Ministry of Higher Education and Scientific Research
وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba

Faculty of Technology

Department of Electronics

جامعة باجي مختار – عنابة
كلية التكنولوجيا
قسم الإلكترونيك

Thesis

Submitted to obtain the diploma of

Doctorate Third Cycle

Field: Telecommunication

Specialty: Multimedia and Digital Communications

By :
BELBEL Amel

Title :

Codage et compression de vidéo multi-vues robuste aux erreurs de transmission

Thesis defended on ..... in front of the jury composed of:

N°	Name and Surname	Grade	Establishment	Quality
01	**FEZARI Mohamed**	Prof.	Badji Mokhtar University – Annaba	President
02	**BEKHOUCH Amara**	Prof.	Mohamed Cherif Massadia University – Souk Ahras	Supervisor
03	**DOGHMANE Noureddine**	Prof.	Badji Mokhtar University – Annaba	Co-Supervisor
04	**BOUMEHREZ Farouk**	MCA	Abbes Laghrour University–Khenchela	Examiner
05	**BOUDEN Toufik**	Prof.	University of Jijel	Examiner
06	**KOUADRIA Nasreddine**	MCA	Badji Mokhtar University – Annaba	Invited

"ترميز وضغط الفيديو متعدد العروض قوي لأخطاء الإرسال "

الملخص:

استغل أحدث جيل من ترميز الفيديو متعدد (MV-HEVC) العروض بشكل فعال المعلومات الزائدة عن العرض البيني في مقاطع الفيديو متعددة العروض ووفر كفاءة ضغط عالية. ومع ذلك فإن زيادة المسافة بين الكاميرات أثناء الحصول على مقاطع فيديو متعددة العروض خلقت تباينًا بين العرض. أثر هذا الأخير على أوجه التشابه بين العرض وأثر على تقدير التباين الذي يتطلب تقديرًا فرديًا لكل وحدة تشفير. وبالتالي، أدى ذلك إلى زيادة التعقيد الحسابي لمشفر(MV-HEVC) .

في هذا البحث اقترحنا نهجًا محسنًا لتقدير التباين من خلال إزالة الإزاحة بين المشاهدات أثناء تشفير مكّن هذا الحذف من الإزاحة من تقليل نافذة البحث لتقدير التباين بناءً على حجم وحدات الترميز. بالإضافة إلى ذلك، قدمنا قرارًا سابقًا لتقسيم وحدات الترميز لتعزيز موثوقية تقدير الحركة. أظهرت النتائج التجريبية أن نهجنا المقترح قلل بشكل كبير من وقت التشفير، حيث تراوح من 20.31٪ إلى 44.35٪ في المتوسط، مع أدنى حد من تدهور الأداء.

علاوة على ذلك، قيمت هذه الأطروحة متانة MV-HEVC ضد أخطاء الإرسال من خلال تحليل حساسية تدفق البتات MV-HEVC في ظل ظروف معرضة للخطأ. على وجه التحديد، ركز التحليل على تأثير انتشار الخطأ في طبقة تجريد الشبكةb (NAL) . تناولت الدراسة نوعين من الوحدات في هذه الطبقة: طبقة ترميز الفيديو (VCL)وطبقة الترميز غير المرئية (Non-VCL)، مما يعرضهما لمعدلات خطأ متفاوتة في البتات. كشفت النتائج التجريبية أن (Non-VCL) والمنطقة النشطة VCL كانت البيانات المشفرة الأكثر حساسية في تدفق البتات، مع التأكيد على الحاجة إلى حمايتها لتقليل انتشار الخطأ في نقل الفيديو متعدد العروض.

كلمات مفتاحية: MVV، MV-HEVC، التعقيد الحسابي، توقع Interview ، تقدير التباين، تيار بت MV-HEVC ، وحداتNAL ، قوة الخطأ.

«Codage et compression de vidéo multi-vue robuste aux erreurs de transmission »

Résumé :

La dernière génération de codage vidéo multi-vues MV-HEVC a exploité efficacement les informations redondantes entre les vues dans les vidéos multi-vues et a fourni une efficacité de compression élevée. Cependant, l'augmentation de la distance entre les caméras lors de l'acquisition de vidéos multi-vues a créé une disparité entre les vues. Ce dernier a influencé les similitudes d'interview et a affecté l'estimation de la disparité qui a nécessité une estimation individuelle pour chaque unité de codage. Par conséquent, cela a augmenté la complexité de calcul du codeur MV-HEVC.

Dans cette recherche, nous avons proposé une approche améliorée pour estimer la disparité en éliminant le décalage entre les vues lors de l'encodage des entrevues. Cette suppression du décalage a permis de réduire la fenêtre de recherche pour l'estimation de disparité en fonction de la taille des unités de codage. De plus, nous avons introduit une décision antérieure de division des unités de codage (EDCS) pour améliorer la fiabilité de l'estimation de mouvement. Les résultats expérimentaux ont démontré que notre approche proposée réduisait considérablement le temps d'encodage, allant de 20,31 % à 44,35 % en moyenne, avec une dégradation minimale des performances.

De plus, cette thèse a évalué la robustesse de MV-HEVC contre les erreurs de transmission en analysant la sensibilité du flux binaire MV-HEVC dans des conditions sujettes aux erreurs. Plus précisément, l'analyse s'est concentrée sur l'effet de la propagation des erreurs dans la couche d'abstraction du réseau (NAL). L'étude a pris en compte deux types d'unités dans cette couche : la couche de codage vidéo (VCL) et la couche de codage non vidéo (non-VCL), en les soumettant à des taux d'erreur binaire variables. Les résultats expérimentaux ont révélé que la NAL VCL et la zone active non-VCL étaient les données codées les plus sensibles dans le flux binaire MV-HEVC, soulignant la nécessité de les protéger pour minimiser la propagation des erreurs dans la transmission vidéo multi-vues.

Mots clés: MVV, MV-HEVC, complexité de calcul, prédiction entre vues, estimation de la disparité, flux binaire MV-HEVC, unités NAL, robustesse aux erreurs.

« Coding and compression of multi-view video robust to transmission errors »

Abstract :

The latest multi-view video coding generation MV-HEVC exploited effectively the inter-view redundant information in multi-view videos and provided high compression efficiency. However, increasing the distance between cameras during multi-view videos acquisition created an inter-view disparity. This latter, influenced the inter-view similarities and affected the disparity estimation which required an individual estimation for each coding unit. Consequently, this increased the computational complexity of the MV-HEVC encoder.

In this research, we proposed an improved approach to estimate disparity by eliminating the offset between views during inter-view encoding. This elimination of the offset enabled a reduction in the search window for disparity estimation based on the size of the coding units. Additionally, we introduced an earlier decision of coding units splitting (EDCS) to enhance motion estimation reliability. Experimental results demonstrated that our proposed approach significantly reduced the encoding time, ranging from 20.31% to 44.35% on average, with minimal performance degradation.

Furthermore, this thesis evaluated the robustness of MV-HEVC against transmission errors by analyzing the sensitivity of the MV-HEVC bitstream under error-prone conditions. Specifically, the analysis focused on the effect of error propagation in the Network Abstraction Layer (NAL). The study considered two types of units in this layer: the Video Coding Layer (VCL) and the Non-Video Coding Layer (Non-VCL), subjecting them to varying bit error rates. The experimental findings revealed that the VCL NAL and the non-VCL active area were the most sensitive encoded data in the MV-HEVC bitstream, emphasizing the need to protect them to minimize error propagation in multi-view video transmission.

Key words : MVV, MV-HEVC, Computational Complexity, Inter-View Prediction, Disparity Estimation, MV-HEVC bitstream, NAL units, error robustness.

Acknowledgments

I would like to express my deepest appreciation and gratitude to the many individuals who have contributed to the successful completion of my PhD thesis. Without their support, encouragement, and guidance, this achievement would not have been possible. I am indebted to the following:

My supervisor, **"BEKHOUCH Amara"**, whose unwavering commitment to excellence, expertise in the field, and invaluable guidance played a pivotal role in shaping the direction and quality of this research. Your constant encouragement, constructive feedback, and tireless efforts have been instrumental in every aspect of this thesis. I am truly grateful for your mentorship and the invaluable lessons I have learned under your guidance.

I would like to express my sincere gratitude for **"DOGHMANE Noureddine"** unwavering support, encouragement, and belief in my abilities. Your mentorship has not only enriched my academic journey but has also played a significant role in shaping my growth as a researcher. I am truly grateful for the trust you have placed in me and the opportunities you have provided for me to learn and excel.

My family and friends deserve special recognition for their unwavering support, encouragement, and understanding during the ups and downs of this challenging endeavor. Especially, **"Mom"**, Your love, patience, and belief in me have been the cornerstone of my perseverance, and I am eternally grateful for your presence in my life.

Publications

Journal papers

- A.Belbel, A. Bekhouch, N. Doghmane, S. Harize, and N.Kouadriaa , "Improved inter-view correlations for low complexity MV-HEVC", Journal of Visual Communication and Image Representation, V (86), 2022, pp. 1–10. doi: 10.1016/j.jvcir.2022.103525

Conference papers

- A.Belbel, A. Bekhouch, N. Doghmane, S. Harize, and N.Kouadriaa , "Exploring the Use of Early Inter-view Offset Estimation for Low Complexity MV-HEVC", 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC), Japan, Okayama, 2020, pp. 1–6. doi: 10.1109/WPMC50192.2020.9309453.

- A.Belbel, A. Bekhouch, and N. Doghmane, "Error Robustness Assessment of MV-HEVC Encoded Data", International Conference on Advances in Electrical and Computer Engineering (ICAECE'2023), Tebessa, Algeria, 2023, pp. 1–6.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-Dimensional
3D-HEVC	Three-Dimensional HEVC
3DTV	Three-Dimensional Television
AMP	Asymmetric Partitions
AMVP	dvanced Motion Vector Prediction
AVC	Advanced Video Coding
AU	Access Unit
BER	Bit-Error Rate
BLA	Broken Link Access
CABAC	Context-Adaptive Binary Arithmetic Coding
CRA	Clean Random Access
CTC	Test Common Conditions
CTU	Coding Tree Unit
CU	Coding Unit
DCT	Discrete Cosine Transform
DCP	Disparity Compensation Prediction
DST	iscrete Sine Transform
DE	Disparity Estimation
EDCS	Earlier Decision for CU Splitting (EDCS)
EDE	Earlier Disparity Estimation

fps	frames per second
FVV	Free Viewpoint Video
EC	Error Concealment
ER	Error Resilience
GOP	Group Of Pictures
HEVC	High Efficiency Video Coding
HDR	High Dynamic Range
HLS	High-Level Syntax
HM	HEVC test Model
HTM	High-Throughput Multimedia
HVC	Hybrid Video Coding
IDR	Instantaneous Decoder Refresh
IEC	International Electro technical Commission
IQ	Inverse Quantization
IRAP	Intra Random Access Picture
IT	Inverse Transform
ITU-T	International Telecommunication Union Telecommunication
ISO	International Standard Organization
JCT-VT	Joint Collaborative Team on Video Coding
MCP	Motion Compensated Prediction
ME	Motion Estimation
MOP	Matrix Of Pictures
MOS	Mean Opinion Score
MPEG	Moving Pictures Experts Group
MV-HEVC	Multi-view High Efficiency Video Coding
MV	Motion Vector
MVC	Multi-view Video Coding
MVD	Multi-view Video-plus Depth
MVV	Multi-View Videos

NAL	Network Abstraction Layer
PLR	Packet Loss Rate
POC	Picture Order Count
PPS	Picture Parameter Set
PTL	Profiles, Tiers, and Level
PU	Prediction Unit
QP	Quantization Parameter
RD	Rate Distortion
RADL	Random Access Decodable Leading
RASL	RandomAccess Skipped Leading
SEI	Supplemental Enhancement Information
SMP	Symmetric Partitions
SPS	Sequence Parameter Set
TSA	Temporal Sublayer Access
TU	Transform Unit
STSA	Stepwise Temporal Sublayer Acces
UHD	Ultra-High Definition
VCL	Video Coding Layer
VPS	Video Parameter Set
VSP	View Synthesis Prediction
VVC	Versatile Video Coding
WPP	Wavefront Parallel Processing

General Introduction

Multi-view videos (MVV), also known as multi-angle or stereoscopic videos, are video contents captured from multiple simultaneous viewpoints. With the advent of technology, these videos have rapidly gained popularity as a legitimate alternative and extension to 2D video. MVVs offer a more immersive and interactive viewing experience, as well as higher video quality. They are also accessible to the general public through various channels such as the internet, cable and satellite transmission, as well as terrestrial broadcasting. The key technology for these different applications is multi-view video, also known as Multi-View Video (MVV). In this introduction, we will explore in more detail what multi-view video is, its advantages and applications, as well as the challenges it faces.

Multi-view video is a form of video captured by multiple cameras from various viewing angles simultaneously. This technique can be used to capture depth in addition to texture, which is known as Multi-view Video-plus Depth (MVD). MVD can provide information about the geometry of the captured scene and can be used like conventional videos for transmission and viewing.

The arrangement of cameras in multi-view video setups can vary based on the application and desired video quality, as well as the display devices being used. Figure 1 is an illustration of a typical multi-view video capture system.

Each camera-captured video can be encoded using any standard encoder such as H263, H264/AVC, etc. However, the individual sequences comprising the multi-view video must be encoded using the same encoder to ensure compatibility. This is essential for new multimedia applications such as 3D television (3DTV) and Free Viewpoint Video (FVV), which provide viewers with a more realistic impression of the visualized scene. This allows for the provision of a completely new stereoscopic

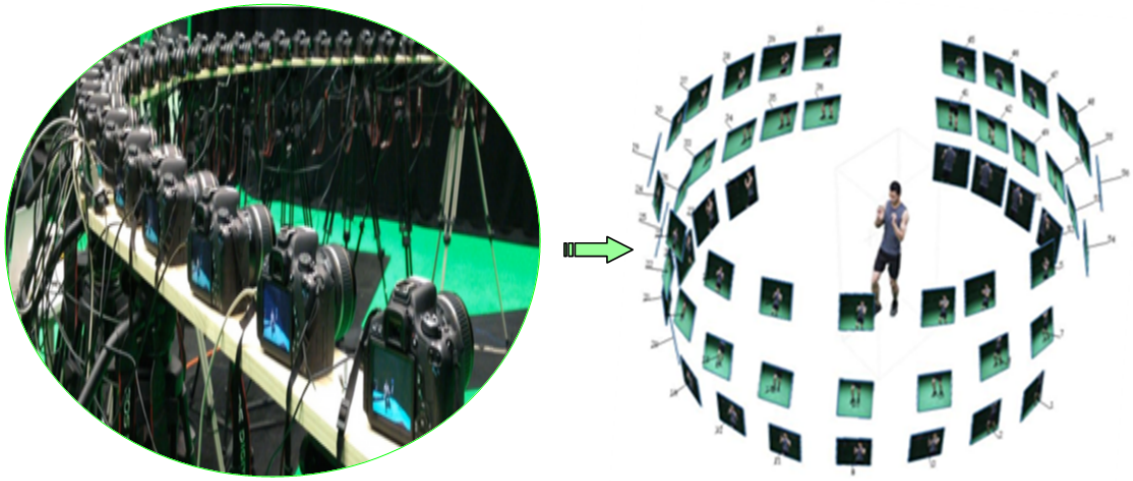

Figure 1: Multi-View Video Capture System.

vision to viewers.

Multi-view video has a wide range of applications across several domains, including entertainment, surveillance, medical imaging, sports, and robotics. With its ability to capture depth and texture, multi-view video has opened up new possibilities for creating immersive and realistic visual experiences.

Multi-view video compression is an area of research that has garnered significant attention in recent years. The compression of MVD sequences, in particular, is a challenging task due to the additional bits required to represent depth information accurately. To achieve efficient compression of multi-view video, it is necessary to leverage the temporal, spatial, and inter-view redundancies present in the different sequences.

Moreover, during the transmission of multi-view video sequences, errors can occur, which can lead to a degradation in the reconstructed video quality. Hence, developing compression techniques that can handle these transmission errors is critical for efficient and reliable multi-view video transmission.

The goal of this thesis is to explore and evaluate multi-view video compression techniques that can provide high compression efficiency and robustness to transmission errors. Specifically, the study aims to investigate methods that can exploit the correlations between multiple camera views to reduce redundancy. Additionally, the

thesis aims to evaluate error robustness in MV-HEVC coding standard.

The proposed research will have significant implications for applications that rely on multi-view video, such as 3D television, free viewpoint video, and virtual reality. By improving the compression efficiency and robustness of multi-view video, it will be possible to deliver high-quality video content over networks with limited bandwidth and to enhance the user experience.

Multi-view video technology has revolutionized the way we perceive and interact with visual media. Unlike traditional single-view video, multi-view video provides the user with the ability to view a scene from multiple angles, navigate between views, and enjoy special visual effects provided by divergent devices.

One of the most exciting applications of multi-view video is FVV, which allows the user to interactively control a viewpoint in a 3D space to view a dynamic scene. The specific feature of FVV is that users can select a viewpoint that does not correspond to any existing camera viewpoints, and generate a new view of the scene from any 3D position. This requires the use of multiple cameras arranged and configured carefully to achieve high-quality video. FVV has many potential applications, such as in sports coaching where a coach can better evaluate their players through multiple angles of view.

Another exciting application of multi-view video is 3DTV, which has become increasingly popular in recent years. 3D video requires capturing depth along with the texture of the scene to be visualized. This means that from each viewpoint, a sequence of texture and depth must be obtained. Several 3D display techniques, such as stereoscopic or auto-stereoscopic systems, can be implemented to provide the viewer with a sense of depth. In stereoscopic systems, special devices like glasses are used to provide the viewer with the sense of depth, while auto-stereoscopic systems use a lenticular network or parallax barrier to distribute different viewpoints intended for each eye.

The compression of multi-view video sequences is a challenging task, and compression of MVD sequences is particularly demanding due to the depth information requiring additional bits for accurate representation. Developing robust compression techniques that can cope with transmission errors is essential for efficient and

reliable multi-view video transmission. Research into multi-view video compression techniques that can provide both high compression efficiency and robustness to transmission errors is an important area of investigation.

Multi-view video compression is a vital aspect of handling the vast amount of information generated by multi-view video capture. Due to the high redundancy of information contained in the different views captured simultaneously, as well as the considerable size of the processed information, efficient compression is indispensable. The importance of compression in this case is evident in the improvement and optimization of storage and transmission of video over computer networks such as the internet. To achieve efficient compression, various constraints need to be considered, including compression efficiency defined by a bitrate/quality compromise, random temporal accessibility requested by all video coding algorithms, random accessibility to the different views of the video, and possible reduction of the complexity of the coding algorithm.

Multi-view video compression (MVC; Multi-view Video Coding) is the process of generating a single binary stream (bitstream) using N views and exploiting the different views after decoding the binary stream sent by the encoder. The general architecture of multi-view video compression is illustrated in Figure 2. The MVC has been standardized as an extension of the H.264/AVC standard by "The ISO/IEC JTC1/SC29/WG11 Moving Pictures Experts Group (MPEG)". All charact eristics and techniques of this video compression standard are used by MVC.

Multi-view video compression has numerous applications, including 3D video, virtual reality, and augmented reality. It enables a high-quality viewing experience with a reduced storage footprint and a lower bandwidth requirement. Additionally, the ability to access multiple views of the video provides enhanced user interactivity and improved overall video quality. However, multi-view video compression still faces challenges, including ensuring compatibility with different devices and video players, as well as ensuring the robustness of the compressed video stream against transmission errors. These challenges require ongoing research and development to ensure that multi-view video compression continues to meet the ever-increasing demands of video processing and delivery.

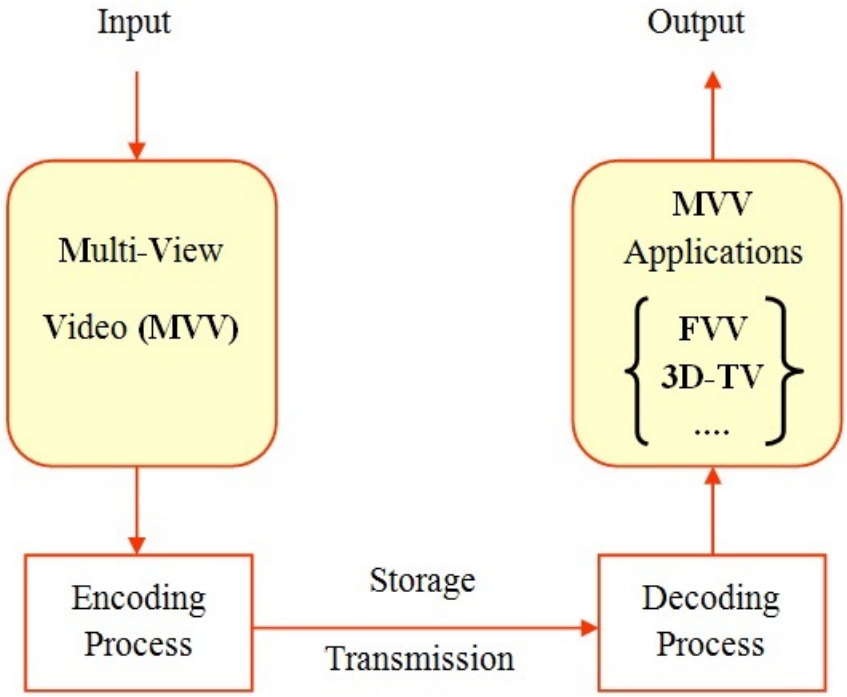

Figure 2: General Architecture of a MVV Compression System.

Thesis Contributions

This thesis mainly addresses issues related to multi-view video compression, with a particular emphasis on reducing computational complexity for the multi-view extension of HEVC (MV-HEVC) and robustness to transmission errors. Specifically, the main contributions of this thesis are as follows:

- Efficiently reducing the computation time for motion and disparity estimations (DE) during the encoding process can help reduce the coding complexity of Multi-View Video (MVV). This thesis proposes an efficient algorithm for Earlier Disparity Estimation (EDE) to further decrease the computational complexity in MV-HEVC. The main contribution of this work is the improved early offset estimation between the different views of the MVV, which takes into account several three-view prediction structures. The EDE algorithm reduces the complexity of disparity estimation by eliminating the inter-view offsets caused by different viewing positions. By considering inter-view simi-

larities and coding unit (CU) split levels, the DE search window is reduced, thereby decreasing the number of searching points within a limited window. Additionally, motion estimation (ME) is improved in this thesis for dependent views only, by introducing an earlier decision for CU splitting (EDCS).

- Due to the scarcity of studies analyzing error-sensitive encoded data in MV-HEVC, this thesis aims to explore the sensitivity of MV-HEVC bitstream under error-prone conditions as a second contribution. The evaluation work examines the effects of different encoded components in the MV-HEVC bitstream, such as the Network Abstraction Layer (NAL) unit types, which include both the Video Coding Layer (VCL) and Non-Video Coding Layer (Non-VCL) samples that are encoded and subjected to various random bit errors and packet loss rates. This study aims to identify the most sensitive compressed data that should be encoded in higher-protected channels, and the less error-sensitive encoded data that can be encoded in dependent views.

