
$$QP_{picture} = QP_b + Level_{3LS} * k,$$
where k is the increase of QP per level

For example, in a test using the 3LS if $k = 2$ the pictures will be encoded four times with the following QP values:
QP_b=30: Levels 0,1 and 2 are coded with QP=30, QP=32 and QP=34 respectively.
QP_b=25: Levels 0,1 and 2 are coded with QP=25, QP=27 and QP=29 respectively.
QP_b=20: Levels 0,1 and 2 are coded with QP=20, QP=22 and QP=24 respectively.
QP_b=15: Levels 0,1 and 2 are coded with QP=15, QP=17 and QP=19 respectively.

Figure 7.4 shows the assignment of pictures within a GOP of the *Variant 1* coding structure to levels by the 3LS.

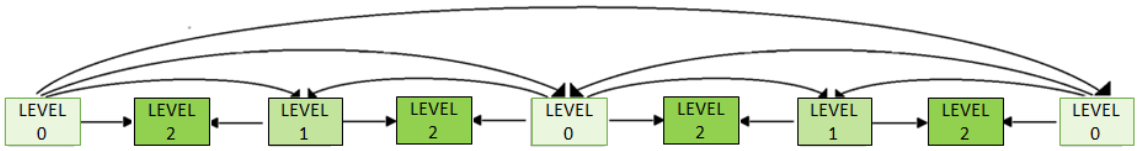


Figure 7.4: Tree Level System

Five Level System (5LS) : The 5LS has two additional levels, created for the I-pictures and P-pictures. The first picture of GOP is coded as intra and will be the only picture in the first level (Level 0). The second encoded picture takes only the intra-picture as reference (P-picture) and will be the only picture assigned to the second level (Level 1). After these two pictures, the renaming pictures are assigned, in a similar way as the 3LS. The number of times a picture is taken as reference ($Nr\_References$) is used to determinate the level, excluding the first two levels. The 5Ls can be expressed with the following equations:

$$Level_{5LS} = \begin{cases} 0, & \textit{if I-Pictures.} \\ 1, & \textit{if P-Pictures.} \\ 4 - \frac{Nr\_References}{2}, & \textit{otherwise.} \end{cases}$$

$$QP = QP_b + Level_{5LS} * k$$

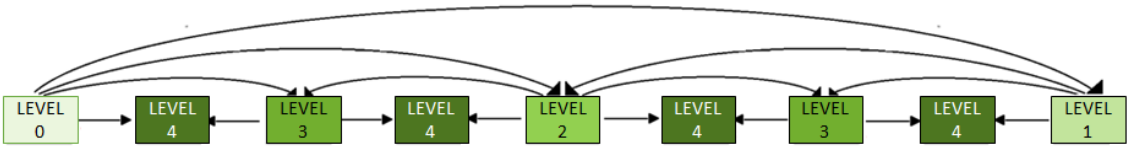


Figure 7.5: Five Level System

Figure 7.5 shows the assignment of pictures within a GOP to levels by the 5LS.

The second heuristic algorithm was designed to find the best increase of QP between levels. The algorithm performs tree tests, each test encodes the LF four times. For each test, the QP value of each picture is assigned conforming to the level system (3LS or 5LS) where k starts at one and is increased by one each test. The pseudo-code of the QP optimization in levels heuristic algorithm for the 3LS is presented next (an equal algorithm is used for the 5LS).

Algorithm 2 QP optimization in levels heuristic algorithm (3LS)

```
for k=1:1:3 do
    for Q_b = 15 : 5 : 30 do
        for all pictures do
            Calculate Level_3LS   (level of each picture)
            QP_Picture = QP_b + Level_3LS * k
            end
            Perform Codification
        end
        Save results as QP_3LS_k
    end
```

The results obtained from the difference of QP in levels are presented in Table 7.2.

It can be observed that the QP optimization in levels achieved lower bitrate savings compared to the QP optimization layers. Nonetheless, the results are mostly positive and show that difference of QP in levels improves the efficiency of the compression method. An average of -11% bitrate gains can be achieved using the 5LS with an increase of one QP by level.

It was also noticed that for both the 3LS and 5LS a high QP difference between levels (k=1,2) produces on average worse results when compared to a small difference (k=1). This shows that the proposed compression method works better with a small differences of the QP between the levels. This effect is more noticeable in the 5LS, where there is a greater loss in compression efficiency from using k=1 to k=3 compared to the 3LS.

Image	BD-Rate Gains (%) vs Fixed QP increase of QP in levels					
	3LS			5LS		
	k=1	k=2	k=3	k=1	k=2	k=3
I01 Bike	-9.36	-9.43	-6.45	**-14.54**	-5.74	8.48
I02 Danger de Mort	-1.61	0.03	5.37	**-4.95**	4.76	26.27
I04 Stone Pillars	-2.60	-1.70	11.22	**-6.74**	-2.16	10.55
I10 Friends	-13.18	-13.86	-12.40	**-20.56**	-18.46	-7.02
HDCA HDCI Greek	-4.49	-6.023	-4.97	**-11.372**	-9.08	7.00
HDCA Legos Knights	-5.24	-4.13	-1.62	**-8.04**	-4.57	10.80
Average	-6.08	-5.85	-1.46	**-11.04**	-5.9	9.35

Table 7.2: BD-Rate Gains of 3LS and 5LS compared to using a fixed QP

7.5 QP optimization in Layers and Levels

In this section, we search for the best QP values for both layers and levels. The QP value of each picture is assigned according to a level system (3LS or 5LS) and is additionally increased if the pictures belong to second, third or fourth layers. The results were obtained by a third heuristic algorithm created by merging the other two previous algorithms. Figure 7.6 presents the bitrate gains versus the random access index of the results obtained form each test.

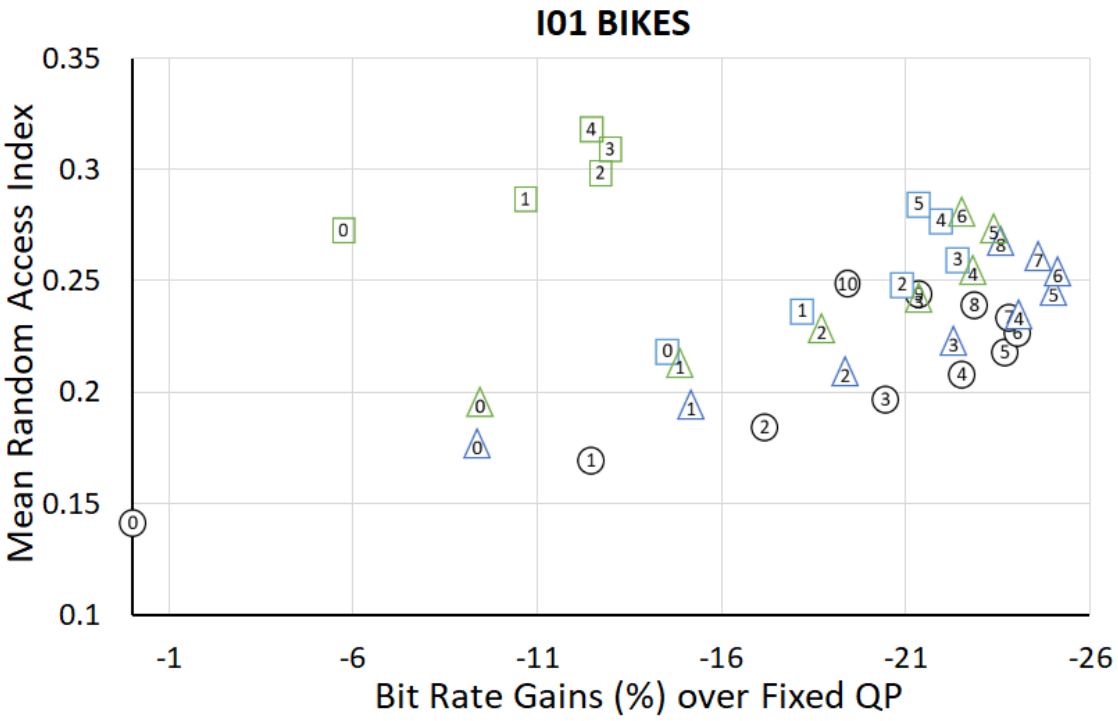


Figure 7.6: Trade-off results obtained using QP dimensioning in both layers and levels.

The key in the figure 7.6 should be read in the following way:

QP+N refers to using only QP dimensioning in layers.

The QP_3LS_k=a+N refers to the 3LS and layers dimensioning

The QP_5LS_k=a+N refers to the 5LS and layers dimensioning

where:

a is the increase of QP by level.

N is the increase of QP in the second, third and fourth layer.

The N value of each result is represented inside the marker.

Since the results obtained from all the LF images have similar a behavior, this section presents only the results for the *I01 Bikes* LF image as they can represent the results from the other LF images. The results of the remaining LF test images are made available in Annex C.

It can be observed that the compression method with a QP level dimensioning (3LS/5LS) reacts the same way to QP difference in layers as a scheme without QP level dimensioning. This is the compression efficiency increases as the difference between the QP in layers increases, up to a certain point. After this point, the compression efficiency begins to decrease. Both the 3LS and 5LS with k=1 or k=2 follow this behavior. The obtained results produce curves in the trade-off figure, where the inflection point of each curve is the most compression efficient QP dimensioning in layers for that specific level system and k factor.

Looking at the trade-off results in the figure 7.6 the 3LS obtains higher bitrate savings for lower random access index values in comparison to the 5LS. Therefore it can be concluded that the 3LS provides an overall better performance compared to the 5LS. The most compression efficient scheme is the 3LS with an increase by level of one (k=1) and an increased of +6 QP in layers. This scheme achieved 25.12% bitrate savings over fixed QP, for the *I01 Bikes* light field image. For all the LF test images the QP_3LS_k=1+6 was consistently the most compression efficient scheme. Therefore the study was successful. An optimization based on dynamically adjusting the QP for each frame was developed that provides significant improvement for our compression method (QSpiral and Variant 1)

It can be seen that it is not possible to simultaneously increase the encoding efficiency and the random access performance. Although the 3LS with k=1 achieved the highest compression efficiency values, it operates consistently at worst random access index values compared to the scheme without level optimization (QP+N). Therefore the decision of the preferred scheme (increase in layers + level system) would vary according to the needs of the user.

8

Conclusion

This chapter concludes our dissertation. The conclusions are subdivided in two sections, the first one regards the developed compression method. The second one regards the optimization of the compression method based on choice of the quantization parameter.