Thesis Outline

In addition to this introduction, which serves as both a motivation and a general presentation of the problem, the manuscript consists of four chapters organized as follows:

Chapter 1 introduces some general information about video compression, particularly the different types of video redundancies and hybrid video coding. We will then provide a brief overview of the technical aspects of the two standards H.264/AVC and HEVC to clarify the specificities of each norm. To position MV-HEVC in relation to HEVC, we will study the profiles and levels of HEVC. Finally, this chapter concludes with the presentation of current multi-view video coding standards, particularly the two standards MVC and MV-HEVC.

In Chapter 02, we will outline the fundamental concepts of multi-view video compression, including the constraints that must be satisfied, the conditions, and the MV-HEVC Hybrid Video Coding. Additionally, we will introduce the construction of the MV-HEVC bitstream and the complexity of this standard. Furthermore, we

will discuss the HTM reference model jointly developed by ISO/IEC JVT, MPEG, and VCEG.

In Chapter 03, we will discuss the proposed approach for reducing the computational complexity of the MV-HEVC standard while maintaining a better trade-off between bitrate and video quality. After presenting a statistical study on the use of blocks in inter-view prediction, we will detail the proposed method based on the elimination of inter-view offset. Finally, to complete the study, we will examine the results on the test data presented in Chapter 02.

The fourth chapter deals with the analysis of the sensitivity of the MV-HEVC bit stream under error-prone conditions. Then, a study will be presented on the effects of different components coded in the MV-HEVC bitstream, such as the types of Network Abstraction Layer (NAL) units, which are samples of Video Coding Layer (VCL) and Non-Video Coding Layer (non-VCL), coded and subjected to error-prone conditions with various error rates on bits and packet loss. The study presented in this chapter helps identify the compressed data that is most sensitive to coding in a more protected channel, as well as the coded data that is least sensitive to errors in dependent views.

Chapter 1

Video Compression Fundamentals

1.1 Introduction

Video compression is the process of reducing the size of a digital video file by removing redundant or irrelevant information, without significantly reducing the visual quality of the video. This is achieved by exploiting several types of redundancy that exist within video content, using specific video coding standards.

The aim of this chapter is to provide a comprehensive introduction to video compression fundamentals and standards that will serve as a foundation for further study and exploration in this area. The chapter will cover the concept of video redundancies, the different types of video compression, and the fundamental principles of hybrid video coding. It will also discuss various video compression standards that have been developed over the years, including H.264/AVC, HEVC, MVC and MV-HEVC.

1.2 Video Redundancies

Streaming high-definition video data requires approximately 2 billion bits per second of data bandwidth. Owing to the significant amount of data required to represent digital video, it is desirable for these video signals to be easily compressible and decompressible. Fortunately, exhibits strong vertical, horizontal, and temporal correlation or redundancy, making it highly compressible. Data compression refers to

the reduction in the number of bits required to store or transmit a video sequence by exploiting four types of redundancies those are Psycho-visual, temporal, spatial and statistical redundancies [1].

1.2.1 Psycho-visual Redundancy

Psycho-visual redundancies refer to the persistency of human eye. The human eye cannot perceive certain details in a digital picture, and those details can be selectively removed from the digital data. However, removing these details does not significantly affect the quality of the digital data, as they are not noticeable to the human eye.

1.2.2 Spatial Redundancy

Video data is typically composed of a sequence of still pictures or frames. To exploit the correlations between pixels, each frame can be divided into smaller pixel blocks. Particularly, when considering the luma and chroma components [2], neighboring pixels within a video frame tend to exhibit significant similarity . Spatial redundancy emerges from the horizontal and vertical correlations between adjacent pixel values within the same frame, commonly referred to as intra-picture correlation. Figure 1.1 illustrates an example of spatial redundancy.

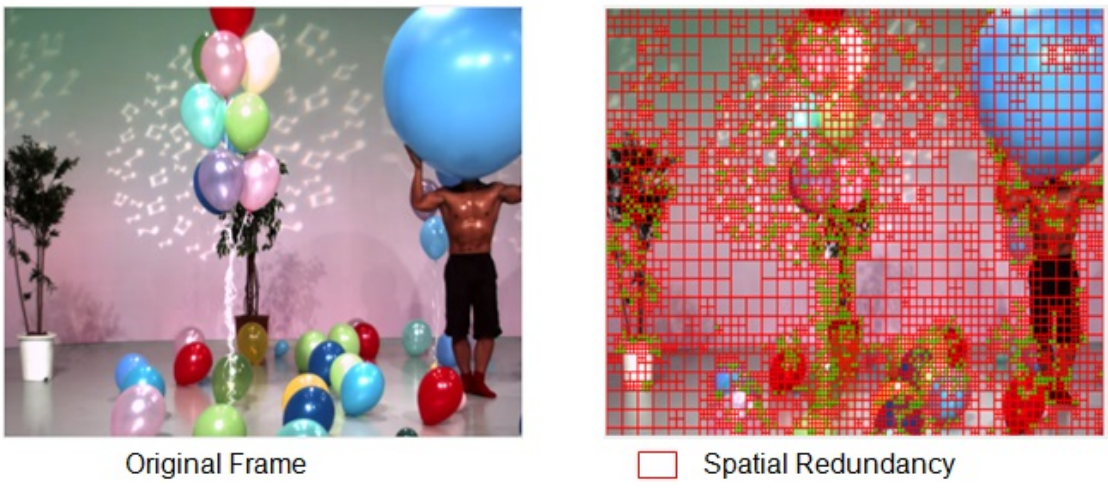

Figure 1.1: Spatial Redundancy Within the Same Video Frame.

1.2.3 Temporal Redundancy

The number of frames per second in a motion video is called as frame rate. Typically, a video is displayed at a frame rate higher than 15 frames per second (fps) to ensure smooth and continuous motion perception by a human observer [3]. This necessitates a high degree of similarity between neighboring frames.Temporal redundancy refers to the correlation or similarities between adjacent frames within the same video sequence, also known as inter-frame correlation.

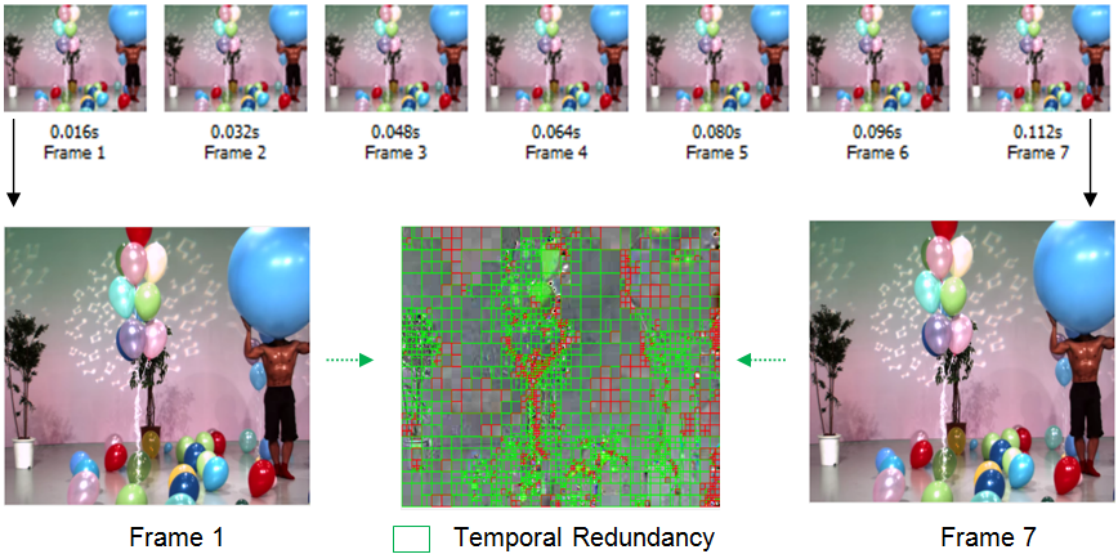

Figure 1.2: Temporal Redundancy Between Successive Video Frames.

1.2.4 Statistical Redundancies

Statistical redundancies in video data typically refer to additional bits or fixed-length binary codes added to encode various components such as transform coefficients, motion vectors, and other data. This redundancy can be eliminated at the last level of video compression before transmission.

1.3 Video Compression

Video compression techniques involve the removal of redundant information that is not considered critical for viewing the video content. In end-to-end video compres-

sion, the process involves an encoder situated at the source, which compresses the input video into a bitstream. At the receiver end, a decoder is utilized to decompress this bitstream, allowing for video playback [4]. The collaboration between an encoder and a decoder is often referred to as a codec or video compression standard, as depicted in Figure 1.3.

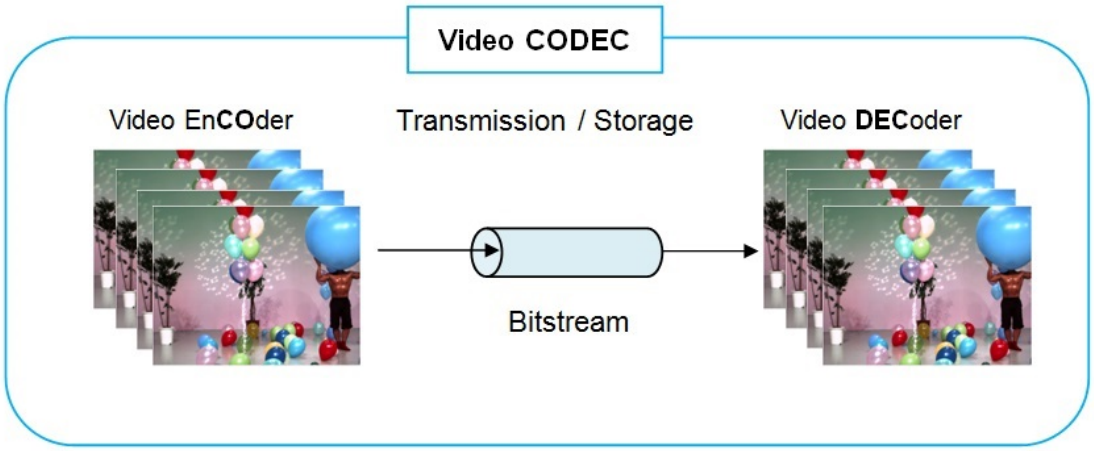

Figure 1.3: Video CODEC (COder/ DECoder).

1.3.1 Lossy Compression

Lossy compression techniques involve a partial loss of information, which means that the compressed data cannot be fully restored during the subsequent decompression process [5]. Lossy video compression takes advantage of psycho-visual redundancies in the human visual system, removing non-relevant data portions. This process retains only the information that can be perceived, resulting in a decoded image that appears almost identical to the original for most observers.

1.3.2 Lossless Compression

Lossless compression serves the objective of reducing redundancy in video data while maintaining the entirety of the information, by encoding it with a reduced number of bits [5]. In which, the compressed data can be fully restored during the decompression process, ensuring that the original and reconstructed video pictures are

identical.

Spatial, temporal and statistical redundancies are the primary types of redundancies visible to the human eye. Such redundancies can be exploited by lossless compression techniques such as efficient prediction, which is commonly found in most current video codecs.

1.4 Hybrid Video Coding (HVC)

Among the widely adopted video compression codecs, Hybrid Video Coding (HVC) stands out as a standardized communication protocol that guarantees interoperability between several types of encoders and decoders [6]. The encoder in the HVC system mainly goes through four steps: prediction, transformation, quantization, and entropy coding. This can be later played by a decoder with a reverse process.

1.4.1 HVC Encoding Process

Figure 1.4 illustrates the fundamental elements of a generic video encoder within the HVC system, represented as a block diagram.

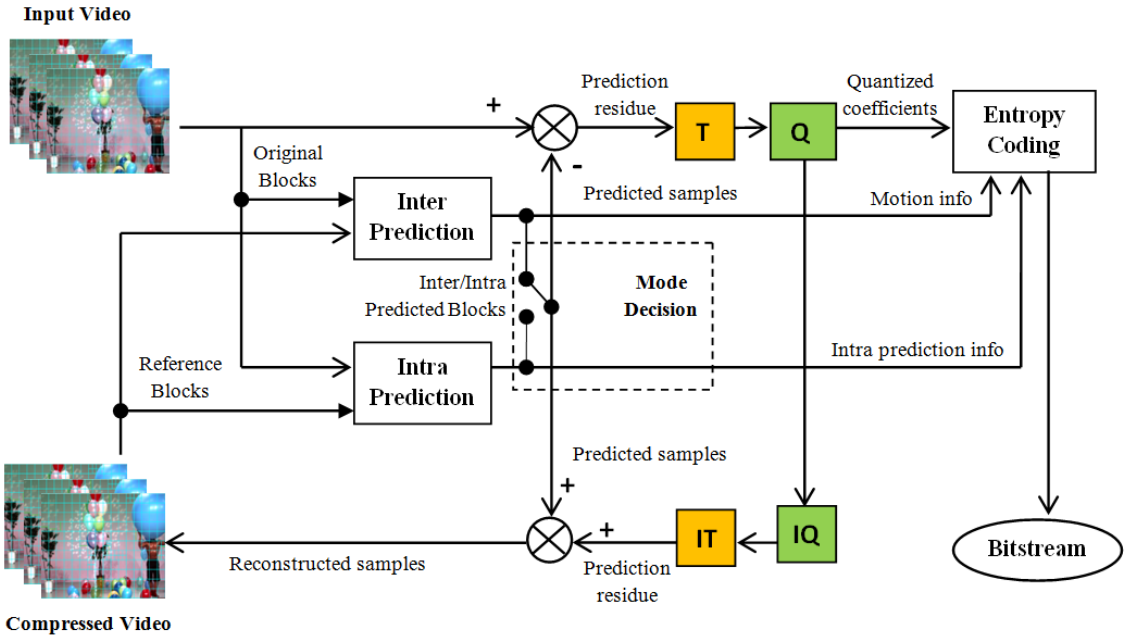

Figure 1.4: Generic HVC System (Encoder).

1.4.1.1 Prediction

Before the encoding process begins, all input video frames are partitioned into rectangular blocks. The performance of video coding largely depends on the effective use of block predictions. The prediction process aims to identify common blocks within the same frames or among neighboring frames .When a reference block is found, the redundant one is removed by either intra-prediction or inter-prediction. Intra-prediction reduces spatial redundancy by predicting blocks in the current frame from neighboring blocks already encoded in the same frame, using a predefined prediction mode. In inter-prediction, redundant information between adjacent frames or so-called temporal redundancy is eliminated by predicting the current block from the blocks belonging to the pre-coded frames. Enhancement of inter-prediction can be achieved through accurate estimation of block motion. This technique is commonly known as Motion Compensated Prediction (MCP) or prediction assisted by Motion Estimation (ME) [7].The ME algorithm is utilized to search for the most similar block within the reference frames, relative to the current block. Once the best matching block is identified, the ME algorithm determines the corresponding motion vector (MV), which consists of a horizontal and a vertical component. This motion vector indicates the relative displacement of the region in relation to the current block position. Given that the predicted and current blocks are rarely identical, the MCP generates a shifted pixel block from the referenced frame. This process involves utilizing the motion vector (MV) to calculate and encode the difference between the shifted block and the current block. This difference is commonly referred to as the prediction residue or prediction error [8]. In the case of inter-prediction during the MCP, the motion vector (MV) needs to be encoded and transmitted alongside the residual. Conversely, for intra-prediction, the residual is encoded and transmitted using different prediction modes [9] [10] [11].

1.4.1.2 Transformation

In the field of image and video coding, effectively encoding spatial data poses inherent challenges due to the high correlation among neighboring samples in the spatial domain and the random distribution of energy across video frames. Consequently,

a mathematical transform (referred to as the T module) is employed to process prediction error blocks. This transformation converts the values from the spatial domain to the frequency domain, thereby mitigating correlation and concentrating energy in a reduced number of coefficients. Among the transformation techniques commonly employed in recent years, the Discrete Cosine Transform (DCT) [12] has gained significant popularity. Typically, it is applied to small, equally sized blocks of image pixels such as 8x8 or 16x16. Additionally, variable block size transforms [13] and 3D transforms [14] have also emerged as widely used alternatives.

1.4.1.3 Quantization

The aim of quantization is to map a set of continuous or discrete values to a smaller set of discrete values, called quantization levels. This is done by selecting one representative value for several values, which is usually the value closest to the original value. The quantization process can eliminate small values of the transformed residuals associated with spectral components that are not perceptibly relevant. For instance, in image compression, high-frequency image components are quantized more aggressively than low-frequency components due to the human eye reduced sensitivity to high-frequency details. Aggressive quantization of high-frequency components helps to eliminate small values associated with these components, leading to more efficient compression. However, it is important to note that improper selection of quantization levels can introduce errors or distortions.

The quantization is commonly used in lossy compression algorithms, in which it is used to reduce the amount of data needed to represent a video signal while preserving its perceptual quality.

1.4.1.4 Entropy Coding

Entropy coding is a highly effective method of lossless data compression, wherein a distinctive prefix code is generated and assigned to every quantized coefficient found in the input. The purpose of entropy encoders is to compress the data by substituting fixed-length input samples with their corresponding variable-length prefix codewords. Subsequently, the processed data is converted into a bitstream and

transmitted through a channel or stored in a medium for decoding purposes.

1.4.2 HVC Decoding Process

The hybrid video decoder is represented by the generalized block diagram illustrated in Figure 1.5.

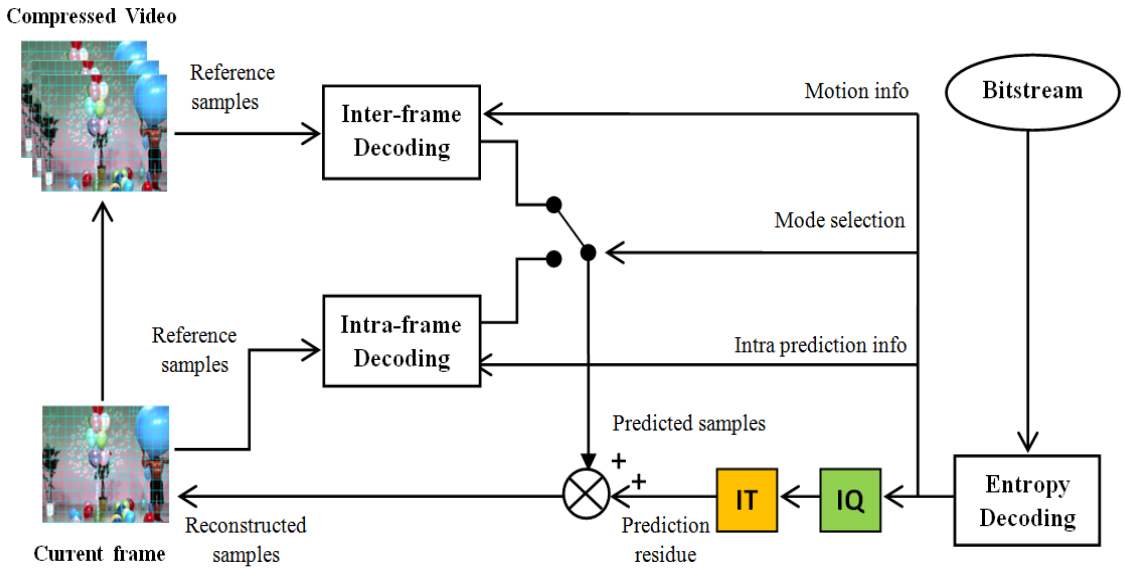

Figure 1.5: Generic HVC System (Decoder).

Within this diagram, the inverse quantization (IQ module) [15] and inverse transform (IT module) [16] are responsible for processing the quantized and transformed coefficients, respectively. Their outputs generate the prediction residue, which is then combined with the predicted samples derived from the inter-frame or intra-frame decoding modules. The resulting combination is stored as the reconstructed residue and serves as a reference for the intra-prediction or inter-prediction modules.

1.5 Video Compression Standards

Numerous video compression standards have been developed by international organizations over the past three decades and continue to remain highly popular today. The primary organizations responsible for the advancement of video coding standards are the International Telecommunication Union Telecommunication sector

(ITU-T), and the International Standard Organization /International Electrotechnical Commission (ISO/IEC) [17]. Collaboratively, ISO/IEC and ITU-T contribute to the creation of the MPEG and H.26x series of video coding standards. These standards are designed to enhance encoding bitrate efficiency while maintaining consistent perceived video quality [18].

Within the JCT-VC team (Joint Collaborative Team on Video Coding), dedicated developers and researchers tirelessly strive to maximize the compression ratio of video coding standards. Simultaneously, other researchers focus on improving the error robustness of video transmission in diverse network environments. Figure 1.6 visually presents the evolution of video coding standards starting from 1990 (H.261) and culminating in the latest generation, VVC (Versatile Video Coding) in 2020.

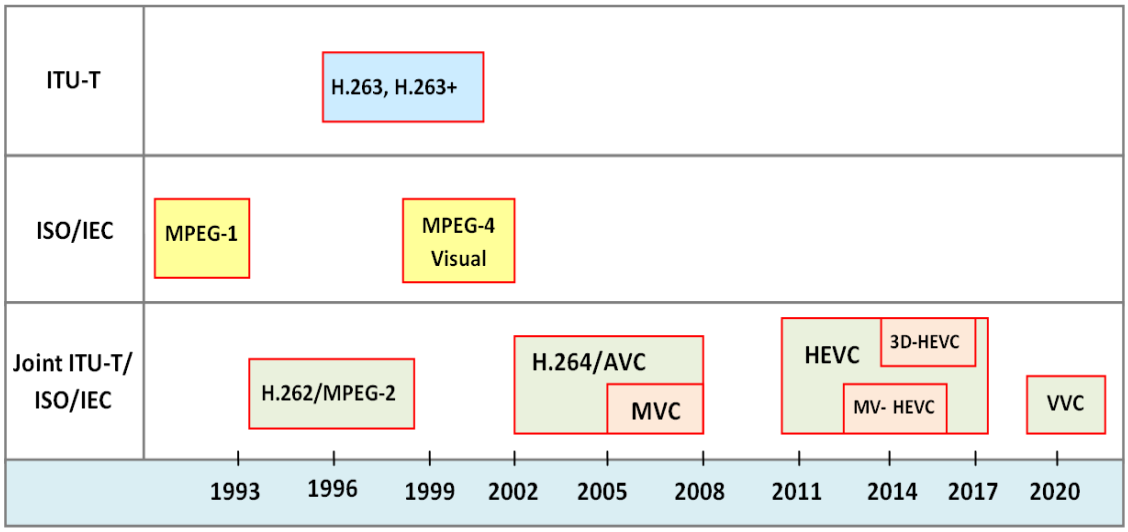

Figure 1.6: Advancements in Video Compression Standards.

Throughout the evolution of video coding standards, spanning from the inception of H.261 to the latest VVC codecs, each successive generation has consistently achieved a remarkable reduction of approximately (50%) in the required bitrate, while maintaining an equivalent level of video quality [19] [20] [21] [22]. However, H.264 and HEVC are the most commonly used standards in the world of video compression and streaming [23] as their HVC fully support layered video coding features and their extensions that are able to generate layered video streams with unequal importance used in 3D video coding.

1.5.1 H.264/AVC

The ongoing collaborative efforts between ITU-T VCEG and ISO/IEC MPEG have led to significant advancements in video coding, resulting in the establishment of the H264/AVC (Advanced Video Compression) standard in 2003 [24]. The AVC standard demonstrates notable enhancements in compression efficiency, surpassing previous standards by up to 50% across various bitrates and video resolutions. Notably, H.264 stands out from other standards due to its exceptional video robustness and the ability to maintain high-quality video at both high and low bitrates. The main improvements added to H.264 hybrid video coding are based on a layered design that allows different aspects of the compression process to be optimized independently.