8.1 Conclusion - Compression Method

This dissertation devises a new method to efficiently compress light field images. These images have a large amount of data and previous compression solutions based on conventional image encoders did not perform well. Therefore, new strategies to compress light field images are relevant and timely. In this thesis, a compressing method based on pseudo-sequence coding using video encoders was developed and successfully applied to encode light field images. The image compression method developed in this work was based on video encoders and it is inspired in previous works published in the literature. From other works, it was perceived that the multi-view extension of HEVC could be advantageous in pseudo-sequences schemes. With the goal to process light fields in applications where speed is important, the compression method was optimized to maximize the compression efficiency but also able to provide random access and scalability.

The method presented in this dissertation is based on the Multi-View extension of HEVC. The proposed method first decomposes the light field image into sub-apertures images, which are subsequently used by a novel scan order, denoted Quadratic Spiral, to form multiple pseudo-sequences to be encoded by an MV-HEVC encoder. In the encoder the pseudo-sequences are encoded according to a coding structure. The proposed coding structure was specifically designed to improve the random access performance, while still exploiting the temporal and inter-view correlations. Furthermore using the layered structure of the MV-HEVC, a scalable representation of the light field can be decoded. The scalable representation is composed of periodic patterns of sub-apertures, the patterns are generated by the way the scan order arranges the sub-apertures in pseudo-sequences.

In order to evaluate the developed compression method the test conditions defined by JPEG Pleno Group were followed and two extra scan orders the Lines and Serpentine scan orders were defined to be used as comparisons. It was concluded that the proposed method is an efficient compression method for light fields, as it achieves PSNR gains as high as 7.8 dB over reference

JPEG for plenoptic images, which is in accordance with the current state-of-art methods. Although the QSpiral scan order is not the most effective compression scan order, the experimental results prove the design choices of the QSpiral order improved the random access performance. The compression method developed in this dissertation is used for the elaboration of a paper[7] used in the 8-th European Workshop on Visual Information Processing.

In summary, the main objectives of this dissertation were fulfilled, an efficient compression method of LF images with good random access performance and scalability was developed and successfully demonstrated.

8.2 Conclusion - QP Optimization

A study of the effects of dynamically adjusting the QP values for each frame was made in this dissertation. The objective of this study was to find the best QP value for each picture that optimizes the efficiency of the proposed compression method. A heuristic study to find the most suitable QP values was made. The study followed two approaches: (i) The first approach was to gradually increase the QP values for the second, third and fourth layers while keeping the first layer QP value constant. (ii) The second approach was to assign a level to all pictures within a layer,

From the study, it was concluded that the proposed compression method has better performance with small differences of QP between pictures, for bigger differences of QP the performance of the method deteriorates. It was also found that the QP dimensioning in layers is more effective than the QP dimensioning using a level system. However, both can be combined for the best optimization model, which is an increase of +6 QP in the second, third and fourth layers combined with a tree level system with an increase of +1 QP per level. The proposed QP optimization achieved significant improvements over fixed QP values with gains as high as -22% bitrate.

In conclusion, the study was successful and the secondary objective was fulfilled. An optimization based on dynamically adjusting the QP for each frame was developed that provides significant improvement for our compression method.

8.3 Future Work

There are some aspects of the presented work carried out in this thesis which may be further explored in a number of aspects that are detailed as follows:

(i) Improvements of the QSpiral scan order.

The developed Quadratic Spiral scan order is very rigid. It requires an even number of rows and columns to properly function. It also forms a fixed number of pseudo-sequences organized in 2x2 blocks of sub-apertures. A more flexible version could be developed. A version where the number of pseudo-sequences and the shape of the block could be adaptably chosen for each LF

image leading to a better performance. As the QSpiral could be made more flexible the coding structure would have to follow the changes in order to keep compression efficiency.

(ii) QP Optimization

The most noticeable aspect that could be further explored is the QP optimization. More experimental tests with a larger dataset should be made to confirm the results obtained. The QP optimization presented in this dissertation, although with good results, is very simple, it should be seen more like the first steps of a future more elaborate optimization. How the QP difference between the reference and predicted pictures affects the distortion and compression rate should be further study and mathematical described. Finally additional research of optimization models, like the ones used in rate control, should be made with the goal of designed one especially for our compression method.

Bibliography

[1] Todor Georgiev. New results on the plenoptic 2.0 camera. *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, pages 1243–1247, 2009.

[2] Matching Light, Field Datasets, From Plenoptic, Waqas Ahmad, Luca Palmieri, and Reinhard Koch. Marten Sjostrom ¨ Department of Information Systems and Technology , Mid Sweden University , Sundsvall , Sweden Department of Computer Science , Christian-Albrechts-University at , Kiel , Germany. pages 18–21, 2018.

[3] Mathias Wien. *Signals and Communication Technology High Efficiency Video Coding*.

[4] Yan-Fei SHEN, Jin-Tao LI, Zhen-Min ZHU, and Yong-Dong ZHANG. *High Efficiency Video Coding*, volume 36. 2014.

[5] Feng Dai, Jun Zhang, Yike Ma, and Yongdong Zhang. Lenselet image compression scheme based on subaperture images streaming. *Int. Conf. Image Process. ICIP*, pages 4733–4737, 2015.

[6] Irene Viola, Martin Rerabek, Tim Bruylants, Touradj Ebrahimi, Fernando Pereira, and Peter Schelkens. PPT-ICME Grand Challenge on Light Field Image Compression. (October), 2016.