1.5.1.1 H.264/AVC Layered Design

The AVC layered design refers to the organization of different functional components in a Network Abstraction Layer (NAL)[25]. Each layer provides a specific set of functionalities and is responsible for a specific process in the hybrid video coding [26].

There are two types of NAL units, Video Coding Layer (VCL) NAL units and non-VCL NAL units as shown in Figure 1.7.

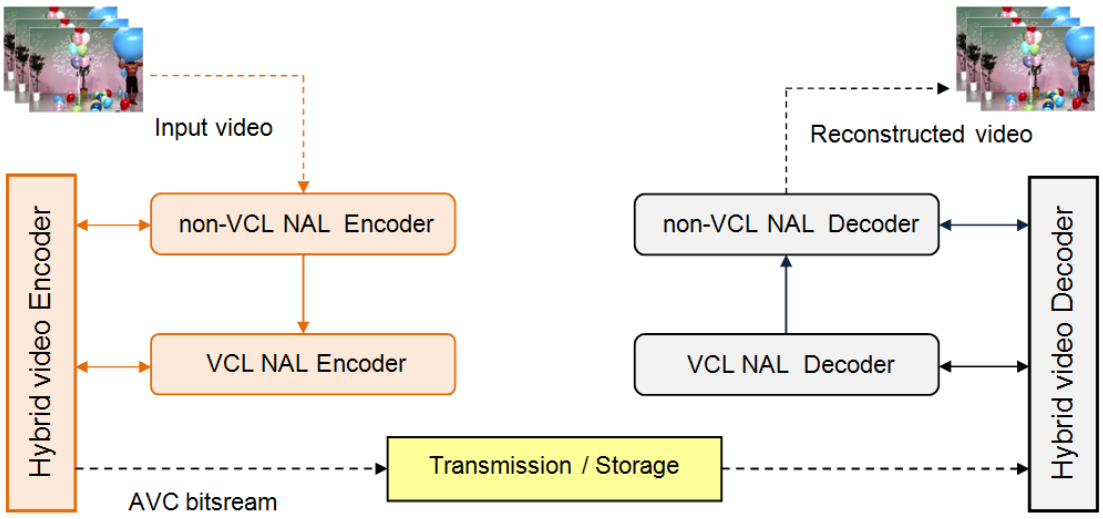

Figure 1.7: H.264/AVC Layered Design.

VCL NAL units are responsible for creating a coded representation of the video pictures based on variable block size to improve prediction processes, ME and MCP. Seven variable block sizes are allowed in VCL NAL which are 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4 [27] as shown in Figure 1.8.

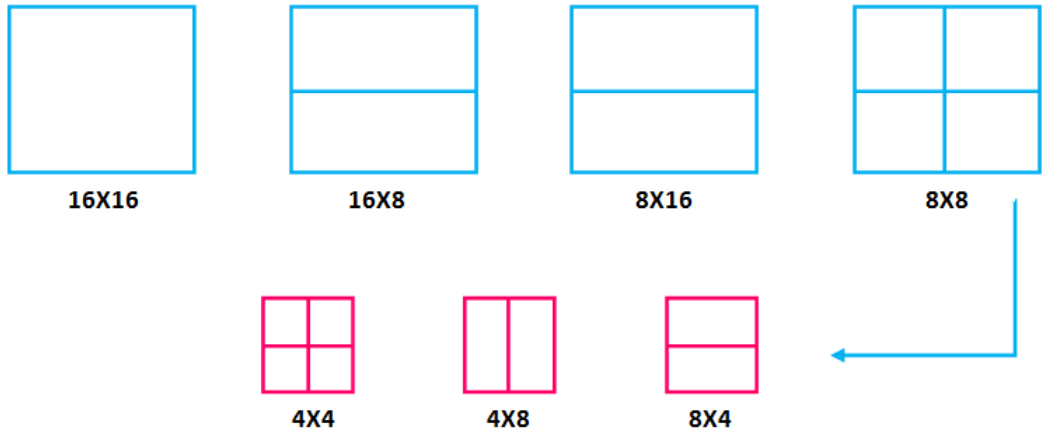

Figure 1.8: H.264/AVC VCL NAL Variable Block Sizes.

In contrast to previous standards that employed fixed blocks [28], the adoption of variable block sizes in AVC has resulted in reduced bitstream sizes. This approach enables smaller block sizes for encoding blocks with higher motion detail, thereby enhancing prediction by capturing fine motion details. Conversely, larger block sizes are utilized to encode blocks with lesser motion detail [29].

The non-VCL NAL units carry information that is needed for decoding and transmission the video stream such as Supplemental Enhancement Information (SEI), parameter sets, picture delimiters and filler data.

SEI is a set of optional metadata that can be included in the H.264 bitstream. It provides additional information about the video content or encoding process, and can be used to signal changes in the video stream or provide metadata about the video, such as color space information or time-code data. SEI can be included in any slice of the video frame and can be inserted at any point in the bitstream, allowing for greater flexibility in the encoding process. Also, SEI are important for improving the quality and efficiency of video encoding and playback, and should be carefully considered when designing an H.264 video encoding system.

Parameters set is a collection of parameters that define the configuration of the video encoder and decoder. There are two types of parameter sets in H.264: the Sequence Parameter Set (SPS) and the Picture Parameter Set (PPS) [30]. The SPS defines the basic properties of the video sequence, such as the image size, aspect ratio, and frame rate. The PPS provides additional information about individual video frames, such as the type of picture (e.g., I-frame, P-frame, or B-frame) and the motion vectors used to encode the frame. This made AVC more flexible for multimedia services, wireless and mobile networks [31].

1.5.1.2 H.264/AVC In-Loop Deblocking Filter

In addition to AVC layered design, H.264 employs an adaptive in-loop deblocking filter in the AVC decoder as shown in Figure 1.9. The deblocking filter is an op-

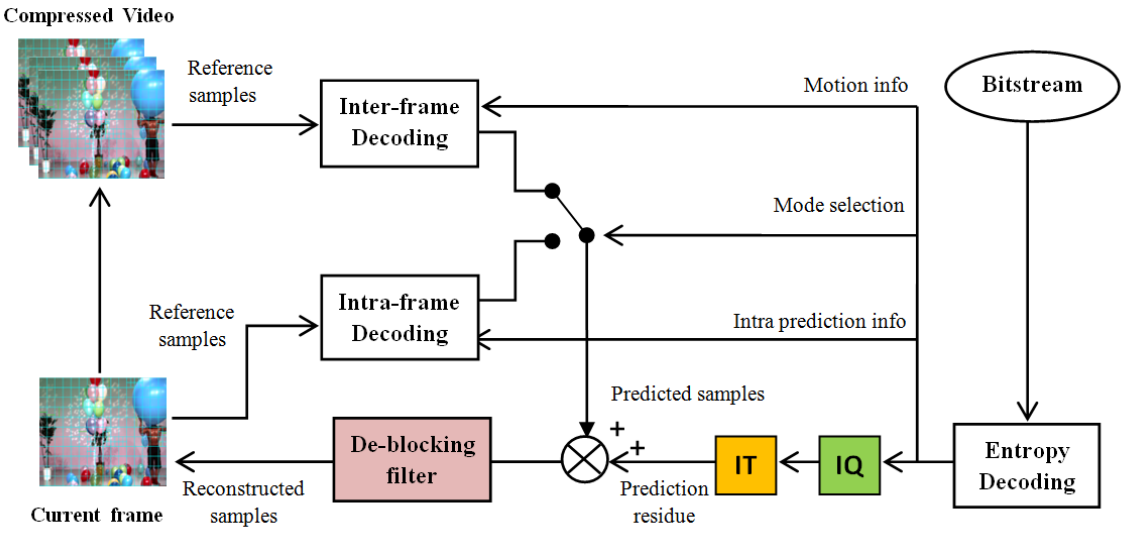

Figure 1.9: In-loop Deblocking filter of AVC.

tional feature in H.264 [32] [33]. It is used to reduce block artifacts that can occur when a video is compressed using block-based motion estimation, compensation and quantization techniques. Block artifacts are caused by the discontinuities between adjacent blocks in a compressed video frame, which can lead to jagged edges or visible blocks in the frame. The deblocking filter smooths out these discontinuities by applying a low-pass filter to the edges between blocks through various parameters such as the strength of the filter, the window size and the threshold for applying the

filter to different block edges [34]. These parameters can be adjusted to balance the reduction of artifacts with the preservation of detail in the compressed video.

1.5.1.3 H.264/AVC Drawbacks

The flexibility and robustness introduced in the newly coding structure of H.264 compression standard make it a popular choice for a wide range of applications, including video telephony, video conferencing, HDTV, storage (DVD and Blu-Ray player), video streaming, digital video authoring, digital cinema, and many others. It is included into approximately one billion devices worldwide. However, the in-loop deblocking filter can increase the computational complexity of the encoding process, leading to longer encoding times and higher processing requirements. There is also a risk of over-smoothing, which can result in a loss of detail and sharpness in the final compressed video. Furthermore, the expanding range of services, the surging popularity of HD video, and the emergence of higher-resolution formats such as 4k×2k or 8k×4k have created a pressing need for coding efficiency surpassing the capabilities of H.264/AVC [35].

1.5.2 H.265/HEVC

High Efficiency Video Coding (HEVC) is the latest standardization efforts of ISO/IEC MPEG and ITU-T VCEG for further improving the coding efficiency of H.264/AVC standard. The first version of HEVC was finalized and published in January 2013 by the (JCT-VC) [36]. It has been reported that HEVC can provide comparable subjective visual quality with H.264/AVC at only half bitrates in many cases. One of the key improvements in HEVC over H.264 is the use of a more complex layered coding structure. It combines various advanced coding tools to further improve compression efficiency and robustness, such as larger block sizes, quadtree-based coding unit partitioning, parallel processing, more advanced prediction methods, new profiles and levels.

1.5.2.1 HEVC quadtree-based coding unit partitioning

Newly emerged video resolutions require larger smooth regions in the picture, which can be encoded more efficiently when using large block sizes. This paved the way for HEVC to introduce an innovative structural syntax with larger block sizes, and more flexible than that defined in earlier H.264/AVC [37].

In HEVC structure, each frame is divided into CUs (Coding Units) for the purpose of encoding processes, and each CU can be further split into PUs (Prediction Units) used for predictions, or into TUs (Transform Units) defined for transform and quantization. CU is based on a quad-tree partitioning architecture, which can be recursively subdivided into four levels, corresponding successively to the depths (D0, D1, D2 and D3) ranging from 64x64 to 8x8, as shown in Figure 1.10 (a).

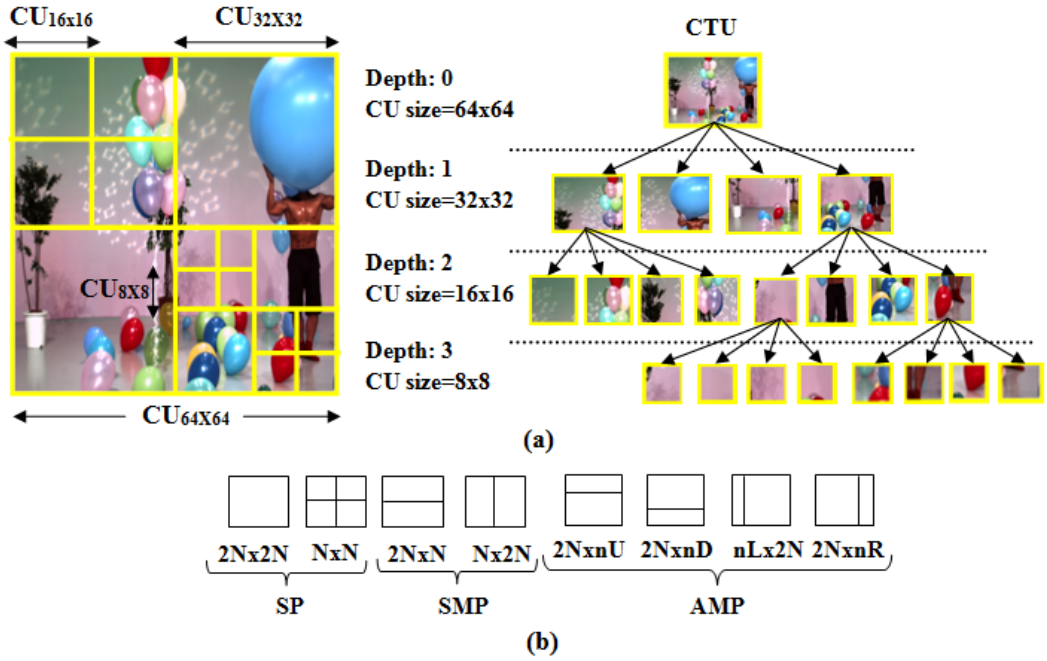

Figure 1.10: (a) CUs Splitting. (b) PUs partitions modes.

CUs of the same depth level have equal sizes, close and do not overlap. In order to enhance intra and inter predictions, each CTU (Coding Tree Unit) of size 2Nx2N is divided into PUs, which are predicted from previously coded data using Symmetric Partitions (SMP) and Asymmetric Partitions (AMP), as shown in Figure 1.10 (b). Square Partitions (SP) with sizes 2Nx2N and NxN are the symmetrical PUs used

in both intra prediction and inter prediction. Whereas, the rectangular SMP and AMP are only used for inter-picture prediction such as SKIP, AMVP (Advanced Motion Vector Prediction) and MERGE modes [38] [39] [40].

The TUs are used to apply Discrete Cosine Transform (DCT) or Discrete Sine Transform (DST) [41] to the pixel values of a CU, in which the residual of prediction is transformed, quantized and entropy encoded. TUs can be rectangular or square, and their size is determined by the CU depth. The smallest size of the TU is 4x4 pixels, while the largest size is 32x32 pixels.

The use of CUs, TUs, and PUs in HEVC partitioning concept provided a much larger number of options in the search for block matches to reduce the spatiotemporal redundancies, which immensely improved the efficiency of HEVC encoding and provided better compression performance, especially for UHD videos.

1.5.2.2 HEVC Parallel Processing

HEVC is designed to offer a significant improvement in coding efficiency compared to the H.264/AVC standard, with a target of doubling the efficiency. However, this higher efficiency comes at a cost of increased computational complexity. As a result, HEVC encoders are expected to be several times more complex than their H.264/AVC counterparts, making parallelism essential for achieving real-time HEVC [42]. Among the various options available, a multi-core platform is considered most suitable for accelerating video compression. HEVC incorporates slices, tiles, and Wavefront Parallel Processing (WPP) to further refine the partitioning structure of CTUs and enable efficient parallel processing. A slice in HEVC is a group of CTUs that can be decoded independently, as illustrated in Figure 1.11 (a).

On the other hand, a tile is a rectangular area of CTUs arranged in horizontal and vertical borders, as shown in Figure 1.11 (b). The number and size of tiles are signaled in the video bitstream using syntax elements. The tile boundaries are determined based on the CU and PU size, and they can be aligned to the raster scan order or specified using tile coordinates [43]. The utilization of slices and tiles for picture partitioning in video coding serves a primary purpose of facilitating parallel processing applications by enabling synchronized multi-threaded processing.

Figure 1.11: Picture partitioning using a) Slices, b) Tiles.

Moreover, tile partitioning in HEVC allows for convenient random access to specific local regions within video pictures through the HEVC bitstream. However, it is important to note that the tile coding structure lacks error robustness when it comes to errors within the HEVC coding standard. In the tile partitioning process, the scanning order initiates from the top-left corner of each divided tile.

In HEVC, a wavefront is another set of consecutive CTUs that are arranged horizontally, as shown in Figure 1.12.

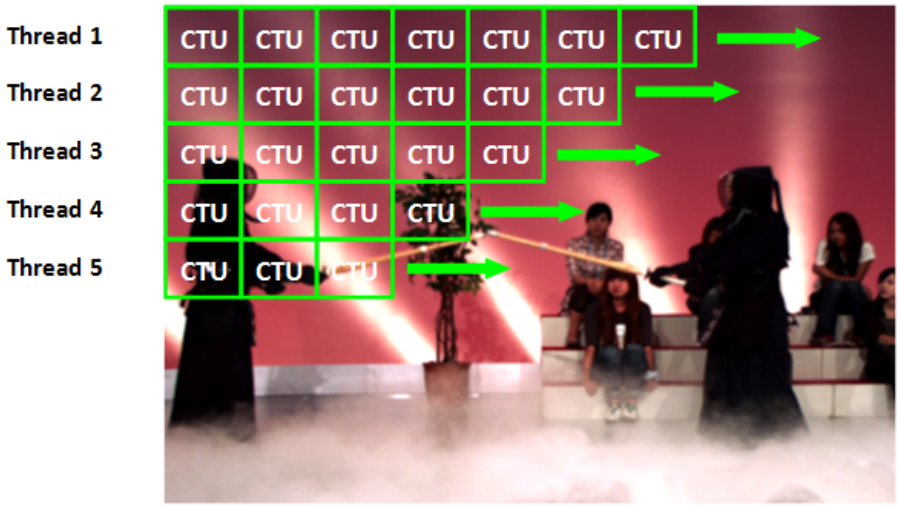

Figure 1.12: Partitioning of CTUs in Wavefront Parallel Processing (WPP).

The encoder processes each wavefront sequentially, but multiple wavefronts can be processed in parallel by different processors. By dividing the frame into wavefronts, WPP can effectively utilize multiple processors [44], leading to faster encoding times.

WPP in HEVC is signaled using a syntax element in the bitstream. The number of wavefronts and the size of each wavefront are determined based on CTU size and the picture dimensions. The number of wavefronts can be adjusted dynamically to optimize the processing performance based on the available computational resources.

One of the main advantages of WPP is that it improves encoding performance without compromising coding efficiency. In fact, WPP can improve coding efficiency by reducing the number of bits required for motion information and reducing the complexity of the motion compensation process at the same time reducing visual artefacts that can be produced using tiles

1.5.2.3 HEVC Layered Design

In HEVC layered design [45], a new layer called Video Parameter Set (VPS) is added to non-VCL NAL, as shown in Figure 1.13. The addition of VPS in HEVC provides more flexibility and efficiency in video encoding compared to H.264, particularly when it comes to handling changes in resolution and frame rate, multiple bitstreams, and video metadata.

More precisely, in H.264 the resolution and frame rate had to be defined in the SPS, which meant that every change in resolution or frame rate required a new SPS to be transmitted. With VPS, changes in resolution or frame rate can be handled more efficiently, as they can be specified separately in the VPS without requiring the transmission of a new SPS.

VPS provides a more efficient way of handling video metadata, such as color space information, bit depth, and chroma subsampling. In H.264, this information had to be specified in the SPS, which could result in inefficient use of bandwidth. With VPS, this information can be specified separately, allowing for more efficient use of bandwidth.

In addition, HEVC allows for the transmission of multiple independent bit-

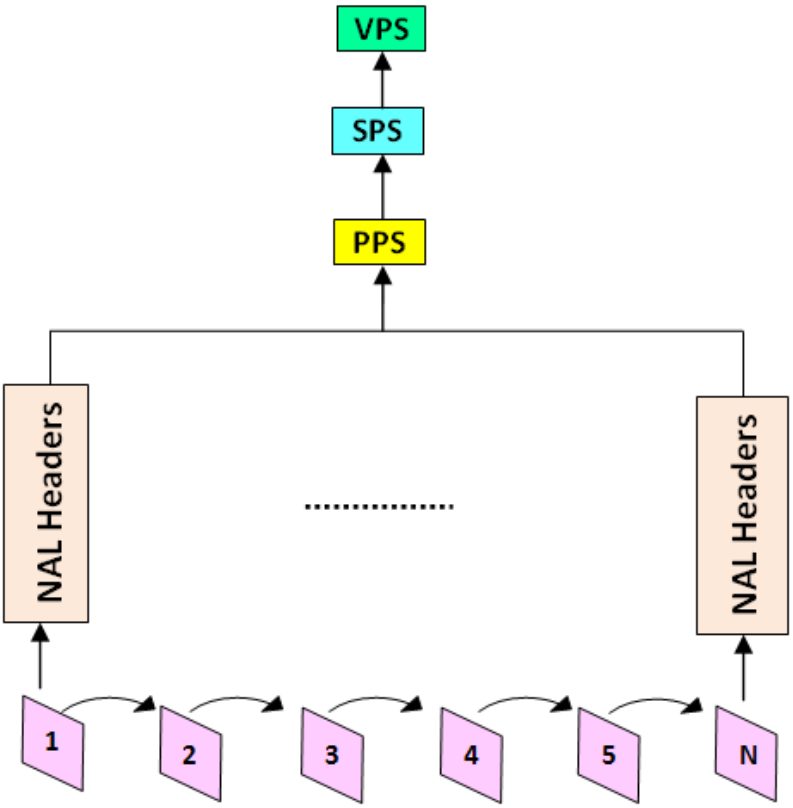

Figure 1.13: HEVC parameters sets.

streams, each with its own VPS, SPS, and PPS. This is particularly useful for applications that require the transmission of multiple video streams, such as multi-view or 3D applications. H.264 also supports multiple bitstreams, but they are not handled as efficiently as in HEVC.

1.5.2.4 HEVC Profiles and Levels

HEVC profiles and levels are standardized sets of specifications for video coding that determine the compression efficiency and compatibility of HEVC encoded video bistreams across different devices and platforms [46].

HEVC profiles define the features and capabilities of the codec, such as the maximum video resolution, bit depth, and chroma sampling formats that can be supported. There are several HEVC profiles, including:

- **Main profile:** is the most widely used HEVC profile, designed for a wide

range of applications, including video conferencing, online video streaming, and digital broadcasting. The Main profile supports video resolutions up to 4096x2304 pixels (4K UHD), and offers a good balance between compression efficiency and computational complexity.

- **Main 10 profile:** is designed for high-quality video applications that require support for 10-bit color depth, such as HDR (High Dynamic Range) video content. The Main 10 profile supports video resolutions up to 4096x2304 pixels (4K UHD) and 10-bit color depth, which allows for a wider range of colors and greater color accuracy.

- **Main Still Picture profile:** is optimized for still image coding, such as in digital cameras or medical imaging applications. It uses advanced coding techniques such as intra prediction and Context-Adaptive Binary Arithmetic Coding (CABAC) [47] to achieve high compression efficiency for still images.

- **Scalable Main profile:** this profile is designed for video coding applications that require scalability, such as adaptive video streaming. It supports multiple layers of video streams, each with a different resolution and bitrate, which can be decoded independently or in combination to provide different levels of video quality.

- **Main Intra profile:** is optimized for video coding applications that require high-quality, low-latency video compression, such as in video conferencing or real-time streaming. It uses only intra-frame coding techniques, which do not require motion estimation and allow for faster encoding and decoding.

Each HEVC profile is designed to cater to specific use cases and applications, and choosing the appropriate profile can help achieve optimal compression efficiency and video quality for a given application.

On the other hand, HEVC levels specify the maximum video parameters that a device can support in terms of video resolution, frame rate, and bitrate. Higher levels require more processing power and bandwidth to decode the video stream, and not all devices may be able to support the highest levels. The HEVC standard

defines ten levels, ranging from Level 1 (for low-resolution, low-bitrate video) to Level 6 (for ultra-high-definition, high-bitrate video) as shown in Table 1.1.