[7] (Accepted) Gomes,Pedro, da Silva Cruz, Luis A. (Accepted) Pseudo-sequence Light Field Image Scalable Encoding with Improved Random Access. In *8-th Eur. Work. Vis. Inf. Process.*, 2019.

[8] Shing Chow Chan. Plenoptic Function BT - Computer Vision. *Comput. Vis.*, (Chapter 7):618–623, 2014.

[9] Marc Levoy and Pat Hanrahan. Light field rendering. *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '96*, pages 31–42, 1996.

[10] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael Cohen. The lumigraph. *Proc. of SIGGRAPH 96*, 96, 08 2001.

[11] Robert C. Bolles, H. Harlyn Baker, and David H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *Int. J. Comput. Vis.*, 1(1):7–55, 1987.

[12] Ren Ng. *Digital Light Field Photography*. Phd thesis, Stanford University, CA, USA, July 2006.

[13] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine. Superresolution with the focused plenoptic camera. page 78730X, 2011.

[14] The (new) standford light field archive. `http://lightfield.standford.edu/`. Accessed: 02-02-2019.

[15] The Digital Michelangelo Project. `https://accademia.stanford.edu/mich/`. Accessed: 02-02-2019.

[16] R Ng, M Levoy, M Brédif, and G Duval. Light field photography with a hand-held plenoptic camera. *Stanford Univ. Comput. Sci. Tech Rep. CSTR 2005-02*, pages 1–11, 2005.

[17] Christian Perwass and Lennart Wietzke. Single lens 3d-camera with extended depth-of-field. *Proc. SPIE*, 8291:4–, 02 2012.

[18] Pelican Imaging Array Camera: Light Field Module For Smartphones. http://lightfield-forum.com/light-field-camera-prototypes/pelican-imaging-array-camera-light-field-module-for-smartphones/. Accessed: 05-08-2019.

[19] Noah Bedard, Timothy Shope, Alejandro Hoberman, Mary Ann Haralam, Nader Shaikh, Jelena Kovačević, Nikhil Balram, and Ivana Tošić. Light field otoscope design for 3D in vivo imaging of the middle ear. *Biomed. Opt. Express*, 8(1):260, 2017.

[20] Gergo Bohner and Gatsby Tea Talk. Light Field Microscopy The Light Field. 25(3):1–11, 2014.

[21] Donald G. Dansereau, Glenn Schuster, Joseph Ford, and Gordon Wetzstein. A wide-field-of-view monocentric light field camera. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 2017-Janua:3757–3766, 2017.

[22] R. Olsson, M. Sjostrom, and Y. Xu. A combined pre-processing and h.264-compression scheme for 3d integral images. In *2006 International Conference on Image Processing*, pages 513–516, Oct 2006.

[23] Gary J. Sullivan, Jens Rainer Ohm, Woo Jin Han, and Thomas Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.*, 22(12):1649–1668, 2012.

[24] Standardization Sector and Itu. High Efficiency Video Coding. Recommendation ITU-T H.265. 265, 2018.

[25] Hanan Samet. The Quadtree and Related Hierarchical Data Structures. *ACM Comput. Surv.*, 16(2):187–260, 1984.

[26] JVT-3V. MV-HEVC and 3D-HEVC Reference Software. `https://hevc.hhi.fraunhofer.de/mvhevc`. Accessed: 05-01-2019.

[27] Miska M. Hannuksela, Ye Yan, Xuehui Huang, and Houqiang Li. Overview of the multiview high efficiency video coding (MV-HEVC) standard. *Proc. - Int. Conf. Image Process. ICIP*, 2015-Decem:2154–2158, 2015.

[28] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light Field Image Processing: An Overview. *IEEE J. Sel. Top. Signal Process.*, 11(7):926–954, 2017.

[29] Harini Priyadarshini Hariharan, Tobias Lange, and Thorsten Herfet. LOW COMPLEXITY LIGHT FIELD COMPRESSION BASED ON PSEUDO-TEMPORAL CIRCULAR SEQUENCING. *2017 IEEE Int. Symp. Broadband Multimed. Syst. Broadcast.*, pages 1–5.

[30] Dong Liu, Lizhi Wang, Li Li, Xiong Zhiwei, Feng Wu, and Zeng Wenjun. Pseudo-sequence-based light field image compression. *2016 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2016*, pages 3–6, 2016.

[31] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[32] Li Li, Zhu Li, Bin Li, Dong Liu, and Houqiang Li. Pseudo-Sequence-Based 2-D Hierarchical Coding Structure for Light-Field Image Compression. *IEEE J. Sel. Top. Signal Process.*, 11(7):1107–1119, 2017.

[33] C . Perra and P . Assuncao. HIGH EFFICIENCY CODING OF LIGHT FIELD IMAGES BASED ON TILING AND PSEUDO-TEMPORAL DATA ARRANGEMENT C . Perra Department of Electrical and Electronic Engineering University of Cagliari Italy P . Assuncao Instituto de Telecomunicacoes IPLeiria Portugal. page 4, 2000.

[34] W. Ahmad, R.Olsson, and M. Sjöström. Interpreting plenoptic images as multi-view sequences for improved compression. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4557–4561, Sep. 2017.

[35] Hadi Amirpour, Antonio Pinheiro, Manuela Pereira, and Mohammad Ghanbari. Fast and Efficient Lenslet Image Compression. pages 1–8, 2019.

[36] Joint Bi-level Image, Experts Group, Joint Photographic, and Experts Group. Coding of Still Pictures. 1:1–20, 2017.

[37] Xu-guang Zuo and Lu Yu. Long-term prediction for hierarchical-b-picture-based coding of video with repeated shots. *Frontiers of Information Technology & Electronic Engineering*, 19(3):459–470, Mar 2018.

[38] HDCI 4D Light Field Dataset. `http://hci-lightfield.iwr.uni-heidelberg.de/`. Accessed: 02-02-2019.

[39] Rocco Goris and Robert Brondijk. ORGANISATION INTERNATIONALE DE NORMALISATION 1 Introduction 2 Tone Mapping process. 3(YCbCr 1886):3–5, 2015.

[40] D G Dansereau, O Pizarro, and Williams. Light Field Toolbox for Matlab. 2015.

[41] G.Bjontegaard. Calculation of average psnr diferences betwen rd-curves (VCEG-M33), Apr. 2001.

[42] Irene Viola, Martin Rerabek, and Touradj Ebrahimi. Comparison and Evaluation of Light Field Image Coding Approaches. *IEEE J. Sel. Top. Signal Process.*, 11(7):1092–1106, 2017.

[43] Amit Anand Ramanand, Ishfaq Ahmad, and Viswanathan Swaminathan. A survey of rate control in HEVC and SHVC video encoding. *2017 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2017*, (July):145–150, 2017.

[44] Woong Lim and Sooyoun Lee. Multi-view Rate Control based on HEVC for 3D Video Services. *J. Inst. Electron. Inf. Eng.*, 50(8):245–249, 2013.

[45] Woong Lim, Ivan V. Bajic, and Donggyu Sim. QP initialization and interview MAD prediction for rate control in HEVC-based multi-view video coding. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, 7122(4):2045–2049, 2013.

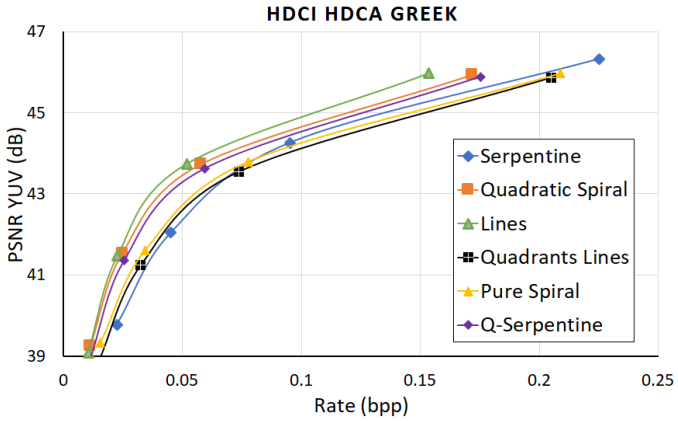
[46] Per-Title Encode Optimization - Netflix Tech blog . `https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2`. Accessed: 01-09-2019.

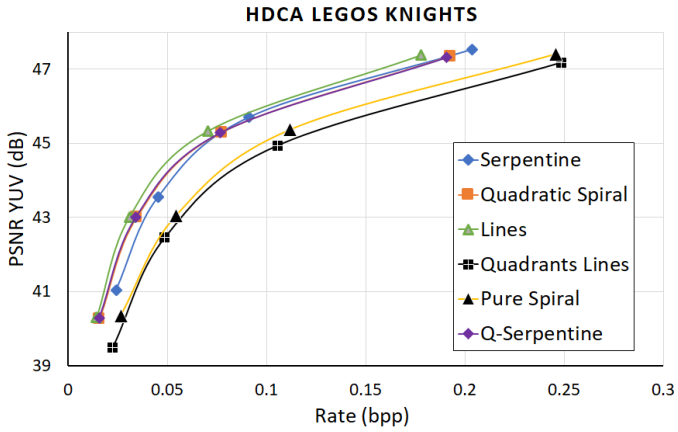
A

Other Scan Order Results

Compression efficiency results from the discarded scan orders .

Image	PSequence	BD-Rate Gains vs Serpentine (%)
HDCI HDCA Greek	QSpiral	-29.29
	Lines	-35.39
	QSerpentine	-23.48
	QuadLines	-1.92
	PSpiral	-7.19
HDCA Legos Knights	QSpiral	-6.98
	Lines	-16.59
	QSerpentine	-7.81
	QuadLines	47.18
	PSpiral	36.26