Table 1.1: HEVC Levels

HEVC Level	Maximum video resolution (pixels)	Frame rate (fps)	Maximum bitrate
Level 1	3840x2160 (4K UHD)	30	128 kbps
Level 2	3840x2160 (4K UHD)	60	1 Mbps
Level 2.1	4096x2160	60	1.5 Mbps
Level 3	5120x2700	60	3 Mbps
Level 3.1	5120x2700	120	4.5 Mbps
Level 4	8192x4320	30	20 Mbps
Level 4.1	8192x4320	60	50 Mbps
Level 5	8192x4320	120	120 Mbps
Level 5.1	8192x4320	240	240 Mbps
Level 6	16384x8640	120	480 Mbps

Choosing the appropriate HEVC level is important for ensuring compatibility between the video encoder and decoder, as well as for achieving optimal video quality and compression efficiency for a given application. Devices that support higher HEVC levels typically require more processing power and bandwidth to decode the video stream, and not all devices may be able to support the highest levels.

1.6 Multi-layer Video Coding

Video compression technologies have made possible various exciting new services since their international standardization started. While the main target has been high compression efficiency, a lot of attention has recently been paid to three-dimensional (3D) applications.

The ability to capture a same scene from multiple camera views simultaneously, often referred to as multi-view video (MVV) has opened up new possibilities for

emerging 3D video technologies, allowing for the creation of more realistic and engaging visual content.

When capturing MVV, each camera view requires its own video stream, and when combined, the resulting video data can be quite large. Although video streams from each camera can be encoded independently using the straightforward coding, such as AVC or HEVC, the exploitation of inter-view dependencies created between cameras views is believed to offer up to 50% or more in bit savings, thereby called Multi-view Video Coding (MVC) [48]. Figure 1.14 illustrates a typical structure of MVC.

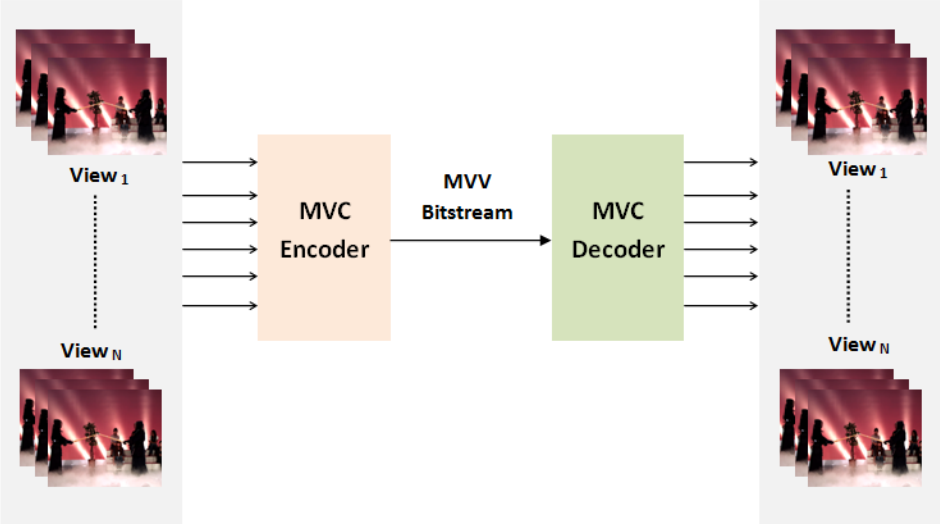

Figure 1.14: MVC Typical Structure.

The MVC encoder receives temporally synchronized video streams and generates one bitstream from each of the N camera views. The MVC decoder receives the bitstream, decodes and outputs the video signals for each of the N views.

Two main approaches that have been reported in the field of MVC. One class of approaches uses adaptive suband and decomposition within and across video sequences from different cameras. This technique is out of the scope of this particular thesis research.

Another class of algorithms extends the use of predictive coding, which is commonly used in single-view video compression, to multiple views. This approach takes advantage of the temporal and inter-view dependencies created between adjacent

views in MVV to reduce redundancy and compress the MVV data. MVC (Multi-view Video Coding) and MV-HEVC (Multi-View High Efficiency Video Coding) are both the latest MVC standards that have explored this approach.

1.7 Current Progress on MVC standards

1.7.1 MVC

Recognizing the growing need for MVC, MPEG published a call for proposals on the MVC standard in July 2005. Since early 2006, The Joint Video Team (JVT) has been devoting on the development of Multi-view Video Coding (MVC) standard as an extension of H.264/AVC, which has been finalized in 2008 [49].

The new MVC standard aims to exploit the high correlation between the multiple views of the same MVV, by exploiting spatio-temporal redundancies between the adjacent frames, as well as by exploiting inter-view redundancies between neighboring views, as shown in Figure 1.15.

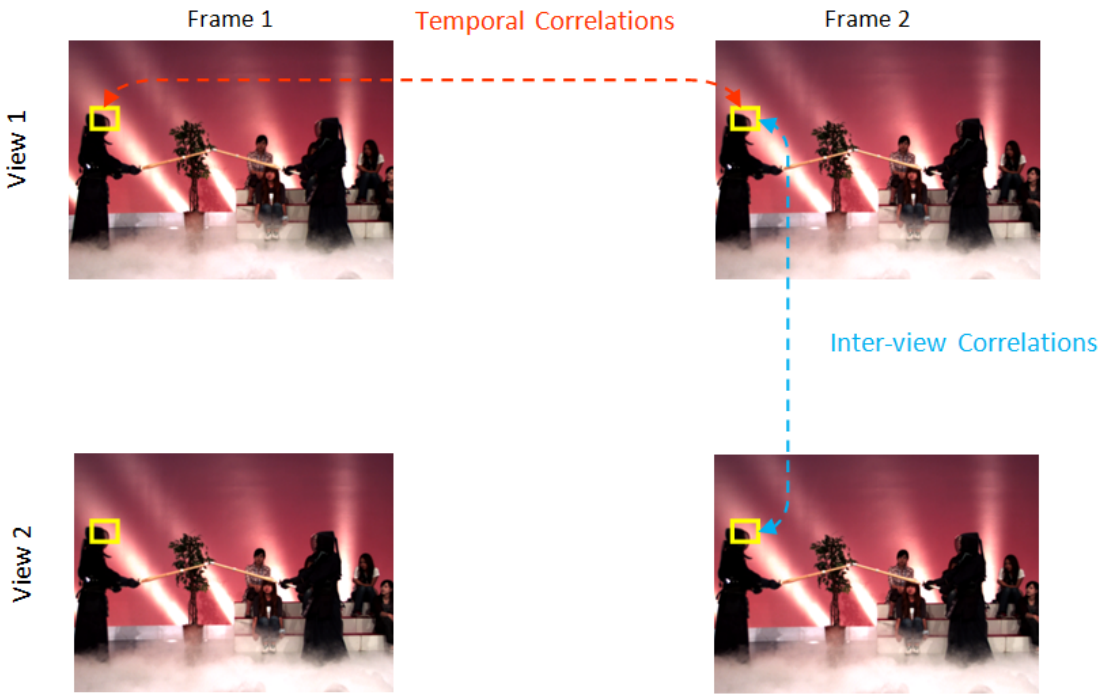

Figure 1.15: Inter-view redundancies between two views of Kendo sequence.

The JVT is also considering the integration of additional features into the codec with minimum changes in macro-block level syntax. Such features include illumination compensation that incorporates illumination changes into the motion compensation process, adaptive reference filtering that compensates for focus mismatches between views, and view synthesis prediction that generate synthesized views from neighboring views.

MVC increases the MVVs compression rate by approximately 30% compared to independently coded multi-view videos using H.264/AVC. These improvements make MVC suitable for use with multi-view 3D displays and the coding format in the majority of 3D Blu-rays. On the other hand, the recent advent of 3D Ultra High Definitions (3D-UHD) technologies and the introduction of HEVC have slowed the success of MVC solutions, as they are intended for compressing high and lower definitions videos.

1.7.2 MV-HEVC

In an effort to satisfy the MVC requirements and to significantly enhance the feasibility of 3D Ultra High Definition (3D UHD) contents, the Joint Collaborative Team on 3D Video (JCT-3V) developed MV-HEVC, a multi-view extension of HEVC. MV-HEVC was published in mid-2014.

MV-HEVC offers a highly efficient video coding solution for MVV contents by combining the capabilities of both HEVC and MVC designs. It offers backward compatibility with HEVC, which enables existing HEVC decoders to decode MV-HEVC bitstreams for the base view. This ensures that MV-HEVC can be adopted gradually, without requiring a complete overhaul of existing MVC inter-layer structure.

In addition to its backward compatibility with HEVC, MV-HEVC allows a View Synthesis Prediction (VSP) in its inter-layer design.The basic idea behind VSP is to use the available views to synthesize a new virtual view. This can be used in cases where inter-view prediction based on motion vectors is not effective, such as in regions with occlusion or where there is significant camera motion. By using view synthesis to generate a synthesized view, VSP can effectively capture the complex

relationships between views By generating a virtual view and using it for prediction and therefore MV-HEVC can achieve higher compression efficiency and better video quality than traditional inter-view prediction techniques.

MV-HEVC improved features make it an ideal choice for applications that require efficient and high-quality video streaming over limited bandwidth, adaptive streaming, and immersive 3D video experiences. Furthermore, MV-HEVC serves as the foundation for 3D-HEVC standard. In which, it is built on the top of the MV-HEVC codec and extends the MV-HEVC view synthesis to support the encoding of depth maps. Depth maps provide essential depth information of a 3D scene and can be compressed more efficiently than full-color video frames. By encoding depth maps, 3D-HEVC achieves more efficient compression of 3D video content compared to traditional 2D video coding standards, resulting in lower bitrates and reduced storage requirements.

Moreover, 3D-HEVC utilizes multi-view video coding to enhance the visual quality of 3D video content, which has been widely adopted for various applications, including virtual reality, augmented reality, and 360-degree video. Its efficiency in compressing 3D video content has made it a popular choice for delivering high-quality immersive media experiences. With ongoing research in this area, 3D-HEVC continues to evolve, providing improved compression efficiency, reduced computational complexity, and higher visual quality for 3D video content.

1.8 Conclusion

Video compression techniques play a crucial role in enabling efficient video streaming over limited bandwidth and improving the overall user experience. Therefore, a thorough understanding of video content redundancies and video coding standards is essential for developing state-of-the-art video compression solutions.

This chapter began by highlighting the video compression process and emphasizing the impact of hybrid video coding on perceived visual quality at both the encoder and decoder sides. A concise review of video coding standards was then provided, with a focus on the primary objectives of each coding tool developed for both cur-

rent and previous video coding standards. Then the chapter presented the current progress of the multi-view video coding standards, particularly the two standards MVC and MV-HEVC.

When comparing the background of video coding standards presented in previous sections, it becomes clear that MV-HEVC offers advantages over traditional single-layer video and MVC coding techniques in terms of the efficiency of video compression and the quality of the encoded video. However, it requires more computational resources for encoding and decoding, and may also require additional bandwidth for transmission. It is therefore typically used in applications where the benefits of multi-view video encoding outweigh these costs, such as in stereoscopic 3D, virtual reality, and immersive video applications. The next chapter of this thesis will present more details about the MV-HEVC standard and its encoding complexity.

Chapter 2

Multi-View High Efficiency Video Coding (MV-HEVC) Standard

2.1 Introduction

The multiview extension of high efficiency video coding (MV-HEVC) offers a significant improvement in compression performance for multi-view and 3D videos when compared to multiple independent single view HEVC coding. However, the high computational complexity of MV-HEVC demands significant optimization of the MV-HEVC codec and poses a significant challenge in terms of real-time processing and error robustness.

This pivotal chapter of the thesis focuses on the various tools and techniques that have been proposed to enhance the performance of MV-HEVC codec. The initial sections furnish an extensive overview of the requirements of MV-HEVC and the diverse optimization tools that have been put forth for this codec.

Moving on, the following sections explore optimization strategies proposed for MV-HEVC codec, such as parallel processing, prediction and mode decision algorithms. The performance of these techniques will be evaluated in terms of encoding time, complexity, and visual quality to identify the most effective approach for optimizing the MV-HEVC encoder.

In addition to exploring optimization techniques for the MV-HEVC codec, this chapter also covers methods proposed in the literature to enhance robustness against

multi-view video coding and communication errors. Specifically, we will investigate error resiliences and error concealments. At the end of the chapter, we will provide an explanation of the Test Common Conditions (CTC) and the MV-HEVC test model.

2.2 General Requirements

Multi-view video is becoming increasingly popular due to its potential for providing an immersive viewing experience. The development of MV-HEVC was driven by several requirements to support efficient coding of multi-view video content. Some of the main requirements for developing MV-HEVC include:

- **Compression efficiency:** one of the main requirements for MV-HEVC was to provide high compression efficiency for multi-view video content. This involves developing new coding tools and techniques to effectively exploit the redundancies between the different views and reduce the overall bitrate required to represent the content.

- **Scalability**: another requirement for MV-HEVC was to support scalability, which allows for the video content to be encoded at different bitrates and resolutions to accommodate different network conditions and playback devices.

- **Low complexity:** MV-HEVC was designed to have low computational complexity to enable real-time encoding and decoding of multi-view video content. This was achieved through the use of efficient coding tools and techniques, as well as the development of parallel processing algorithms to take advantage of multi-core processors.

- **Flexibility:** MV-HEVC was designed to be flexible to support different types of multi-view video content, including different numbers of views, view geometries, and camera setups.

- **Interoperability:** MV-HEVC was developed with interoperability in mind, to ensure compatibility with existing video coding systems such as H.264/AVC

and H.265/HEVC, as well as support for different video formats and transport protocols.

2.3 MV-HEVC Hybrid Video Coding

Fundamental principle of MV-HEVC hybrid video coding is to reuse the same coding tools of HEVC with only inter-layer processing and High-Level Syntax (HLS) changes.

For inter-layer processing, the MV-HEVC encoder should include a combination of motion estimation and disparity estimation (DE) techniques. Figure 2.1 illustrates the processing blocks typically included in the HEVC encoder to build an MV-HEVC encoder.

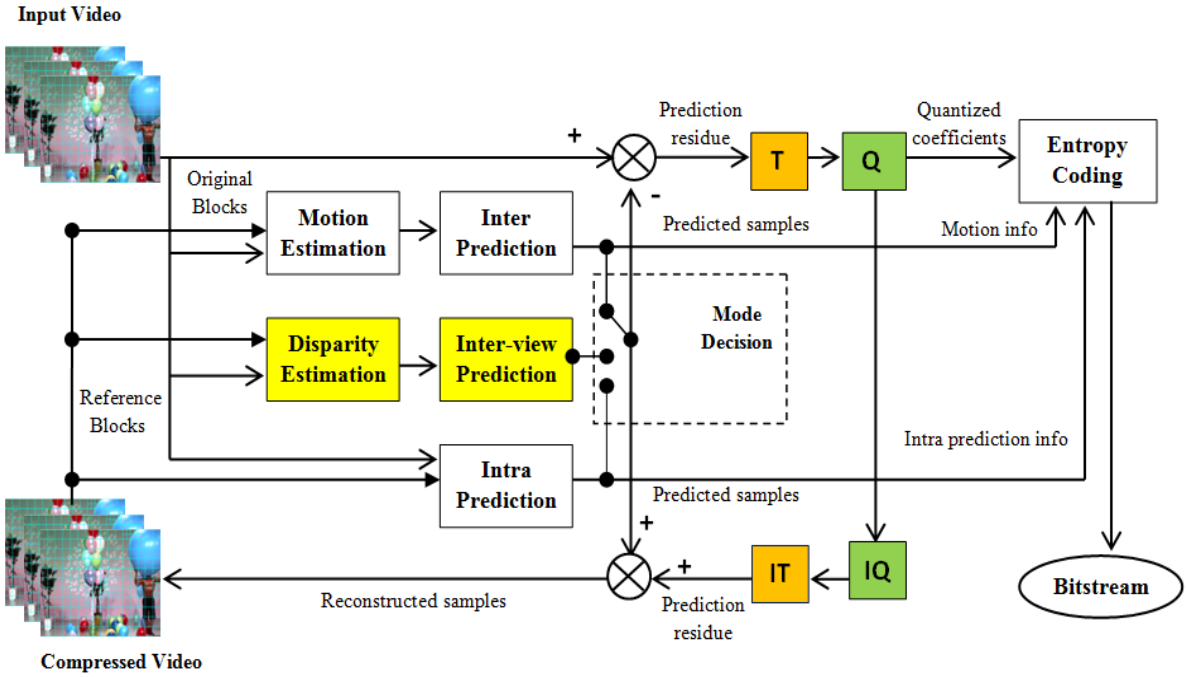

Figure 2.1: Diagram of the MV-HEVC Encoder.

2.3.1 Disparity Estimation

In multi-view video coding, different cameras capture the same scene, which means that the content of the videos is almost the same. The slight difference is due to

the change in angle of capturing between the scene and the cameras. This can be observed from the CU splitting of the first frame for three views shown in Figure 2.2.

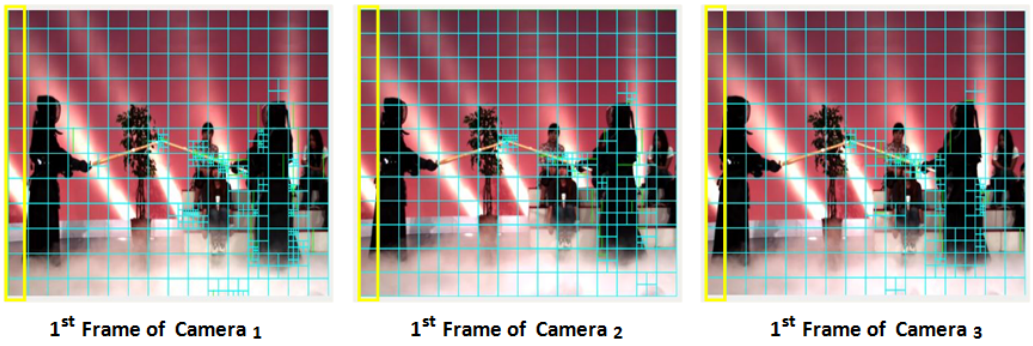

Figure 2.2: Three views from video sequence "Kendo", showing all CU splitting of each First Frame and the inter-view offsets.

Disparity estimation involves estimating the spatial offsets between corresponding points in two different camera views, as shown in Figure 2.3.

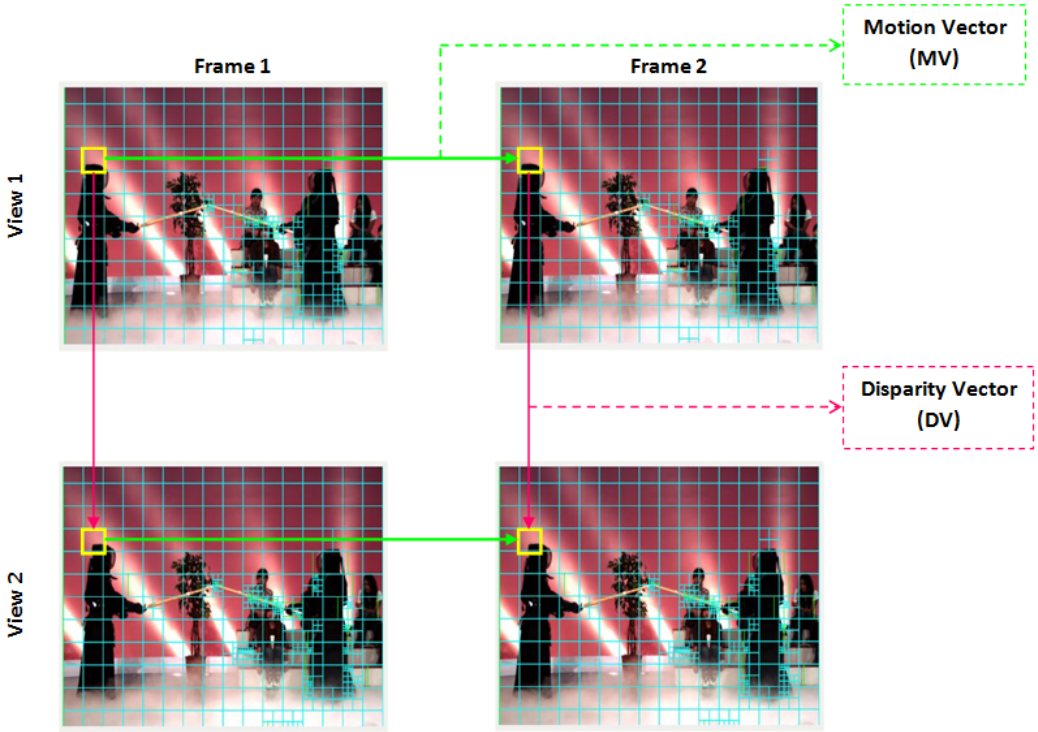

Figure 2.3: Motion and Disparity Vectors.

Also, it can estimate the disparity vectors for each block in the current view by

comparing it with the corresponding blocks in the reference views. The disparity vectors can be represented using a set of syntax elements in the bitstream. Once the disparity vectors are estimated, the reference frames can be transformed to match the current view. This can be done using techniques such as block-based motion compensation. The resulting compensated reference frames are subsequently employed for inter-view prediction [50].

2.3.2 Inter-view Prediction

Unlike a single-view sequence, a multi-view sequence exhibits high correlations between views in addition to spatio-temporal correlations within each frame. Enhancements in subsequent standards and recently in MV-HEVC [51] have reduced the data rates of MVV sequences and achieved a high coding gain. MV-HEVC used an efficient coding structure to exploit inter-view and spatio-temporal dependencies. This is predicted by a hierarchical coding concept known as a Matrix Of Pictures (MOP), as shown in Figure 2.4. Each column of the MOP represents the time do-

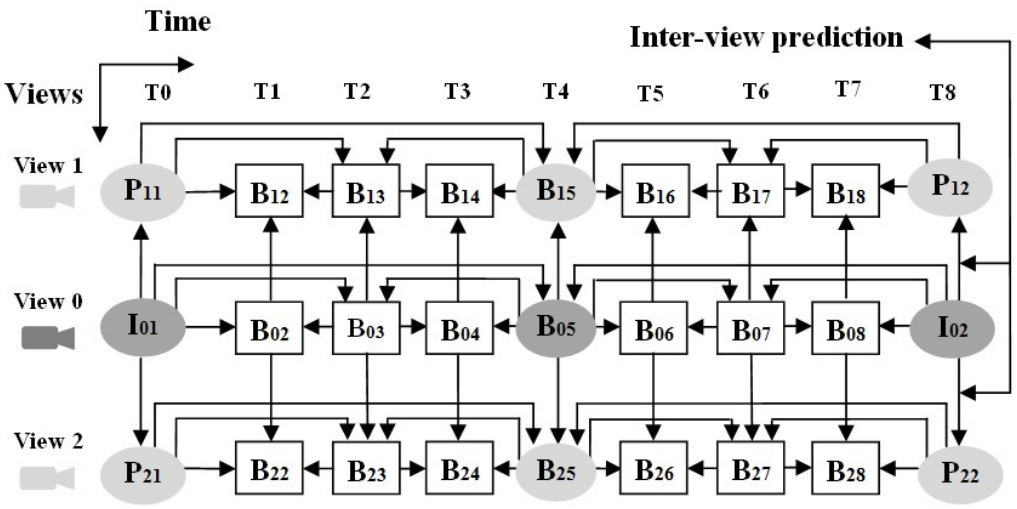

Figure 2.4: MV-HEVC Prediction structure for three-views.

main of the video while each row consists of Group Of Pictures (GOP). A GOP is a fixed pattern which is periodically repeated along the video sequence with the encoding of pictures in Intra mode (I), Predicted mode (P) and Bi-predicted mode (B).

2.3.2.1 Intra-coded Frames (I Frames)

I frames are a type of frame that contains the full image information and can be decoded independently without any dependency on other frames. In other words, each I-frame is a self-contained image that can be decoded without any reference to any other frame. I-frames are used as the starting point for a group of pictures (GOP) in video compression.

The use of I-frames at the beginning of each GOP facilitates random seeking, which means that users can quickly jump to a specific part of the video without having to decode every frame that comes before it. This is because I-frames provide a reference point for decoding other frames in the GOP. In contrast, without I-frames, seeking to a particular part of the video would require decoding every frame up to the desired point, which would be time-consuming and computationally intensive.

Another benefit of I-frames is that they provide significant coding robustness to transmission errors. This is because each I-frame is coded with only moderate compression, reducing the spatial redundancies in the image. As a result, if there is a transmission error that affects a particular frame, only that frame needs to be retransmitted, and the rest of the video can continue to be decoded using the I-frame as a reference point. In contrast, without I-frames, a transmission error in one frame would affect the decoding of all subsequent frames, leading to a loss of image quality or even complete video failure[52].

2.3.2.2 Predictive-coded Frames (P Frames)

P frames are a type of frame that is more efficiently compressed than I-frames by using motion compensated prediction from a past I or P-frame. This is achieved by analyzing the motion of objects within the video frames or the video views and predicting their location in subsequent frames.

By only transmitting the difference between the predicted and actual image data, P-frames can achieve higher compression than I-frames, which require transmitting the full image data [53]. P-frames are typically used as a reference for further prediction in video compression. For example, a P-frame might predict the motion of a moving object in one frame and use that prediction to generate the next frame. The

next frame might then be used as a reference for further prediction, with subsequent frames predicting the motion of objects relative to the previous frame. This process can continue until the next I-frame is reached, at which point a new reference point is established.

By using motion compensated prediction and referencing past frames, P-frames can achieve high compression while maintaining good image quality. However, this approach also introduces some challenges, such as the need to handle errors and discrepancies in motion prediction. In addition, because P-frames depend on past frames, they may not be suitable for certain types of video content, such as live streaming or real-time video, where delays in transmission can be problematic.

2.3.2.3 Bidirectional-predictive-coded frames (B Frames)

In MVV compression, B-frames are used to provide the highest degree of compression. B-frames are compressed by analyzing the motion of objects within the video frame and predicting their location in both the past and future frames. By transmitting the difference between the predicted and actual image data, B-frames can achieve even higher compression than P-frames.

However, in multiview video compression, B-frames also require both past and future reference pictures for motion compensation. This means that the encoder must use information from both the previous and next frames in multiple viewpoints to generate the B-frame in a specific viewpoint. B-frames are typically placed in between two P-frames in the GOP structure to ensure that both past and future reference pictures are available.

The use of B-frames in multi-view video compression can introduce some challenges. One of the challenges is that the use of both past and future reference pictures can introduce delays in video decoding, which can be more significant in multi-view video compression due to the increased complexity of the decoding process. Another challenge is that the use of future reference pictures can make B-frames more vulnerable to transmission errors, as the future reference picture may not be available at the time of decoding in all viewpoints [54].

Despite these challenges, B-frames remain an essential part of multiview video

compression, particularly in applications where high compression is essential, such as virtual reality or 3D video. The use of B-frames can significantly reduce the amount of data that needs to be transmitted, enabling efficient use of limited bandwidth and reducing storage requirements while maintaining high-quality video.

2.3.2.4 Access Unit (AU)

The pictures captured at the same time are grouped together and stored in an Access Unit (AU), with a uniform Picture Order Count (POC). The first view or base layer within an AU must comply with the single-layer coding standards of HEVC. The layers of other views within the same AU are known as enhancement or non-base layers and must conform to the specifications for MVC extension. In the case of the prediction structure for a three-view MV-HEVC, as shown in Figure 2.5, View 0 is the base view, which is encoded first without using inter-view prediction information. On the other hand, View 1 and View 2 are dependent views and use inter-view prediction from the base view to improve coding efficiency.

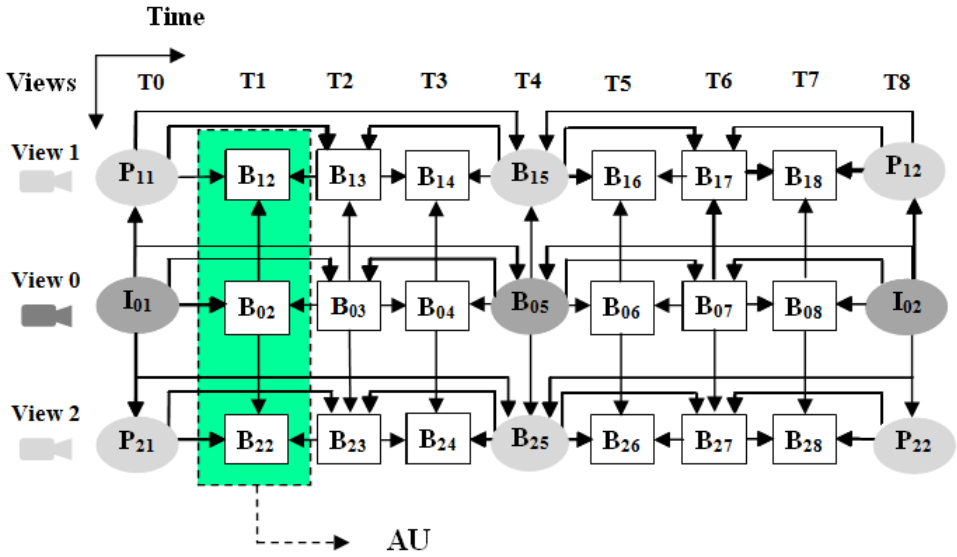

Figure 2.5: Access Unit in Three-View MV-HEVC Prediction Structure.

In HEVC single-layer coding, motion information MV and reference index for a current prediction unit (PU) can be coded in merge mode or using advanced motion vector prediction (AMVP). In both modes a list of candidates is created

from the motion information of spatial or temporal neighboring PUs. In this process MVs from neighboring blocks may be temporally scaled, by multiplying the POC difference between the picture of the current PU and its reference picture; and dividing the POC difference between the picture of the neighboring PU and its reference picture.

Within the context of MV-HEVC, pictures within the same AU are assigned identical POC values. Consequently, the POC of an inter-view reference picture matches that of the current picture, potentially leading to MV scaling issues arising from division or multiplication by zero. To circumvent this problem, inter-view reference pictures in MV-HEVC are consistently designated as long-term reference pictures. This designation allows for the utilization of a block-level design mechanism from HEVC version 1, known as the motion hook [55]. The motion hook effectively disables MV scaling for long-term reference pictures. Furthermore, since DVs correspond to spatial distance while MVs correspond to temporal distance, performing prediction between DVs and MVs proves inefficient. Hence, in HEVC single-layer coding, an MV from a neighboring PU can only be considered as a candidate for the current PU if the related reference pictures for both the current PU and the neighboring PU are marked as either long-term reference pictures or short-term reference pictures.

By employing the motion hook, MV-HEVC achieves efficient reuse of motion prediction from HEVC single-layer coding, requiring modifications solely in the HEVC bitstream design. More details about this design will be explained in the following section.

2.3.3 MV-HEVC Bitstream Construction

MV-HEVC compressed bitstream must include the base view bitstream encoded independently of other views in a format compatible with decoders for the single view profile of HEVC. Figure 2.6. shows the proposed scheme explaining the construction of the MV-HEVC bitstream.

The encoded multiple bitstreams are assembled into one bitstream and the synchronization between bitstreams as well as the connection between the base and

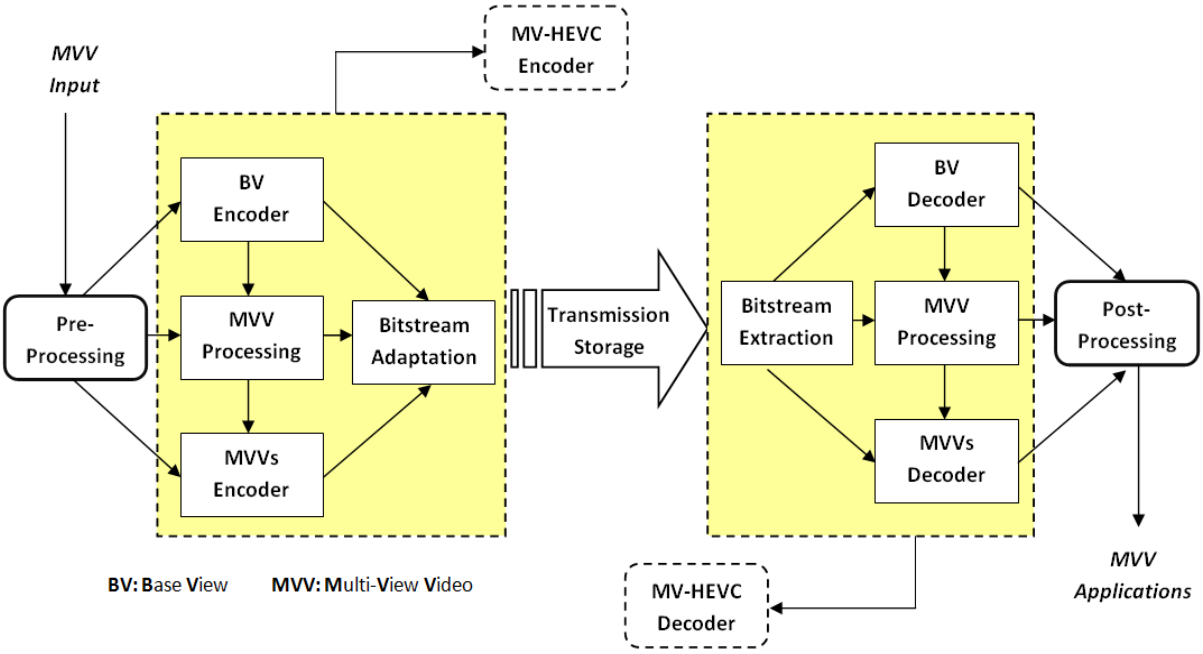

Figure 2.6: MV-HEVC Bitsream construction.

dependents views decoding processes are provided by the layer system.

In which, an encoded MV-HEVC video bitstream can be obtained by encapsulating the video data in a series of NAL units. MV-HEVC shares the same NAL unit header as that specified in HEVC standard. Each one contains an integer number of bytes. The first two-byte constitutes the NAL unit header that is designed to make it easy to parse the main properties of a NAL unit such as type layer and temporal sub-layer. The rest of the NAL unit contains the payload data. Figure 2.7 provides a visual representation of the NAL unit header used in the MV-HEVC. The

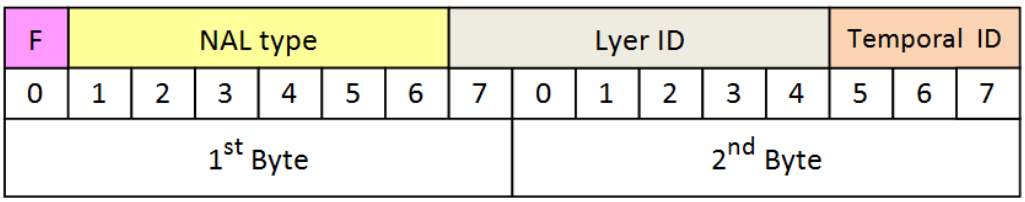

Figure 2.7: MV-HEVC NAL unit header structure.

NAL units that represent compressed video pictures, are the Video Coding Layer (VCL) NAL units. The slice data within the VCL NAL units is encoded with larger

Coding Tree Units (CTU) according to a particular scanning order. Each CTU can split into square regions of equal sizes, known as Coding Units (CUs) that contain samples of the luma and the chroma color planes. The CUs have a size selected by the encoder ranging from 8×8 to 64×64 pixels for all HEVC extensions. There are three basic slice coding classes that specify the types of coding supported for the CUs:

1. Random access point pictures, where a decoder may start decoding a coded video sequence. These are referred to as Intra Random Access Pictures (IRAP). There are three types of IRAP pictures; Instantaneous Decoder Refresh (IDR), Clean Random Access (CRA), and Broken Link Access (BLA). IDR pictures are self-contained pictures that can be decoded independently of any other picture in the sequence. They are used to start a decoding process and are typically placed as the first picture in a GOP. CRA pictures, on the other hand, allow a decoder to start decoding at any point within a GOP without any error propagation from previously decoded pictures. These pictures are useful in scenarios where the video stream may be damaged or interrupted, as they allow the decoder to recover quickly. Finally, BLA pictures are used to recover from a broken link in the video stream. These pictures can help recover the video stream after transmission errors or packet loss.

   The decoding process for a coded video sequence always starts at an IRAP.

2. Leading pictures, which precede a random access point picture in output order but are coded after it in the coded video sequence. In other words, they are pictures that come after an IRAP picture in the coded video stream but appear before the IRAP picture when the video is played back. Leading pictures that are independent of pictures preceding the IRAP in coding order are called RADL (Random Access Decodable Leading) pictures. These pictures can be decoded independently of the pictures that precede the IRAP in coding order, which allows for faster decoding and random access to the video stream. Random Access Skipped Leading pictures (RASL) is another type of leading picture that are not decodable on their own and require other pictures to be

decoded before they can be used. RASL pictures are typically used for random access points that are not IDR frames, such as recovery points, which can be used to recover from errors in the video stream. RASL pictures can be used as reference pictures for decoding other frames in the stream.

3. Trailing pictures, which follow the IRAP and the leading pictures in both, output and display order. Pictures at which the temporal resolution of the coded video sequence may be switched by the decoder: Temporal Sublayer Access (TSA) and Stepwise Temporal Sublayer Access (STSA). TSA allows the decoder to access specific sub-layers of the video sequence based on the temporal resolution, which can improve the overall decoding performance and reduce the computational complexity. This is achieved by coding specific information in the trailing pictures that allows the decoder to identify and access the desired sub-layers. On the other hand, STSA provides a more fine-grained temporal resolution switching capability by allowing the decoder to switch the temporal resolution at specific intervals in the video sequence. This is achieved by coding additional information in the trailing pictures that allows the decoder to identify the specific switching intervals.

2.3.4 MV-HEVC High Level Syntax (HLS)

The structuring of the MV-HEVC bitstream into NAL units and all of the syntax outside the VCL NAL or the slice data is considered as the High-Level Syntax (HLS) or the non-VCL NAL. The interconnection between MV-HEVC VCL and Non-VCL NAL units is highlighted in the proposed scheme depicted in Figure 2.8 An example of the HLS outside the slice data is the slice headers that precede the slice data within the VCL NAL units and all of the syntax in the Non-VCL NAL units, such as Parameter sets and Supplemental Enhancement Information (SEI) messages. There is a hierarchy of parameter sets in MV-HEVC, including the Video Parameter Set (VPS), the Sequence Parameter Set (SPS) and the Picture Parameter Set (PPS).

The VPS is the newly parameter introduced in HEVC and it has been extended

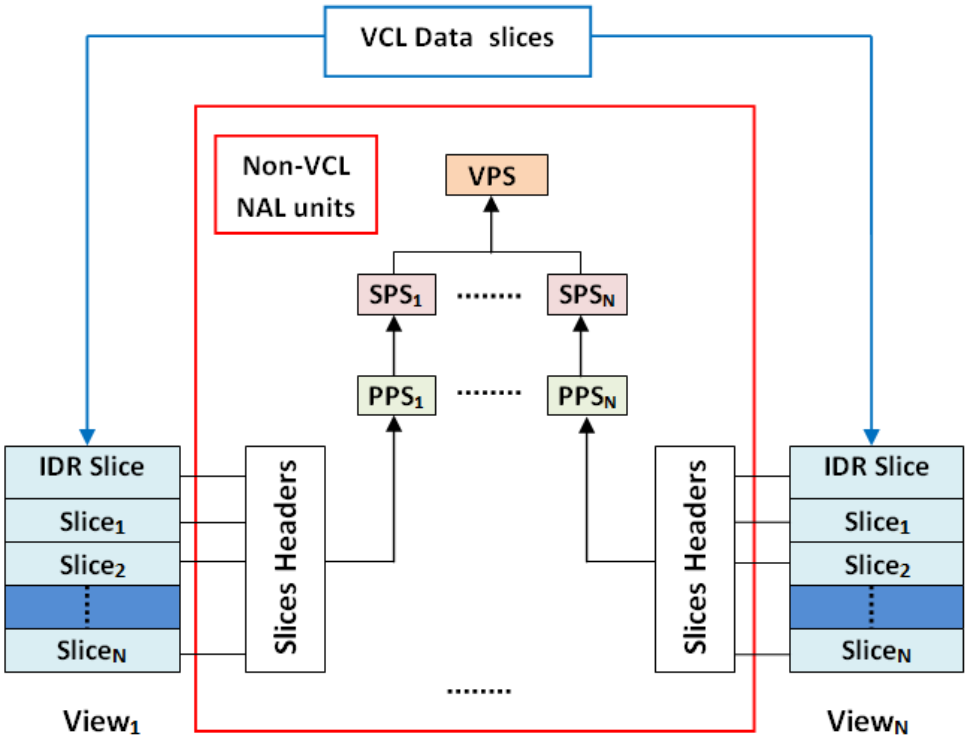

Figure 2.8: MV-HEVC NAL Units types (VCL and Non-VCL).

in MV-HEVC by adding the number and format of views, views dependency, layer sets and output layer sets. A VPS contains two parts, the base VPS and the VPS extension. The base VPS, as per the first version of HEVC, includes information relevant to the HEVC version 1 compatible layer, as well as operation points that correspond to layer sets. Moreover, the base VPS contains temporal scalability details, such as the maximum number of temporal layers.

On the other hand, the VPS extension holds information related to additional views beyond the base view. The syntax used in the VPS extension can flexibly associate each layer ID with scalability parameters and inter-view dependencies. When coding the current view, view dependencies are signaled to indicate which views are used as reference views for inter-view prediction. The pictures are encoded in increasing order of view ID, meaning that a view can only depend on another view with a lower value of layer ID. VPS is usually transmitted once per sequence.

The SPS contains parameters that do not change from picture to picture within

coded views. This include picture format (size and the color sampling format), PTL indicators (Profiles, Tiers, and Level), block structures (coding, prediction, transform), etc. However, the PPS contains parameters that may change for different pictures within the same view such as initial quantization parameter (QP), chroma QP offsets, slice configurations, deblocking filter controls, and scaling list data. Each slice header information is stored in a single active PPS, SPS. Finally, the VPS contains shared information of the PPS and the SPS. SPS and PPS are typically transmitted once per view.

By transmitting these parameters separately, MV-HEVC enables more efficient coding of MVV sequences, where different views may have different coding parameters. This helps to reduce the bitrate while maintaining the desired video quality.

2.4 MV-HEVC Complexity

The incorporation of new coding tools and parameter sets in MV-HEVC can increase the computational complexity of the MV-HEVC encoder, potentially affecting the encoding time and requiring more processing resources. The multiple parameter sets, such as VPS, SPS, and PPS, require the encoder to handle and transmit more information than in previous video coding standards. Additionally, the use of multiple views and frames in inter-view prediction and the new coding structures, such as the inter-view prediction structure, the disparity and motion compensation require the encoder to estimate and manage more motion vectors and reference frames, further increasing the computational complexity. Therefore, the MV-HEVC encoder may require additional resources and optimizations to maintain real-time encoding performance, especially for high-resolution and multi-view video sequences.

Modelling the computational complexity of video codecs is essential to estimate the computational requirements of a particular operation before its execution. However, this task is not straightforward as video encoders consist of multiple interdependent tools with parameterised functionalities that can significantly impact the level of computational complexity.

Another challenge of MV-HEVC is the propagation of errors between different

views. Since the encoder uses information from other views to compress a given view, any errors in the encoding of one view can propagate to other views. The propagation of errors can be particularly problematic in cases where the video content has large depth variations or when there is a significant disparity between the views. In such cases, the errors can become more visible, leading to a loss of quality in the reconstructed video.

A range of techniques and approaches have been put forward by the research community to enhance the efficiency and reliability of the MV-HEVC encoder, as well as to ensure that the resulting video output adheres to the highest standards of quality and clarity. The following sections provide a detailed exploration of these approaches, with particular attention to their ability to reduce computational complexity of the MV-HEVC encoder and mitigate the effects of errors on the final video output.

2.4.1 MV-HEVC Computational Complexity

A number of studies have been conducted to evaluate the computational complexity of MV-HEVC. In a study by W. Ahmad et al [56]. It was found that the computational complexity of MV-HEVC was significantly higher than that of HEVC for low-resolution videos, but the difference decreased as the resolution increased.

Another study in [57] found that the computational complexity of MV-HEVC was higher than that of HEVC for both high and low-resolution videos, but the difference was more pronounced for high-resolution videos.

The computational complexity of MV-HEVC is generally higher due to the increased computational requirements of processing multiple views. In particular, the processing of motion and disparity estimations, which are the more complex parts in MV-HEVC, as they involve encoding and decoding multiple views and predicting the motion of objects across different views. The complexity of motion and disparity estimation can be divided into two main stages: search and refinement.

The search stage involves searching for the best motion or disparity vectors for each CU or pixel in the current view, based on a search range and a cost function. The cost function is used to measure the quality of the predicted motion or disparity

vector and is based on the difference between the current block or pixel and its predicted value based on the motion or disparity vector.

The refinement stage involves refining the motion or disparity vectors found in the search stage to improve the overall quality of the encoded video. This may involve using more advanced techniques such as interpolation or filtering [58] to reduce artifacts or improve the accuracy of the predicted vectors.

Several fast algorithms have been proposed recently to reduce the computational complexity of the HEVC encoder [59, 60]. However, a limited number of methods have been devoted to the two extensions MV-HEVC [61, 62] and 3D-HEVC [63] of this standard. In fact, the majority of the proposed algorithms for the improvement of the HEVC standard can also be applied on the base view or the dependent views of the two extensions MV-HEVC and 3D-HEVC. The algorithms proposed for the reduction of the complexity of HEVC and its extensions have generally been based on three aspects: the CU, the Prediction Unit (PU) and the Transformation Unit (TU). This was done primarily by prediction of the best CU, PU and TU for improved encoding time.

Pan et al. [59] described a Fast motion estimation algorithm for improving the encoding time of the HEVC encoder. This approach can be used without inter-view encoding for the different views of the MV-HEVC. Basic thought behind this approach is the use of the best motion vector selection correlation for the various sizes of the prediction modes. This choice of using the best movement selection is made in [59] through two major steps. The initial step is to organize all PUs into two classes depending on the partition type. In the second step, when the parent PU uses the initial search point, in this case the block matching search process of the children PUs must be ignored. An important gain in ME encoding time is achieved by combining these two phase which can reach 20% on average. Nevertheless, in the case of MV-HEVC this method can be useful only for the simulcast prediction structure. In [60], Ahn et al presented a fast and an efficient CU encoding algorithm using the spatiotemporal encoding parameters of HEVC encoders. The idea on which this method is based is the prediction of the best CU split in advance. Therefore, the spatiotemporal encoding parameter data is used to evaluate the texture complexity

of encoded CU. Before this step, the CU block must be encoded in HEVC inter-slice coding. The number of the sample adaptive offset (SAO) edge categories 1 and 4 is then compared with that of categories 2 and 3 for the current CU. This comparison is made for making a rough CU classification. This approach allowed reducing up to 35% of the encoding time compared to the reference software HM 12.0. Nevertheless, using only the SAO parameters make this method limited in terms of using the encoding parameter in the procedure of CU depth decision.