B

Matlab Scripts

```
% Author: Pedro de Medeiros Gomes
% script of the Quadratic Spiral Scan order
% Input: LF image
% Output: Pseudo-Sequences rearranged by Quadratic Spiral.

clear all; clc; close all;
delete('Sequencias/Pozan_lab1_30x30/
    Pozan_1936x1288_QuadEspiral_8bit_420_vista_1.yuv');
fid=fopen('Sequencias/Pozan_lab1_30x30/
    Pozan_1936x1288_QuadEspiral_8bit_420_vista_1.yuv','a');

%Define the center
l=14;
c=14;
%example for the first pseudo-sequence
%to form the other sequences change the center coordinates.
dir=0; % 0 up, 1 right | 2 down | 3 left
for i=1:1:225 %nr de frames
    if (c < 10 & l<10)
     path= sprintf('/media/one_tera/Pedro_Data/poznanlab1_TAU
        /00%d_00%d.ppm',l,c); end
    if (c > 9 & l<10)
      path= sprintf('/media/one_tera/Pedro_Data/poznanlab1_TAU
        /00%d_0%d.ppm',l,c); end
    if (c < 10 & l> 9)
      path= sprintf('/media/one_tera/Pedro_Data/poznanlab1_TAU
        /0%d_00%d.ppm',l,c); end
    if (c > 9 & l> 9)
      path= sprintf('/media/one_tera/Pedro_Data/poznanlab1_TAU
        /0%d_0%d.ppm',l,c); end
    img= imread(path);
```

```
26      img=im2uint8(img);
27
28      %Write frame to video sequence
29              img_yuv=rgb2ycbcr(img);
30              img_yuv_420 = downsample(img_yuv);
31              Y=img_yuv_420(1);
32              U=img_yuv_420(2);
33              V=img_yuv_420(3);
34              Y=cell2mat(Y);
35              U=cell2mat(U);
36              V=cell2mat(V);
37              Y=rot90(Y);
38              U=rot90(U);
39              V=rot90(V);
40              fprintf('l = %d  c= %d    i=%d\n', l  ,c,i)
41              fwrite(fid,Y,'uint8');
42              fwrite(fid,U,'uint8');
43              fwrite(fid,V,'uint8');
44
45      if( dir ==0 ) l=l-2;end %up
46      if( dir ==1 ) c=c-2;end %left
47      if( dir ==2 ) l=l+2;end % down
48      if( dir == 3 ) c=c+2;end %right
49
50      if(i==1 |i == 9 | i==25 | i==49 | i==81 | i==121 | i==169)
           dir =1; end
51      if(i==2 | i == 12 | i==30 | i== 56 |i==90 | i==132 | i==
           182) dir =2; end
52      if(i==4 | i== 16 | i== 36 | i==64 |i==100 |i==144 |i==196)
           dir =3; end
53      if(i==6 | i== 20 | i==42 | i==72 | i==110 | i== 156 | i
           ==210) dir =0; end
54  end
```

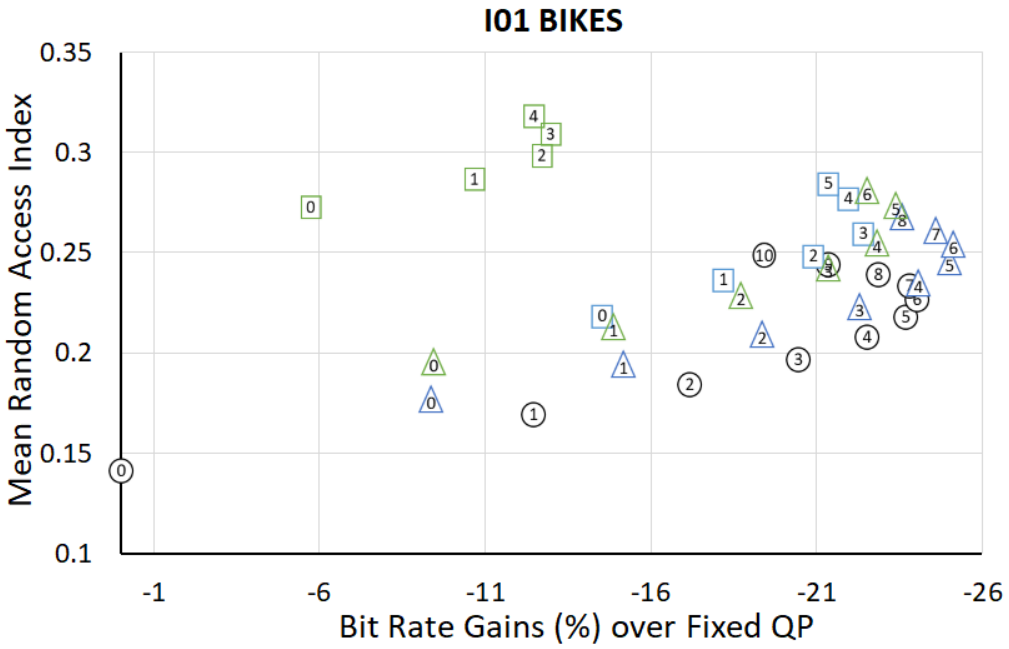
```
% Author: Martin Rerabek (martin.rerabek@epfl.ch)
% Copyright(c) Multimedia Signal Processing Group (MMSPG),
%              Ecole Polytechnique Federale de Lausanne (EPFL)
%              http://mmspg.epfl.ch
% All rights reserved.

% script for RGB 444 to YCbCr 444 color space conversion
% Input: RGB image in uint8 representation [0-255]

% Output: YCbCr image in uint8 representation [0-255]

function [ycbcr] = rgb2ycbcr(rgb)

M = [    0.212600  0.715200  0.072200 ;
        -0.114572 -0.385428  0.500000 ;
         0.500000 -0.454153 -0.045847   ];

ycbcr = reshape(double(rgb)/255, [], 3) * M';
ycbcr = reshape(ycbcr, size(rgb));
ycbcr(:,:,1) = 219*ycbcr(:,:,1) + 16;
ycbcr(:,:,2:3) = 224*ycbcr(:,:,2:3) + 128;
ycbcr = uint8(ycbcr);
```

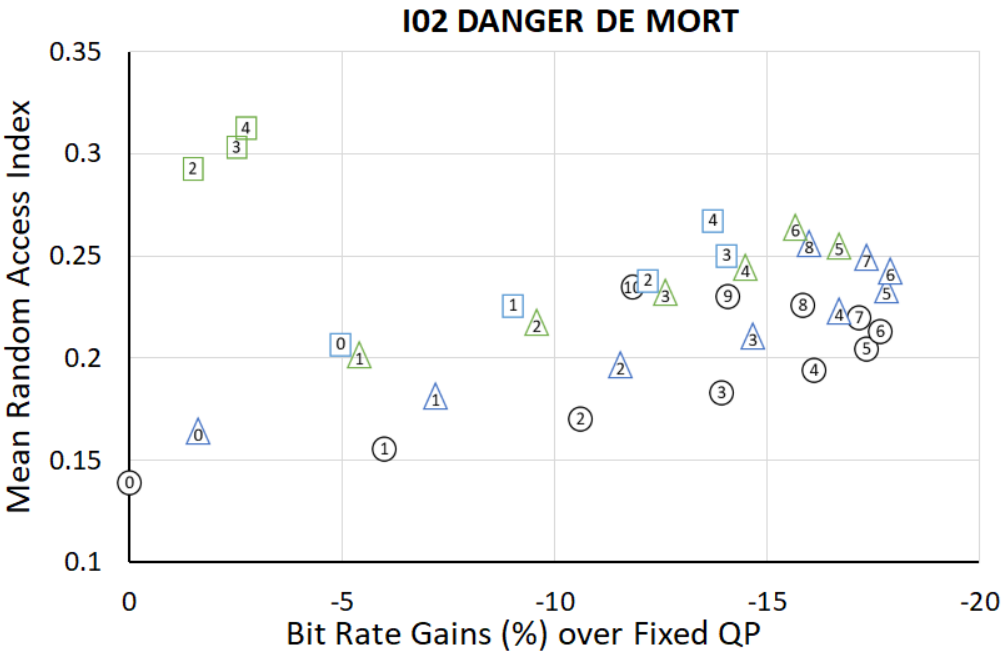
```
% Author: Martin Rerabek (martin.rerabek@epfl.ch)
% Copyright(c) Multimedia Signal Processing Group (MMSPG),
%              Ecole Polytechnique Federale de Lausanne (EPFL)
%              http://mmspg.epfl.ch
% All rights reserved.

% script performing 444 to 420 downsampling

% Input: image in YCbCr 444 sampling 8 bits in uint8
    representation [0-255]

% Output: 3x1 cell, each cell item contatin one channel Y,Cb, Cr
% coresponding to 420 solor sampling

function [out] = downsample(in)

out = cell(3,1);
for i=1:3
    out{i} = in(:,:,i);
end

for i=2:3
    tmp = imfilter(double(out{i}), [1 6 1], 'replicate', 'same')
        ;
    out{i} = tmp(:,1:2:end);
end

for i=2:3
    tmp = imfilter(out{i}, [0 ; 4 ; 4], 'replicate', 'same');
    out{i} = uint8((tmp(1:2:end,:) + 32) / 64);
end
```