G. Jiang et al. [61] explored the use of a 3DSobel model [64] for the proposal of a novel perceptual distortion threshold model (PDTM). The PDTM model consists primarily in reducing the computational complexity of the dependent view coding using a new fast inter-frame prediction algorithm in MVC. Then, the developed fast MVC algorithm is integrated into the MV-HEVC encoder to improve MVC coding efficiency. In practical coding, the thresholds derived from the PDTM are used to terminate the unnecessary mode selection for inter-frame prediction of dependent view. The computational complexity is reduced by 52.9% compared with the HTM14.1 algorithm. This gain is estimated only in relation to the dependent view. In [62], P. Wan et al presented a fast coding unit (CU) decision algorithm based on the use of spatiotemporal and inter-view information in addition to the maximum depth level of current CUs and motion vectors. The first step is to predict the maximum depth level of the related CU by exploiting spatiotemporal and inter-view CUs. Then, a check between the current CU and the nearby spatial CUs in terms of motion feature is performed in order to adjust the maximum depth level of the related CU. A promising encoding time gain of 50.02% on average is reported based on experiments carried out using this approach.

In order to reduce the 3D-HEVC computational complexity, Z.Pan et al. [63] proposed a motion and disparity vectors early determination algorithm. This approach is applied on the texture views which mean that it can be applied also on MV-HEVC. This method starts by adaptively reducing the spatial and temporal motion vector candidates for the prediction unit with the Merge mode. Then, a combination of spatial and temporal candidates is performed to early determine the final Motion Vector (MV) for the PU with the inter mode. The next step in this al-

gorithm relates to the inter-view encoding where an adaptive optimization algorithm is used for the selection of the valid inter-view disparity vecied also on MV-HEVC. This method starts by adaptively reducing the spatial and temporal motion vector candidates for the prediction unit with the Merge mode. Then, a combination of spatial and temporal candidates is performed to early determine the final Motion Vector (MV) for the PU with the inter mode. The next step in this algorithm relates to the inter-view encoding where an adaptive optimizattor. Another test used to skip unnecessary coding process is the difference between candidate vectors. If this difference is within a conditional range, the current PU must be encoded with the Merge mode. The encoding time saving obtained by this algorithm can reach 33.03%.

In [65], an Efficient Inter-Prediction Mode Decision (EIPMD) technique was proposed to decrease the complexity of the MV-HEVC encoder by presenting a novel approach for inter-prediction mode decisions of dependent views. The technique utilizes CU depth levels information of the co-located area of the base view to optimize decision-making for dependent views, with the observation that the views are only horizontally displaced and disparity is only in the horizontal direction. The proposed technique also utilizes the correlation between the temporal levels of pictures and the mode levels to optimize mode decisions and avoid intra-prediction in temporal level-4 pictures. Furthermore, the technique uses the average rate distortion cost of the CUs of the decision window and the correlation between current CU depth and mode decisions to enhance its performance. This technique decreased the encoding time by 66.56% on average compared to the reference encoder.

The previously mentioned work is a valuable contribution to the realm of low complexity MV-HEVC techniques. However, during our investigation, we have identified additional parameters that can be explored to optimize the MV-HEVC encoding process in a simpler manner. Hence, the focus of our research thesis is on a novel idea to reduce the encoding time of the MV-HEVC encoder.

To the best of our knowledge, there is no previous literature on removing the inter-view offsets to minimize the encoding time of both disparity estimation and inter-view prediction. Our proposed technique not only will introduce this novelty

but also will generate better results than the existing methods for low complexity MV-HEVC. Further details on this technique will be introduced in the next chapter.

2.4.2 MV-HEVC Error Propagation Effects

The main purpose of using NAL units in the MV-HEVC standard is to support various video transmission systems. However, in the event of transmission errors, compressed MVV bitstreams become more susceptible to these errors, due to saptio-temporal and inter-view dependencies in the MVV sequences as shown in Figure 2.4. In which, an important dependency structure is introduced by motion compensation, where each reference frame is used to predict another frame. Another set of dependency structures is introduced in layered video coding where the base view is referenced by enhancements views, and each enhancement view can be further referenced by other enhancement views.

The I frame at the beginning of a base view serves as a basic entry point to facilitate random seeking or channel switching and also provides coding robustness to transmission error, but it is coded with only moderate compression to reduce the spatial redundancies. P frames are coded more efficiently using motion compensated prediction from a past I or P picture and are generally used as a reference for further prediction. B frames provide the highest degree of compression, but require both past and future reference pictures for motion compensation. In error-prone environments, error happening in the previous frames in a GOP may propagate to all the following frames and views. The effects of displaying one of these frames reconstructed from corrupted data can adversely degrade visual perception in the decoded video.

Various techniques have been proposed in the literature to enhance the robustness against video coding and communication errors. Error Resilience (ER) is one of the most effective solutions for mitigating the effects of errors in real-time video transmission. Which are applied at the encoder side. One of the approaches based on the ER technique is the Feedback Channel (FC) [66], in which the receiver can request the retransmission of some missing data. Channel Coding or Forward Error Correction (FEC) [67] is another approach based on the ER technique. The basic

idea of this approach is to protect video sequences with Error Correction Codes (ECCs). This technique inserts redundant data into the encoded video for detecting missing data on the decoder side, and to recover them without retransmission. Multiple Description Coding (MDC) or Layered Coding (LC) [68] is another group of ER without using complex channel coding schemes. MDC and LC encode a media source into multiple sub-bitstreams. The main difference lies in the dependency among bitstreams. In which, all sub-bitstreams for LC are dependently generated while each sub-bitstream for MDC can be independently encoded. The robustness comes from the fact that a high quality reconstruction can be achieved from all sub-bitstreams at the decoder, as shown in Figure 2.9.

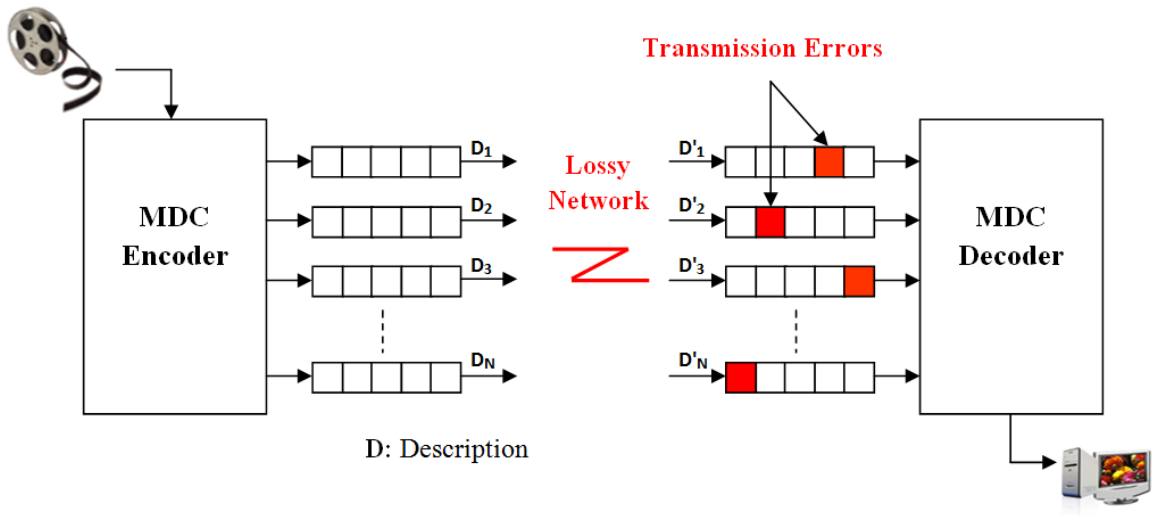

Figure 2.9: Enhancing Error Resilience with Multiple Description Coding (MDC).

However, the coding efficiency is reduced depending on the amount of redundancy between sub-bitstreams. Several works exploit intra refresh [69]-[70] techniques to improve video robustness. In video coding, an inter-coded frame can be affected by a reference frame with errors, and it can propagate the error to other inter-frames which use it as reference. Thus, defining appropriate intra refresh encoding modes and intra-coded frames with an adequate selection of reference frames can make bitstreams more robust.

Unlike ER methods that actively place error protection at the encoder side, Error Concealment (EC) techniques estimate and predict the missing data in an attempt

to conceal errors in the decoded stream. ECs predict the missing parts and correct them by exploiting spatial and temporal similarities in the video sequence. Spatial prediction can be only used for small areas, whereas the temporal prediction can replace a whole missing frame with a previously received frame [71]. This is known as Frame Copy Concealment. A variety of EC techniques based on motion vector (MV) information can also be used to estimate the content of a missing area [72].

The approaches discussed in the first and second paragraph of this section have been devoted to improve the quality and transmission of single-view videos, while little efforts have been reported in the literature on error robustness of multi-view and 3D video data. Song [73] proposed an efficient EC algorithm in Multi-view Video Coding (MVC) to conceal blocks losses by exploiting intra-view and inter-view correlations using three EC modes and hierarchical B prediction structures [74]. Pan et al [75] described an efficient mode switching algorithm to improve 3D video reconstruction qualities. The idea is based on the analysis of the impact of texture and depth map video reconstruction errors on the quality of the virtual view. In [76] the multi-view video end-to-end distortion is estimated at the encoder according to the inherent error resilience property of the hierarchical B prediction. In this model, blocks at regions that will cause significant drifting errors can be refreshed with intra blocks to achieve the trade-off between coding efficiency and error robustness. Dissanayake [77] presented an error resilient method by incorporating the disparity vectors into MVC to generate a redundant data stream. Another work in [78] examined the effect of GOP size on erroneous multi-view Data Partition (DP) bitstream when transmitted over error-prone networks. Two different techniques namely H.264 DP and multi-layer DP are used to analyse the bitrate performance for different GOP and error rates on MVC. In this study, the Multi-layer DP technique showed a better performance at higher error rates with different GOP.

ER and EC techniques are crucial for minimizing the impact of error propagation in video standards. However, the development of a reliable ER or EC algorithm relies heavily on an in-depth analysis of error propagation effects and identifying the most vulnerable components of the video data. Given the lack of error analysis in the MV-HEVC standard, chapter 4 of this study will propose an analysis of error

robustness in the MV-HEVC standard to fill this gap in the literature.

2.5 Data and Test Conditions

During the standardization of the HEVC video coding standard, the (JCT-VC) group coordinated the development of a reference software encoder and decoder referred as HTM (High-Throughput Multimedia), as well as a set of testing guidelines known as the Common Test Conditions (CTC) document.

2.5.1 HTM Reference Model

The High-Throughput Multimedia (HTM) is a widely used test model for evaluating the performance of video compression algorithms, especially for HEVC extensions. HTM provides a set of software implementations specifically designed for this purpose, including three separate software packages that are used to compare and evaluate the efficiency of video compression algorithms. These packages provide a comprehensive framework for testing various aspects of the compression process, including encoding speed, complexity, and visual quality.

By leveraging the HTM test model, researchers and developers can ensure that their video compression algorithms meet the high standards of efficiency and effectiveness required for modern video applications.

The following subsections will provide an explanation of the three distinct software implementations included in HTM.

2.5.1.1 HEVC Test Model (HM)

HM is the first and most widely used software implementation of the HEVC standard. HM is a reference software that was developed by the Joint Collaborative Team on Video Coding (JCT-VC), which is a joint committee of the ISO/IEC MPEG and ITU-T VCEG.

HM provides a flexible framework for evaluating the performance of various HEVC algorithms and encoding techniques. It includes a set of configurable encoding parameters that enable researchers and developers to experiment with different

encoding strategies and compare their results. Additionally, HM supports a range of different video formats and resolutions, making it suitable for testing and evaluating different applications and use cases.

2.5.1.2 MV-HEVC Test Model

MV-HEVC is an extension of HM that supports multi-view video coding (MVC) and scalable video coding (SVC) in addition to HEVC. MV-HEVC is designed to test the performance of these additional features and to facilitate the development of new algorithms for these types of video coding. It includes in HM framework the encoding and decoding of multiple views of a scene in a single bitstream. It also supports different coding structures for MVC, including the inter-view prediction and the view synthesis prediction.

2.5.1.3 3D-HEVC Test Model

3D-HEVC Test Model is a specialized software extension of HM designed specifically for 3D video encoding. It supports advanced features of the HEVC standard, such as multi-view video coding and depth-based coding for 3D video content. Unlike the MV-HEVC Test Model, the 3D-HEVC Test Model does not support scalable video coding but provides additional tools and features that are specifically designed for 3D video coding. These features include support for depth maps and more efficient inter-view and intra-view dependencies coding.

All three implementations are freely available for download and can be used to test and evaluate the performance of different video coding algorithms. The source code is also available, allowing developers to modify and customize the software to suit their needs.

2.5.2 Common Test Conditions (CTC)

In video compression, CTCs are a set of standardized test conditions used to evaluate the performance of a video codec. For MV-HEVC, the CTCs would typically include standardized test conditions such as the video content used for testing, the number

and type of views being encoded, the frame rate, the bitrate, and other relevant parameters.

These CTCs are important to ensure that different implementations of the MV-HEVC codec are evaluated under the same conditions, which makes it easier to compare and benchmark their performance. This can help identify areas where improvements can be made to the codec, as well as provide guidance to users on how to optimize their encoding settings for the best possible video quality and compression efficiency. Common test conditions of MV-HEVC typically include:

- **View configuration:** MV-HEVC is designed to support stereoscopic and multi-view video content, so testing should include multiple views of the same scene or object. Different view configurations can be tested, including side-by-side, top-and-bottom, and interlaced views.

- **Frame rate:** The frame rate refers to the number of video frames per second. MV-HEVC can support a range of frame rates, so testing should include different frame rates to evaluate its performance at different frame rates.

- **Bitrate:** The bitrate refers to the amount of data that is transmitted per second. MV-HEVC should be tested at different bitrates to evaluate its compression efficiency and to determine the appropriate bitrate for a given video quality.

- **Video content:** MV-HEVC should be tested using a variety of video content, including video with high and low motion, varying levels of detail, and different color palettes. This helps to evaluate MV-HEVC performance across different types of video content.

- **Encoder settings:** Different MV-HEVC encoders may have different settings that can affect the output video quality and compression efficiency. Testing MV-HEVC with different encoder settings can help determine the optimal settings for a given application.

- **Disparity estimation:** Disparity estimation is a key component of MV-HEVC, as it is used to identify and compensate for the differences between

different views of the same scene or object. Testing should include different disparity estimation algorithms to evaluate their performance and to determine the optimal algorithm for a given application.

2.6 Conclusion

This chapter discussed the challenges posed by the high computational complexity of the MV-HEVC encoder and the need for optimization to achieve real-time processing and error robustness.

The various optimization techniques proposed for the MV-HEVC codec were explored, targeting various aspects such as Intra-prediction, Quad-tree CU and PUs splitting process, and inter-prediction modes.

Our upcoming research work, however, will focus on reducing the encoding time of the inter-view prediction by optimizing disparity and motion estimations of MV-HEVC encoder. By doing so, we aim to effectively reduce the complexity of the MV-HEVC encoder.

Furthermore, the chapter also covered the methods proposed in the literature to enhance robustness against multi-view video coding and communication errors, with a particular focus on error resiliences (ER) and error concealments (EC), emphasizing their crucial role in minimizing the impact of error propagation in video standards. It was further highlighted that developing a robust ER or EC algorithm requires a thorough analysis of error propagation effects and identifying the most vulnerable components of the video data. In light of this, chapter 4 will propose an analysis of error propagation in the MV-HEVC standard to address the lack of error analysis in the literature.

Finally, an explanation of the common test conditions and the MV-HEVC test model was provided, which are essential for evaluating the performance of the MV-HEVC codec.

Chapter 3

Improved Inter-view Correlations For Low-Complexity MV-HEVC

‘

3.1 Introduction

The latest advanced multi-view extension, MV-HEVC, effectively exploits visual similarities between multi-view videos and enables high compression efficiency. Each view in the multi-view sequence depends on the captured scene, the distance between cameras and recording angles. Increasing the distance between dependent viewpoints generates an inter-view disparity. This impacts the inter-view similarities, affects the disparity estimation and further increases the computational complexity of the MV-HEVC encoder.

In this chapter, an efficient Earlier Disparity Estimation (EDE) algorithm is proposed for low complexity MV-HEVC. This algorithm is based on reducing the complexity of disparity estimation by eliminating the inter-view offsets. Moreover, the inter-view similarities are controlled by considering the reliability of each coding unit size in the search range. This reliability is estimated by reducing the number of searching points within a new limited window.

For reliable motion estimation, we further proposed an Earlier Decision of Coding units Splitting (EDCS) in the dependent views according to those in the ref-

erence views. Experimental results show that the proposed algorithm can achieve an average encoding time saving of 20.37% - 40,61% with marginal performance degradation. This research work is published in [79].

3.2 Inter-view Prediction: Motivation and Statistical Analyses

In preliminary work, an experimental analysis of MV-HEVC inter-view prediction is performed to study the rate of CUs used during inter-view prediction. The objective of this analysis is to assess the inter-view coding efficiency during MV-HEVC encoding process as well as to evaluate the disparity between the dependent views and in particular between the more distant ones. This assessment was conducted using multiple HD-MVV sequences with different contents and qualities. Table 3.1 describes the used videos sequences and their parameters. Also, two prediction structures are used in this analysis: IP by the use of two views and IPP using three views. Analysis results are given in Table 3.2 and 3.3.

Sequences	Resolution	Frames	Frame Rate (fps)	Cameras spacing (cm)
Champagne-tower	1280x960	300	30	Stereo
Panatomim	1280x960	300	30	Stereo
Dog	1280x960	300	30	Stereo
Balloons	1024x768	300	30	5
Kendo	1024x768	300	30	5
Sharek	1920x1088	300	30	Stereo

Table 3.1: HD-MVV sequences used in experimental analysis

As can be seen in Table 3.2, the rate achieved by IP structure is very high, exceeding 37.50%, 40.30% and 40.70% for Dog, Pantomim and Kendo sequences, respectively. Furthermore, it can be clearly seen from Table 3.2 that the rate for high-definition videos such as 1920 x 1088 is much higher than that for low-definition

	Sequences					
Views	Champagne-tower	Pantomim	Dog	Balloons	Kendo	Sharek
V0&V1	9.04	40.39	37.52	22.15	40.70	91.61
V0&V2	8.91	39.68	37.37	20.86	37.07	84.27
V0&V3	8.80	30.57	33.63	20.56	37.17	78.36
V0&V4	8.68	30.34	31.55	20.48	36.57	72.92
V0&V5	8.45	28.22	31.05	20.18	34.32	69.33
V0&V6	7.48	25.44	30.89	19.31	34.60	65.51

Table 3.2: Inter-view prediction rate in MV-HEVC for IP structure

	Sequences					
Views	Champagne-tower	Pantomim	Dog	Balloons	Kendo	Sharek
V0,V1&V2	8.97	40.03	37.44	21.50	38.37	87.94
V0,V2&V4	8.79	35.01	34.46	20.67	36.81	78.59
V0,V3&V6	8.08	28.01	32.25	19.93	35.88	71.93

Table 3.3: Inter-view prediction rate in MV-HEVC for IPP structure

videos of 1280 x 960 and 1024 x 768, whereby a gain of 91.61% is obtained for Sharek sequences. Similarly, Table 3.3 shows that the rate obtained using three views is also high. It exceeds 37.44%, 38.37% and 40.03% for Dog, Pantomim and Kendo sequences, respectively. The important rate is always obtained for Sharek sequences, which exceeds 87.94%.

The aforementioned results have shown that a large part of the CUs is obtained only through inter-view prediction during the MV-HEVC encoding process. This means that the inter-view similarity is very important. Accordingly, if we can take full advantage of this similarity to simplify Disparity Compensation Prediction (DCP) process, the encoding time will be greatly reduced in the whole MV-HEVC encoding process.

Table 3.2 and Table 3.3, also show that the rate of CUs at the inter-view level decreases with the increase in the distance between the different views, whether for

the structures of two or three views. The inverse relationship between interview distance and CUs rate is mainly due to the diminution of the inter-view similarity each time. More precisely, the capture of the same scene simultaneously from different angles creates an interview offset between the adjacent views as shown in Figure 3.1. This may be accompanied by low precision of CUs in the search range during disparity estimation.

In this research, we exploit the high rate of CUs at the interview level and also the interview offset between the adjacent views to reduce the computation time for motion and disparity estimation during MV-HEVC encoding process.

3.3 MV-HEVC Disparity Estimation and Motion Estimation Improvements

3.3.1 Elimination of the Interview offsets

The multi-view video coding is mainly based on three coding levels: spatial, temporal and inter-view. The coding quality in each level depends on certain criteria. At the temporal level, this quality strongly depends on the similarity of the information between the images, in particular the successive ones. Within a MVV sequence, using multiple viewing angles to capture the same scene results in a slight offset between adjacent views as shown in Figure 3.1.

The similarity between the views is therefore considered very important, especially other than the inter-view offset zones. The offset between simultaneously captured scenes varied depending on several settings of the camera array, such as the distance between cameras and the difference in viewing angle between views, as well as the depth of the objects in the scene. To increase the accuracy of inter-view motion estimation, disparity vectors are required to identify PUs of the same objects between adjacent views. Knowing the disparity for a given PU allows the prediction of that PU from the corresponding disparity-compensated PU in the neighboring view. Table 3.4 shows the rate of PUs predicted only from the offset part.

As it is evident in Table 3.4, the rate of PUs predicted from this inter-view offset

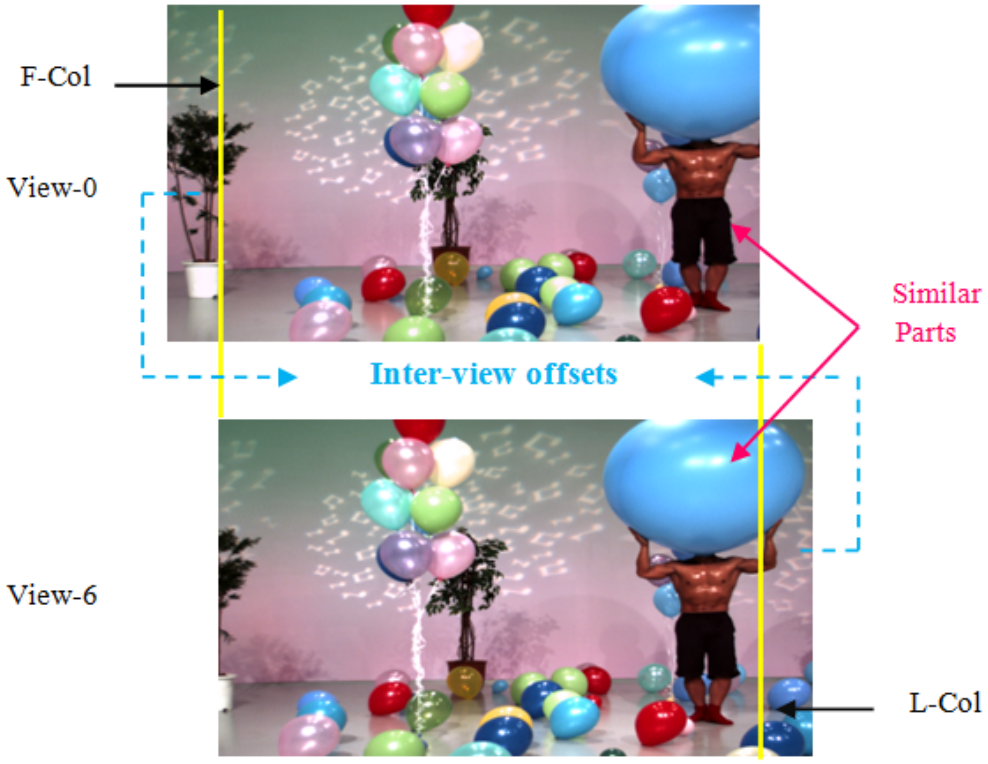

Figure 3.1: Inter-view offsets between the adjacent views.

	Dependent Views					
Sequences	View 1	View 2	View 3	View 4	View 5	View 6
Champagne_tower	0.60	0.72	1.32	2.17	2.89	4.08
Pantomim	0.79	1.17	1.58	1.70	2.80	3.69
Dog	1.18	2.79	3.66	4.47	4.66	5.18
Balloons	2.75	3.22	3.31	4.88	6.43	9.92
Kendo	1.08	1.77	4.47	5.01	6.09	7.2
Sharek	0.03	018	0.40	0.68	0.84	0.95

Table 3.4: The percentage of predicted units from shift parts

increases with increasing distance between views, which the highest rate is observed between views 0 and 6. It starts from 0.95% for Sharek sequences and achieves 9.92% for Balloons sequences. This is a negligible rate, but can consume a substantial amount of time in inter-view encoding process which increases the complexity of the

MV-HEVC encoder. Additionally, increasing the inter-view displacement between views affects the disparity estimation, in which a small difference in estimation of a single pixel may vary the inter-view prediction. The main focus of our proposed approach is essentially based on this constraint of inter-view similarity.

The basic idea of the proposed approach in this research is to make the inter-view information visually similar as shown in Figure 3.1. This can be done according to the previous analysis by eliminating the inter-view offset. The main aim of this visual adjustment between the different views is to improve the inter-view correlation.

Figure 3.1 illustrates an example of the inter-view offset between two images of the same temporal level. Elimination of inter-view offset is achieved by cutting out both sides of the offset from both views. That is, the first part to be eliminated is located at the beginning of the images of the views on the left while the second part is at the end of the images of the views on the right. To clarify, the right and left views are obtained by using the right and left cameras. The parts eliminated from the two views must then be analyzed and processed separately.

3.3.2 Earlier Disparity Estimation Algorithm (EDE)

Disparity estimated between two images of two views that are captured simultaneously can be done using several approaches. Usually, this disparity estimation results in a disparity map for all pixels. Note that the proposed method seeks to find the offset portions between the two views, not the disparity for all pixels. In fact, the proposed approach focuses on the detection of the first and the last column delimiting the similar part between the two images in the used views. An example of the two columns to be detected is shown in Figure 3.1. These two columns should be determined from the left and right views respectively. The detection of these two columns delimiting the inter-view offset is performed using the similarity between all the columns of the two frames in the two views. This method is mainly applied on MVVs captured using linear arrangement of cameras where the similarity between the different views is very important. The first and the last column to be detected are called $F-Col$ and $L-Col$ respectively. The $F-col$ is always in the first view which is on the left while $L-Col$ is in the second view which is on the right. In this

research work, the $F-Col$ is always shown in view-0. The $L-Col$ can be found in the other views from view-1 to view-6. The similarity measure used in this case between the different columns is the Euclidean distance. The $F-Col$ is the most similar column to the first column of the second view. Indeed , the $L-Col$ is the most similar column to the last column of the first view considering that $L-Col$ is still in second view. The Euclidean distance can be, therefore, calculated between the first column of the second view and all the columns of the first view. In this case, the $F-Col$ is the one that has the minimum distance. The $L-Col$ can be obtained by the same method using the last column of the first view for the calculation of the Euclidean distance. Using the first column only for $F-Col$ detection does not always give good results. We thus calculated the Euclidean distance for all the columns of the second and the first view for the detection of the $F-Col$ and $L-Col$ respectively. Therefore, the general steps followed in this method to obtain the $F-Col$ can be summarized as follows:

1. Calculate the Euclidean distance between column number i of the second view and all the columns of the first view; i starts with the value 1. We consider that the minimum distance is Dmin.

2. Store the number of the column that has the minimum distance Dmin in a vector V of size equal to the column number of the video. This number must be stored in position i of the created vector.

3. Repeat the two steps 1 and 2 for all the columns of the second view, adding the value 1 to the variable i each time.

4. In this step, we apply a subtraction between the content of the vector V and its indices, starting with the index 1.

5. Thus, F-Col corresponds to the most repeated number in the result vector V which is obtained after step 4.

These steps can be given by,

$$VRes(i) = EDis(CR_g(i), Frame_{Left}),\ i = 1..M \tag{3.1}$$

$$F_{Col} = Occ(VRes(i) - i),\ i = 1..M \tag{3.2}$$

The *VRes* is the vector of size equal to the column number of the video. This vector contains the column numbers of minimum distances. The *CRg* corresponds to the columns of the right views while *Occ* denotes the function that returns the most repeated number in *VRes* and M is the column number of the video. The Euclidean distance is represented in 3.1 by *EDis*. $L - Col$ can be obtained by the same method using all the columns in the left view.

Using a prediction structure of three views with two dependent views requires the estimation of the inter-view offset between the base view and the two dependent views and also between the dependent views them selves. Figure 3.2 illustrates an example of the IBP prediction structure which uses three views.

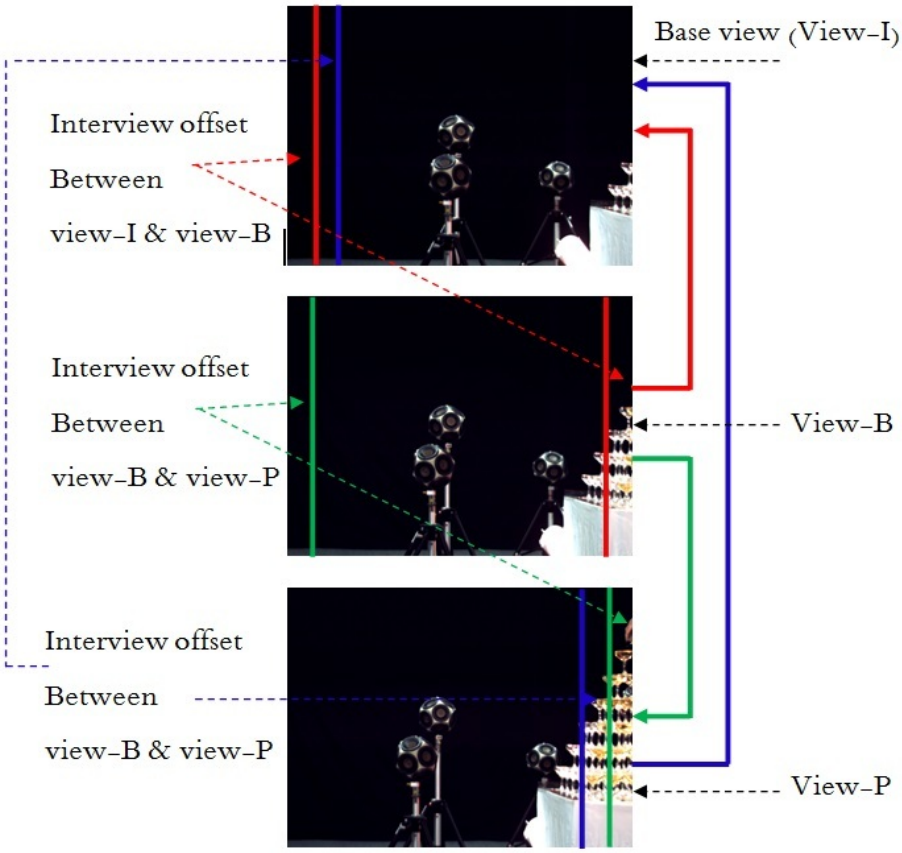

Figure 3.2: Inter-view offsets for three views.

This figure shows the different inter-view offsets where the largest offset is between the base view and the view-P. The size of the offset between views is mainly

due to the distance between these views. Increasing the distance between cameras gives a greater offset between the views. The greatest distance is between the base view and the view-P.

The coding order in the case of three views is started with the base view, then the view-P uses the base view as the reference view for its coding and finally the view-B uses the base view and the view-P as reference views for its coding. The elimination of the inter-view offset is applied during the coding of the two views P and B which use the inter-view coding. In the case of view-P, the same steps previously presented for the case of two views are applied for the designation of the $F-Col$ and $L-Col$. The same method is applied when coding view-B which uses two reference views. In this case, determining $F-Col$ and $L-Col$ is carried out first between view-B and the basic view ($F-Col$ in the basic view and $L-Col$ in the view-B) and then between the view-B and view-P ($F-Col$ in view-B and $L-Col$ in view-P) as shown in Figure 3.2.

To sum up, the inter-view correlation can be improved by eliminating the inter-view offset. The most evident advantage of improving the inter-view correlation is the improvement of the encoding time of the disparity estimation process. This improvement is essentially based on reducing the size of the search window as presented in the next section.

3.3.3 Improved ME/DE for dependent views

The disparity estimation is the most important process in the greater part of MVC standards. Nevertheless, the computational complexity of the encoder significantly increases using this technique. It is worth mentioning that 70% of the total encoding time in the HEVC coding standard is consumed in the ME/DE process. That has been tied to the fact that this technique is the most important block finding process for the different blocks of the current frame. The block finding process is carried out in a search window of the reference frame. The computational complexity has a direct relation with the size of the search window in the ME/DE process. The encoding process thus consumes more computing time for a large search window. An example of a finding process in a search window is shown in Figure 3.3.

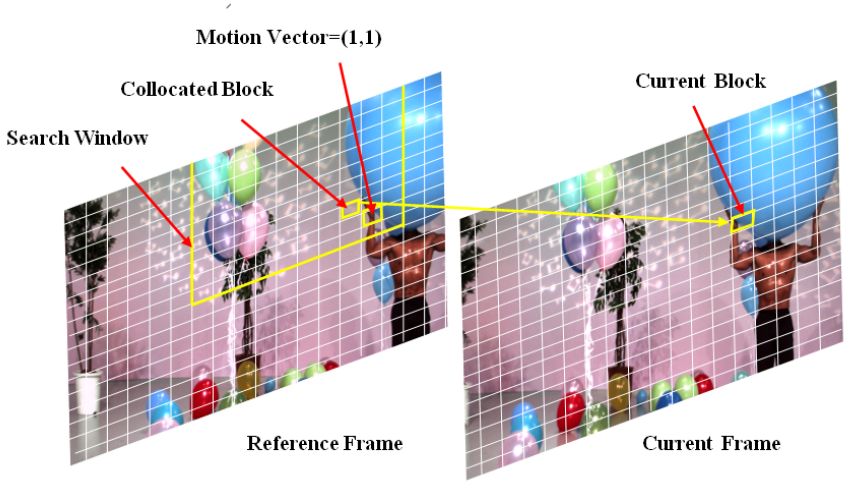

Figure 3.3: Illustration of the search window in ME / DE processes.

The algorithm presented in this research essentially aims to decrease encoding time for DE process. Basically, the main idea revolves around the visual adjustment between views by offset elimination as presented in previous section. After eliminating the inter-view offset, the first step in the coding process is to reduce the search window in DE process. This reduction is achieved depending on the CUs depth levels for the current frame from the reference image which is located in the base view. In fact, this reduction can be justified by the visual adjustment between the current frame and the reference one by the elimination of the inter-view-offset. To this end, we opted to increase the size of the search window according to the four CUs depth levels D0, D1, D2 and D3. Therefore, the four sizes of the search window that are used for the four depth levels are 4,8,12 and 16, respectively. Note, that these four sizes are only used at the inter-view level for the disparity estimation. In fact, the reduction in the size of the search window especially for CUs depth levels D0 and D1 is justified by the inter-view visual adjustment applied in our approach.

At the temporal level, the reference frames can be either similar or completely different from the current frame. Unlike the temporal level, the reference frame at the inter-view level is certainly similar to the current frame. This property is justified by the fact that the base view and the dependent view which contain the reference frame and the current frame respectively are captured simultaneously

from two different angles. Thus, this capture constraint generally results in very similar video sequences. In the process of disparity estimation for the D0 level for example where CU size is 64x64, the use of a small search window is due to the prior knowledge that the most similar information of the current CU in the base view cannot be moved away. This is due to the fact that we have already eliminated the inter-view offsets as mentioned earlier. The larger the CU, the less likely it is that the most similar CU will be found far from it in the base view. This explains the relationship between the search window size and the CUs depth level.The use of the small sizes of the search window, in particular for the two levels D0 and D1, can achieve a significant reduction in computational complexity.

3.3.4 Earlier Decision of Coding units Splitting Algorithm (EDCS)

After visual adjustment and decreasing the size of the search window in the DE process, the second step is the earlier decision for CU splitting (EDCS). The EDCS are performed according to the decisions obtained from basic view in terms of selection modes as the content of the base view and the dependent views is almost the same as presented above. Selecting the best match is considered as one of the most computationally time consuming processes. The selection of the best match is made in this step without having all possible combinations of sizes in the available reference frame. This is mainly due to the visual adjustment made in the first step. Earlier decisions are made in this step in dependent views only for the temporal level, that is to say just for the ME process.

Several previous studies [62, 80, 81] proved that during the CU size selection process, the highest percentages are best matched at level D0 and D1.This is also due to increased similarity between base view and dependent views.These studies also demonstrated that the CU splitting decisions made by the encoder for the basic views are very similar to those made for the dependent views. For those purposes and in order to achieve a good compromise between encoding time and quality, only the CUs depth levels D0 and D1 are concerned by the EDCS. This means that in

the reference frame of the base view if the size of the used CU is 64x64 or 32x32, the current block of the dependent view must therefore use the same mode directly.

The visual adjustment between the base and the dependent view is obtained by cutting out the inter-view offset from the different used views. As argued in the previous section, the first part of the inter-view offset is at the start of the base view while the second is located at the end of the dependent view. To further reduce the computational complexity in the MV-HEVC, we have eliminated the inter-view prediction for the offset parts. This action is justified by the total absence of information from these parts in the reference frames at the base view. By removing the inter-view prediction for the offset parts, we significantly reduce the computational complexity involved in analyzing and processing regions that lack any meaningful reference information, thereby improving the overall encoding efficiency of MV-HEVC.

3.4 Experimental results

3.4.1 Experimental settings

The proposed algorithm is implemented on the recent MV-HEVC reference software, the HTM version 16.2. Experimental results presented in this study were obtained using six HD-MVV sequences(Champagne-lower, Panatomim, Dog, Balloons, Kendos, and Sharek) as shown in Table 3.1. The six MVV sequences used in the evaluation of the proposed approach are captured by linearly arranged cameras where each video is made up of seven views. The values of the quantization parameters (QP) used during this evaluation are 25, 30, 35 and 40.

The detailed simulations were performed using the common test conditions (CTC) [82] required by JCT-3V. In this research study, we have utilized three inter-view prediction structures for the evaluation of the proposed method which are IP, PIP and IBP. These prediction structures use two (IP) or three views (PIP and IBP). Table 3.5 illustrates a number of configuration parameters that we used during the encoding process. As mentioned in the previous section, the size of the search window changes according to the four CUs depth levels D0, D1, D2 and D3. The values

thus used representing the size of this search window are 4,8,12 and 16.

Setting	Value
Prediction Structure	IP and PIP
GOP length	8
Intra period	24
Coding tree block	64x64 pixels
CU depth level	4
Search range	Variable (4,8,12 and 16)
Entropy coder	CABAC

Table 3.5: Encoder configurations

3.4.2 Experimental results using tow views (V1 & V2)

The distance between the different cameras used when capturing views directly influences the disparity between these views. That is, the inter-view disparity changes depending on the distance between the cameras. Elaborate analysis is devoted on the relationship between the disparity eliminated each time from two views and the saving encoding time.

The experiments are performed using the prediction structure called IP. The base view used in this case is always V0 while the dependent view can be any of the V1 to V6 views. The evaluation of the proposed approach is carried out by the three parameters: change in encoding time $\Delta time$ (%), percentage change in the bit rate $\Delta Bitrate$ (%) and the change in PSNR $\Delta PSNR$ (dB). These parameters can be calculated using Eq.3.3.

$$gain = \frac{Proposed - HTM}{HTM} \tag{3.3}$$

Where *Proposed* denotes the coding results of the proposed approach and *HTM* is the coding results of the HTM software.

Table 3.6 summarizes the obtained results by the proposed approach (EDE) with respect to the HTM reference software using two views V-0 and V-1.

As shown in Table 3.6, the proposed method reduces the total encoding time from 20.37% to 29.26% using QP equal to 25 for the six used videos. Increasing the value of QP gradually improves the saving encoding time. Nevertheless, this increase in the value of QP causes reduction of the encoding quality and the used CUs depth levels. This implies that the size of the CUs increases according to the value of QP. Indeed, the size of the search window is also affected by any change in the value of QP and CUs. This volume decreases with each increase in QP, and vice versa. The values presented in Table 3.6 in terms of $\Delta time$ can thus be justified by this relation. The obtained results show that the $\Delta time$ varies from one video to another may be due to a number of reasons, including the nature of the video itself and the distance between the cameras. The $\Delta time$ is less important for the two videos "Kendo" and "Balloons" that the distance between the different cameras of these two videos is slightly lower than the other two videos.

The rate distortion (RD) performance degradation of the proposed method in terms of $\Delta PSNR$ and $\Delta Bitrate$ is negligible as shown in Table 3.6.

3.4.3 Experimental results using two views (V1 up to V6)

Figure 3.4 illustrates the obtained results by moving the dependent view further away from the basic view.

The base view is always V0 while the dependent views are V1, V2 up to V6. When the distance between the cameras increases, this makes the inter-view disparity more important. The information removed from the inter-view encoding has a relation of proportionality with this disparity.

The elimination of information during the inter-view compression process means a gain in search time in this eliminated part and therefore a reduction in encoding time. As results, we got a $\Delta time = 39.20\%$ on average for the "Pantomime" video using the V-6 depend view and a QP of 25. As shown in Figure 3.4, in the case where QP is equal to 40, this gain exceeds 48%. Figure 3.5 shows the obtained Rate Distortion (RD) performance degradation for the IP prediction structure using the dependent views from V1 to V6.

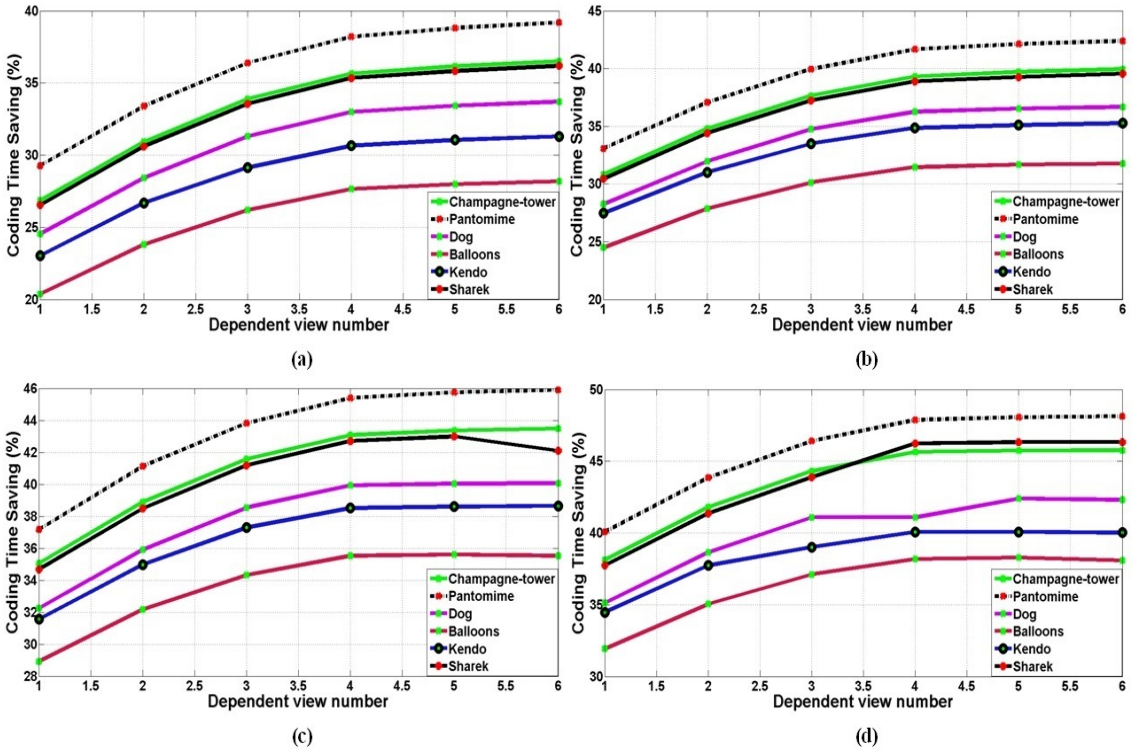

Figure 3.4: The saving encoding time according to the increase in inter-view offset: (a) QP = 25 (b) QP = 30 (c) QP = 35 (d) QP = 40.

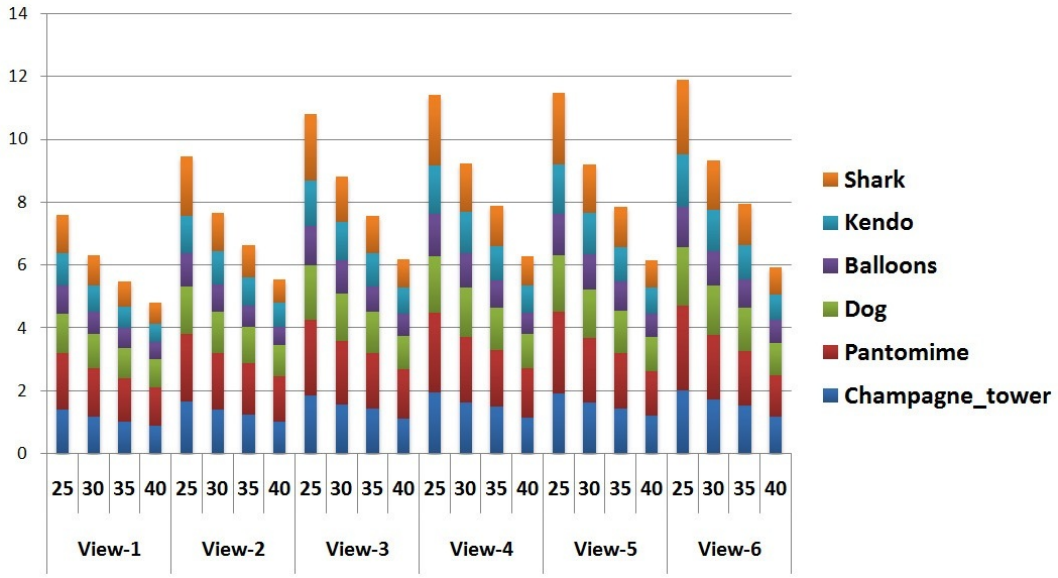

Figure 3.5: The RD performance degradation using two views.