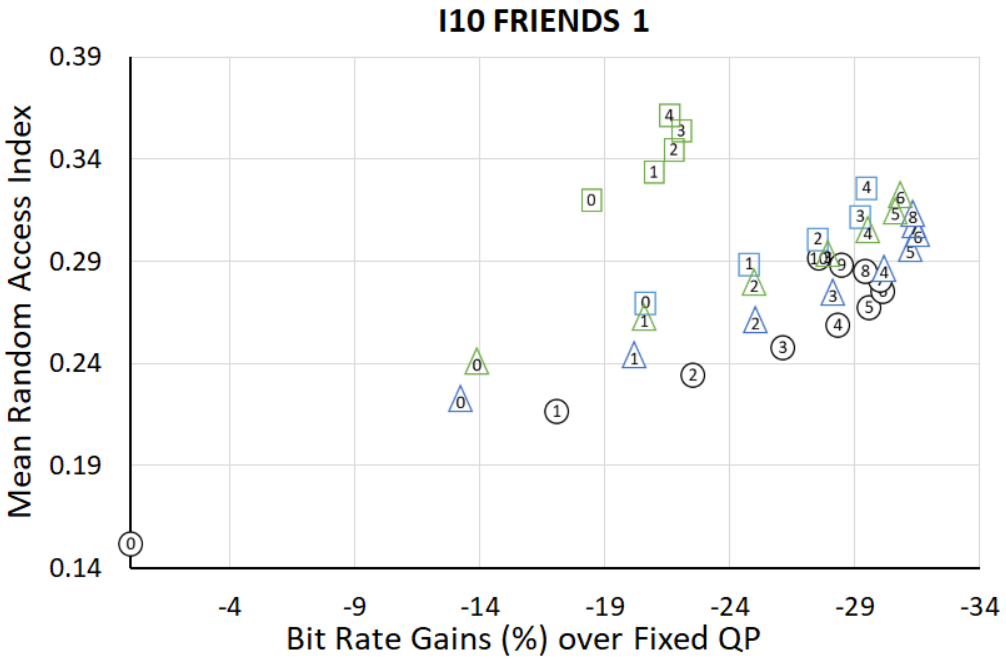
C

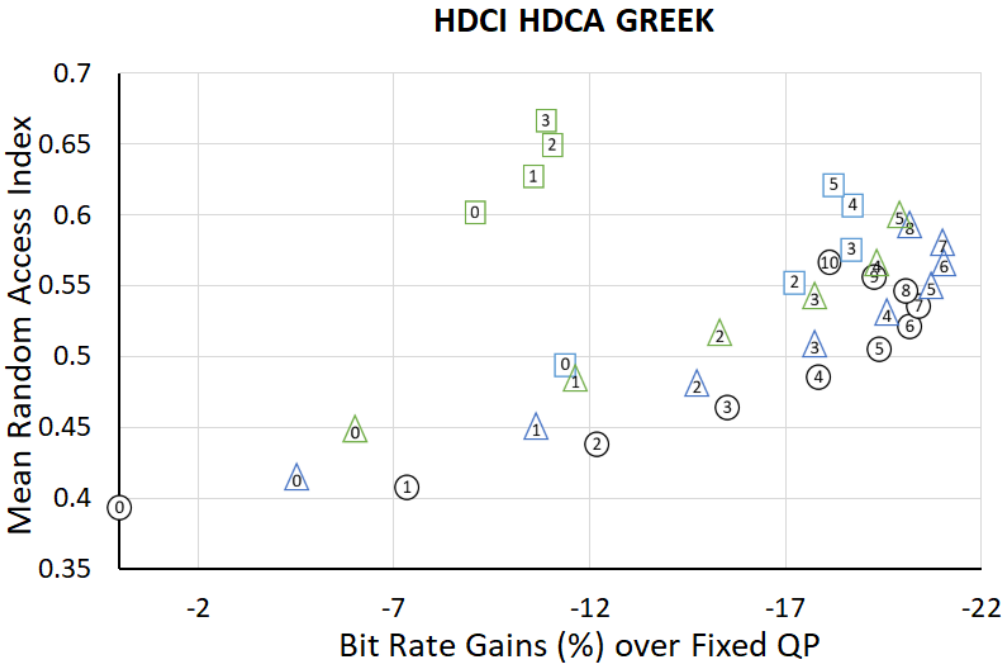
QP Optimization Results

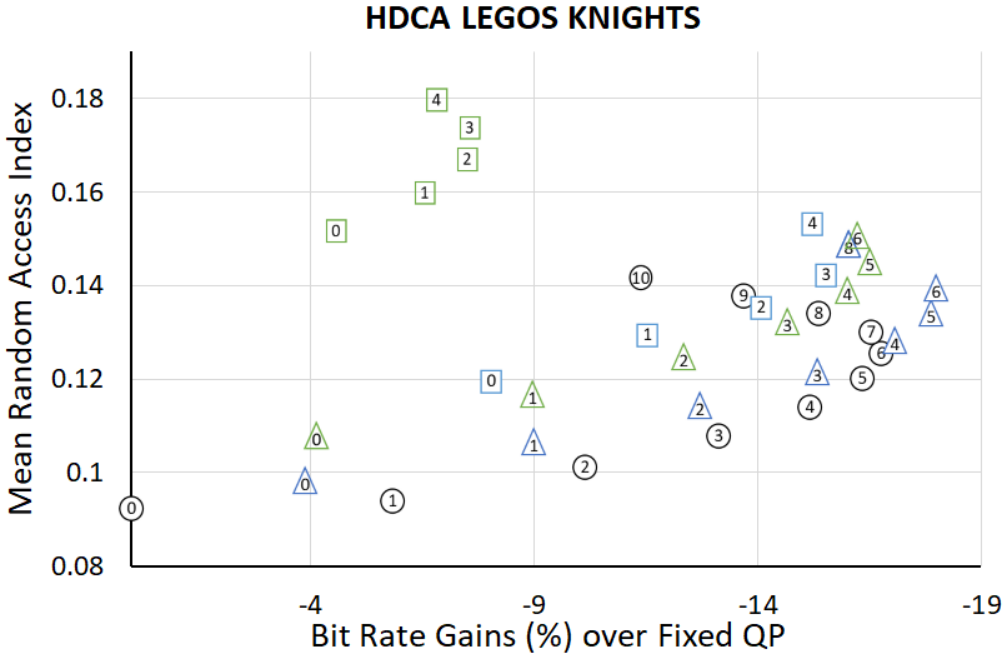
Trade-off results between compression efficiency and random access for all the test images with different QP dimensioning.




I02 DANGER DE MORT
Mean Random Access Index
0.35
0.3
0.25
0.2
0.15
0.1
0
-5
-10
-15
-20
Bit Rate Gains (%) over Fixed QP


I10 FRIENDS 1
Mean Random Access Index
0.39
0.34
0.29
0.24
0.19
0.14
-4
-9
-14
-19
-24
-29
-34
Bit Rate Gains (%) over Fixed QP


HDCI HDCA GREEK
Mean Random Access Index
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
-2
-7
-12
-17
-22
Bit Rate Gains (%) over Fixed QP


HDCA LEGOS KNIGHTS
Mean Random Access Index
0.18
0.16
0.14
0.12
0.1
0.08
-4
-9
-14
-19
Bit Rate Gains (%) over Fixed QP