3.4.4 Experimental results using three views

This section presents the obtained results using three-view prediction structures. The objective is therefore to analyze the elimination of disparity before encoding using three views. The first three-view prediction structure used in this experiment is PIP. As with the IP prediction structure, the parts to be eliminated during inter-view encoding are located in the three views used which are the base view I and the two dependent views P. The removal of these parts depends on the order of each view in the prediction structure. This analysis is applied first on adjacent views V0, V1 and V2 where V1 is the base view and V0,V2 are dependent views. Then, the case where the dependent views are far from the base view is tested. In this case, the new base view becomes V2 and dependent views are V0 and V4. the obtained results in terms of $\Delta time$, $\Delta Bitrate$ and $\Delta PSNR$ are summarized in Table 3.7.

The obtained results show that the use of three views significantly increases the $\Delta time$. This increase can be attributed to the use of two dependent views where the number of parts to be eliminated is doubled compared to two views. The obtained saving time from the two dependent views is almost similar because the parts removed from these two views are equal. The best saving encoding time in the case of three views is also obtained for video "Panatomim" from 35.41% to 46.13% on average respectively for QP equal to 25 and 40. Spacing the cameras from each other increases the disparity between the views which has increased the saving encoding time as shown in Table 3.7. The obtained $\Delta time$ in this case exceeds 46.00% on average for QP equal to 40.

In order to analyze the impact of the proposed approach for bidirectional prediction of B-views, the IBP prediction structure is used in this experiment. This impact can be expressed by the efficiency of the compression in terms of trade-offs between bit-rate and quality and also encoding time $\Delta time$. The process of eliminating the inter-view offset in the case of the IBP prediction structure is different from that of the PIP structure, especially for the view-B. As presented in Section 3, the elimination of inter-view lag is based on visual adjustment which is essential in this approach for improving the disparity estimation. Each frame in view-B uses two inter-view reference images from the base and the dependent views respectively.

In the disparity estimation from the base view the eliminated parts from B frames are not the same as those in the estimate from the P-view. The change between the two is mainly in the position of these parts in the reference views and also in the B-view. Tables 3.7 and 3.8 show that the obtained saving encoding time is greater compared to the PIP prediction structure.

Parts eliminated from view-B in addition to base view I and view-P greatly improves the disparity estimation time. the visual adjustment obtained after elimination of the inter-view offset also makes it possible to bring similar inter-view blocks closer together, which reduces the block search window and thus the encoding time in the temporal level. Another important criterion also influences the encoding time is the distance of the view-P from the base view I beside PIP structure. The results presented in Table 3.8 show that the bit rate increment is greater in the IBP prediction structure compared to the PIP structure. The bit rate increment in this case is due to the increase in the eliminated parts of the inter-view prediction. Obviously, the bit rate decrement with the use of inter-view prediction in addition to the temporal prediction.

3.5 Conclusion

The compression efficiency of MVVs particularly that concerning the inter-view prediction is greatly improved with the coding tool-sets of the MV-HEVC standard. Hence, improvements built into the MV-HEVC encoder have significantly increased the computational complexity of the encoding process.

In this chapter, we proposed the Efficient Earlier Disparity Estimation (EDE) approach for MV-HEVC standard by improving the inter-view prediction and disparity estimation processes. Our approach involves several steps:

- Firstly, an inter-view visual adjustment is performed by eliminating inter-view offsets caused by different viewing positions. The elimination of the inter-view offsets is justified by the negligible rate of the PUs predicted from these inter-view offsets.

- Then, we decrease the DE search window based on the inter-view similarity and the split levels of CUs.

- At last, we use the CUs splitting decisions that are obtained from the reference views for the proposal of the Earlier Decision of CUs Splitting (EDSC) in the dependent views.

Experimental results demonstrated that the proposed algorithms provided promising encoding saving time using inter-view visual adjustment with negligible effect on RD performance. Furthermore, the proposed algorithms can be extended to the 3D version of the HEVC compression standard, offering a decrease in computational complexity.

Sequences	QP	$\Delta Time$	$\Delta Bitrate$	$\Delta PSNR$
Champagne_tower	25	-26.87	1.39	-0.05
	30	-30.82	1.17	-0.03
	35	-35.05	1.02	0.02
	40	-38.11	0.89	0.02
Pantomime	25	-29.26	1.79	-0.09
	30	-33.06	1.53	0.00
	35	-37.17	1.36	-0.03
	40	-40.07	1.23	0.07
Dog	25	-24.53	1.27	-0.07
	30	-28.26	1.09	0.02
	35	-32.25	0.98	0.01
	40	-35.12	0.88	0.04
Balloons	25	-20.37	0.89	-0.05
	30	-24.49	0.73	-0.03
	35	-28.91	0.63	-0.02
	40	-31.93	0.55	0.02
Kendo	25	-23.03	1.02	0.01
	30	-27.47	0.82	0.07
	35	-31.56	0.69	-0.01
	40	-34.47	0.59	0.02
Sharek	25	-26.54	1.22	-0.04
	30	-30.45	0.98	-0.03
	35	-34.68	0.80	-0.05
	40	-37.74	0.66	0.00
Average	25	-23.31	1.04	-0.02
	30	-27.47	0.84	0.00
	35	-32.73	0.70	-0.02
	40	-34.71	0.60	0.01

Table 3.6: The obtained results with respect to the HTM encoder using the IP prediction structure

Sequences	QP	Used Views					
		1-2-3			1-3-6		
		$\Delta Time$	$\Delta Bitrate$	$\Delta PSNR$	$\Delta Time$	$\Delta Bitrate$	$\Delta PSNR$
Champagne	25	-33.07	2.57	-0.11	-37.24	2,94	-0.13
	30	-36.95	2.27	-0.09	-41.17	2.62	-0.07
	35	-41.29	1.91	0.02	-45.29	2.28	-0.01
	40	-44.42	1.69	0.05	-48.09	2.09	0.03
Panatomim	25	-35.41	2.82	-0.17	-39.72	3.25	-0.14
	30	-39.21	2.55	-0.05	-43.37	2.93	-0.07
	35	-43.52	2.18	0.02	-47.41	2.51	-0.04
	40	-46.13	1.87	-0.01	-50.15	2.32	-0.01
Dog	25	-30.69	2,53	-0.13	-34.81	2,82	-0.12
	30	-34.38	2,11	-0.04	-38.25	2,43	-0.07
	35	-38.47	1,86	0.01	-42.33	2,25	-0.03
	40	-41.29	1,62	0.05	-44,92	1,96	0.02
Balloons	25	-26.59	2.08	-0.07	-30.21	2.38	-0.10
	30	-30.60	1.85	-0.05	-34.26	2.19	-0.08
	35	-35.17	1.59	0.02	-38.48	1.91	-0.05
	40	-38.09	1.35	0.05	-41.35	1.73	0.02
Kendo	25	-29.31	2.23	-0.06	-33.09	2.49	-0.08
	30	-33.64	1.98	-0.04	-37.35	2.33	-0.04
	35	-37.81	1.78	0.03	-41.28	2.16	-0.04
	40	-40.59	1.48	0.04	-44.02	1.88	-0.02
Sharek	25	-32.78	2.39	-0.07	-36.95	2.89	-0.09
	30	-36.58	2.01	-0.07	-40.71	2.36	-0.05
	35	-40.80	1.77	-0.04	-44.83	2.14	0.02
	40	-43.91	1.50	0.03	-47.68	1.87	0.04
Average	25	-31,31	2,44	-0,10	-35,34	2,79	-0,11
	30	-35,23	2,13	-0,06	-39,18	2,48	-0,06
	35	-39,51	1,85	0,01	-43,27	2,21	-0,02
	40	-42,40	1,58	0,03	-46,03	1,97	0,01

Table 3.7: The obtained results with respect to the HTM encoder using the PIP prediction structure.

Sequences	QP	Used Views					
		1-2-3			1-3-6		
		$\Delta Time$	$\Delta Bitrate$	$\Delta PSNR$	$\Delta Time$		$\Delta PSNR$
Champagne	25	-38,18	3,11	-0.10	-43,53	3,49	-0.13
	30	-41,83	2,69	-0.04	-47,24	3,08	-0.10
	35	-45,73	2,44	0.01	-50,91	2,78	-0.05
	40	-48,46	2.05	0.02	-53,52	2,29	0.01
Panatomim	25	-40,61	3,56	-0.13	-46,12	3,89	-0.10
	30	-43,99	3,25	-0.04	-49,56	3,56	-0.07
	35	-47,92	2,83	-0.04	-53,37	3,17	-0.05
	40	-50,81	2,34	0.03	-55,71	2,57	0.03
Dog	25	-35,81	2,89	-0.11	-40,88	3,22	-0.09
	30	-38,62	2,56	-0.04	-44,09	2,83	-0.06
	35	-42,84	2,31	-0.03	-47,80	2,65	-0.04
	40	-45,26	2.02	0.02	-50,21	2,28	0.04
Balloons	25	-31,22	2,39	-0.08	-35,55	2,72	-0.12
	30	-34,70	2.12	-0.07	-39,31	2,53	-0.07
	35	-38,85	1,91	-0.04	-43,47	2,20	-0.04
	40	-41,85	1,65	0.00	-46,02	1.92	0.01
Kendo	25	-34,12	2,53	-0.06	-38,54	2,81	-0.08
	30	-37,76	2.19	-0.01	-42,74	2,49	-0.05
	35	-41,71	1,81	-0.02	-46,39	2,10	-0.01
	40	-44,53	1,64	0.02	-47,91	1,87	0.00
Sharek	25	-37,94	2,83	-0.07	-43,24	3,12	-0.13
	30	-41,11	2,52	-0.05	-46,75	2,82	-0.10
	35	-45,23	2.19	-0.01	-50,57	2,49	-0.06
	40	-48,15	1,87	0.03	-54,16	2,11	0.01
Average	25	-36,31	2,88	-0,09	-41,31	3,21	-0,11
	30	-39,67	2,55	-0,04	-44,95	2,88	-0,07
	35	-43,71	2,25	-0,02	-48,75	2,56	-0,04
	40	-46,51	1,93	0,02	-51,25	2,17	0,02

Table 3.8: The obtained results with respect to the HTM encoder using the IBP prediction structure

Chapter 4

Error robustness evaluation in MV-HEVC coding standard

4.1 Introduction

In multi-view video coding, using a large GOP size can efficiently improve MVV compression and enable transmission of higher video content for a given bitrates. However, when transmitting MVV data, particularly in the presence of errors such as random bit error, burst error, or packet loss error, using a very long GOP can have adverse effects that propagate error spatially, temporally, and in inter-view directions [83]. As a result, even a single lost or corrupted bit in the MVV bitstream can propagate to adjacent views and subsequent frames, making the bitstream challenging, if not impossible, to decode. This situation can lead to a loss of encoder-decoder synchronization, affecting all slices that follow. For instance, in the MV-HEVC standard, the MVV codec takes multiple synchronized bitstreams captured from different cameras as input and generates a single bitstream for storage or transmission. However, when transmitting compressed MVV data over noisy channels, the video decoder decoding the corrupted video bitstream can lose synchronization and leading to errors propagating to the start of the next GOP. This poses two main problems:

Firstly, there is weak error detection capability in enhancement layers or dependent views, causing the decoder to include a lot of incorrect data in the decoded

video before detecting an error.

Secondly, once the decoder detects an error, it would discard the rest of the bitstream in that frame.

As there are limited studies on error-sensitive encoded data in MV-HEVC, this research aims to analyze the sensitivity of the MV-HEVC bitstream under error-prone conditions. The evaluation examined the effects of different components in the MV-HEVC bitstream, including the Network Abstraction Layer (NAL) unit types for both Video Coding Layer (VCL) and Non-Video Coding Layer (Non-VCL) samples encoded, which were subjected to varying bit error and packet loss rates. This study aims to identify the most sensitive compressed MVV data that should be encoded in a higher protected channel, while encoding the less error-sensitive data in dependent views.

This chapter is structured to provide a comprehensive understanding of MVV errors transmission types and their impact on MV-HEVC decoding process. The first subsections provide an introduction to MVV errors transmission types, along with a discussion of MVV quality assessment methods. The following sections investigate the impact of Bit-Error Rates (BERs) and Packet Loss Rates (PLRs) on NAL units in MV-HEVC Bitstream, analyzing the effects of these transmission errors on MVV quality. These sections include an in-depth examination of the relationship between transmission errors and MV-HEVC decoding efficiency. Finally, the chapter concludes with a presentation and discussion of the results obtained from the conducted research, along with an analysis of their significance and implications for future research in this area.

4.2 MVV Errors Transmission Types

The main purpose of using NAL units in the MV-HEVC standard is to support various video transmission systems. However, in the event of transmission errors, compressed MVV bitstreams become more susceptible to these errors, due to saptio-temporal and inter-view dependencies in the MVV sequences.

In error-prone environments, such as video transmission or storage, error hap-

pening in the previous frames in a GOP may propagate to all the following frames and views within that GOP, resulting in a degradation of the overall video quality. Figure 4.1 showes an example of MVV error propagation effect.

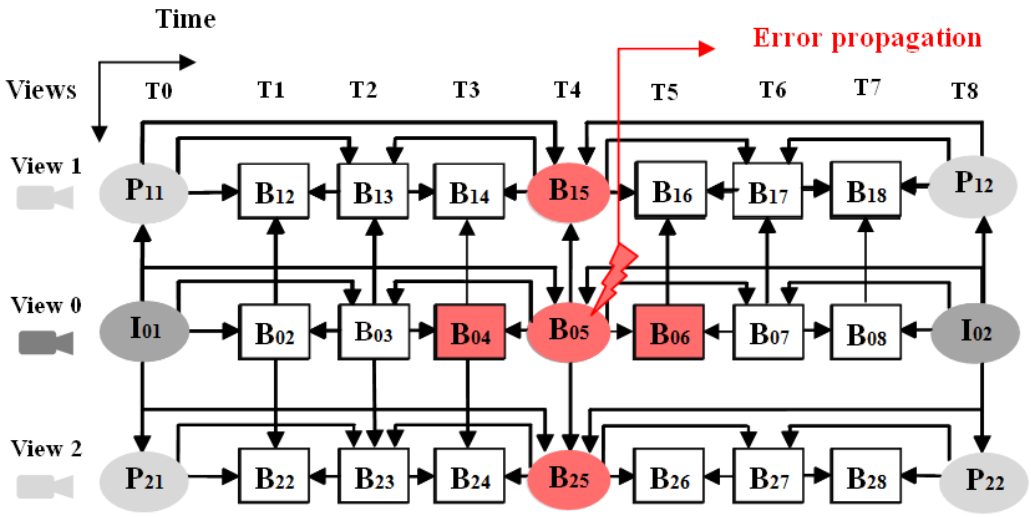

Figure 4.1: MVV error propagation effect.

The effect of displaying one of these frames that has been reconstructed from corrupted data can have a significant impact on visual perception in the decoded video. This degradation can be observed in various ways, such as blockiness, blurring, and artifacts. These visual impairments can not only affect the subjective quality of the video, but they can also impact the performance of various computer vision applications, such as object detection and recognition, which rely on the accuracy of the visual information.

MVV errors transmission can occur due to various factors, including deficiencies in physical channels. These errors can be classified into *random bit errors* and *erasure errors* [84].

4.2.1 Random Bit Error

Random bit errors can occur during transmission when there are issues with the physical channels, causing bit insertion, deletion, or reversal. These errors can also arise when one or more bits in the transmitted data are corrupted, flipped or lost during the transmission process, as shown in Figure 4.2.

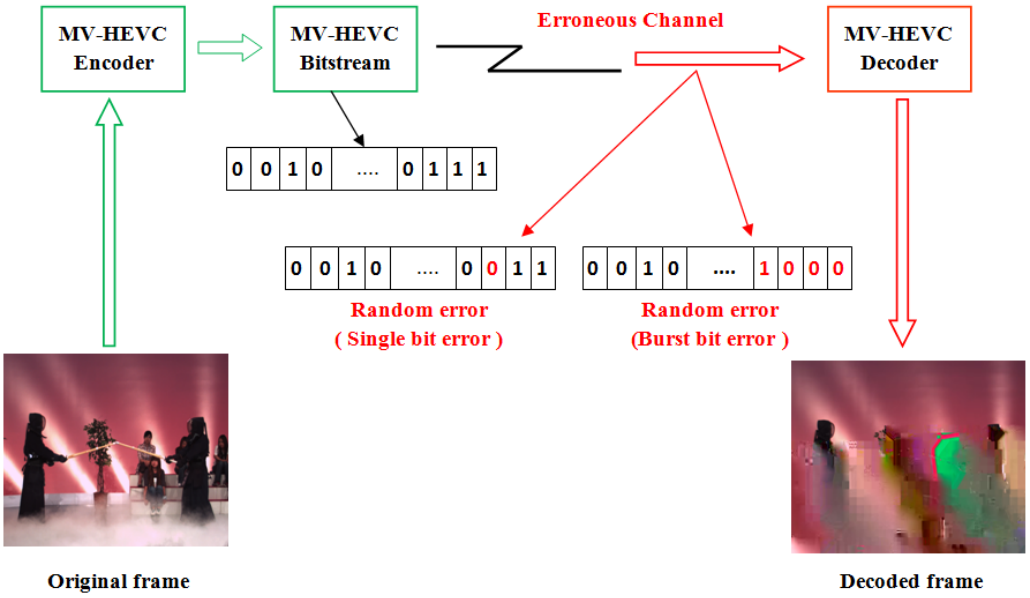

Figure 4.2: Random Error effect on MVV transmission.

The causes of these errors can be varied, including factors such as the signal-to-noise ratio, the quality of the transmission channel, the distance between the transmitter and receiver, electromagnetic interference, and the hardware used in the transmission process.

MVV transmission involves the transmission of large volumes of data at high speeds over long distances. As such, the probability of random bit errors occurring during transmission is high. The impact of random bit errors on the transmitted data can be significant, potentially leading to distortion, loss of critical information, or even complete failure of the transmission.

4.2.2 Erasure Error

Erasure errors occur when packets or group of packet that contain information about a specific view are lost or corrupted during MVV transmission , as shown in Figure 4.3

As result, the receiver knows which packets are missing and where they should be, but the missing packets cannot be recovered. Erasure errors are common in packet-based networks where packets can be lost due to network congestion or errors in the transmission medium. These errors can be more challenging to correct as they

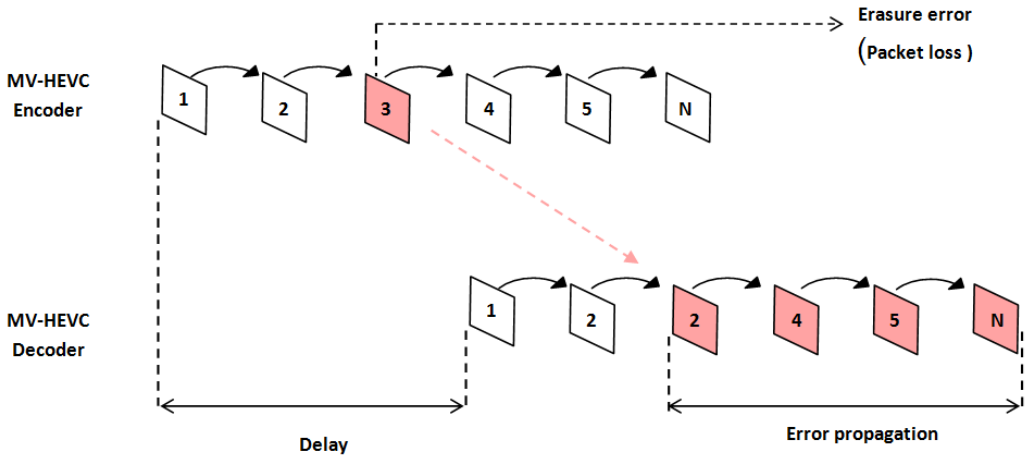

Figure 4.3: MVV Erasure Error propagation effect.

involve the loss of entire packets of data.

The main difference between erasure errors and random errors is that erasure errors result in the loss of specific packets of information, while random errors result in the corruption of some or all of the transmitted data. Understanding the different types of errors that can occur during MVV bitstream transmission is crucial in developing effective methods to correct and prevent such errors.

4.3 MVV Quality Assessment

The transmission of MVV signals involves a complex chain of processing components related to video codec and transmission. However, this chain of processing components can lead to a degradation of the video quality. This, in turn, can result in a low quality of experience for the end-users. Therefore, it is vital to evaluate the perceived visual quality to assess the coding efficiency of a particular video coding system. In this experiment study, two primary methods are used for evaluating video quality: objective and subjective evaluation methods.

4.3.1 Objective Assessment

Objective evaluation is a method used to measure the quality of video signals based on mathematical algorithms and models. Objective evaluation can be used to com-

pare different video codecs, compression techniques, or transmission systems, and it can also be used to evaluate the performance of different video processing algorithms.

Peak Signal-to-Noise Ratio (PSNR) is a widely used metric in objective evaluation for measuring the quality of video signals. PSNR is a mathematical formula that measures the difference between the original video signal and the compressed video signal. The PSNR value is expressed in decibels (dB) and it indicates the level of degradation in the video signal due to compression.

In MVVs, PSNR is used to evaluate the quality of the reconstructed views from multiple cameras. The PSNR is calculated by comparing each reconstructed view to the original view from the corresponding camera. The average PSNR value across all the views is used to determine the overall quality of the reconstructed multiview video.

In this experiment, the video signal is first compressed using a MV-HEVC encoder, and then the compressed signal is transmitted or stored with a certain level of error or loss. The resulting video signal is then compared to the original signal using PSNR, and the difference in quality is measured.

By varying the level of error or loss, it is possible to determine the error sensitivity of the system. If the video quality degrades significantly at low levels of error or loss, the system is considered to be highly error-sensitive. On the other hand, if the system can maintain video quality even at high levels of error or loss, it is considered to be less error-sensitive.

It is important to note that while PSNR is a widely used metric in objective evaluation, it has its limitations. PSNR does not always correlate well with subjective quality ratings, meaning that a high PSNR value does not necessarily indicate that the video quality is perceived as good by humans. Therefore, it is often necessary to complement objective evaluation with subjective evaluation, where human observers provide ratings of the video quality based on their perception.

4.3.2 Subjective Assessment

Subjective assessments are a fundamental evaluation method that involves human participants providing feedback on the quality of multimedia content such as MVV.

These ratings are utilized to determine the Mean Opinion Score (MOS), representing the average perceived quality of the video by the observers. Subjective evaluation methods are considered more accurate in assessing the true quality of experience for end-users since they incorporate human perception, which can identify factors that objective metrics may miss.

In the context of MVV, which involves capturing a scene from different angles and presenting the video with multiple viewpoints, subjective evaluation involves having a group of participants view the MVV and rate the quality of the video based on their subjective impressions. This can be done using various methods, such as rating the video on a scale of 1 to 5 or providing qualitative feedback about the video. Through analyzing this feedback, specific areas requiring improvement, such as the smoothness and clarity of transitions between different viewpoints, can be identified.

Subjective assessments are often used in conjunction with objective quality metrics, which are more quantitative measures of video quality. By combining both subjective and objective assessments, researchers and developers can gain a comprehensive understanding of the overall quality of the MVV and identify areas for improvement. Therefore, subjective evaluations are crucial in MVV research to provide an accurate and holistic evaluation of the end-users perception and satisfaction with the MVV.

4.4 Exploring the Impact of Bit-Error Rates on NAL Units in MV-HEVC Bitstream

The objective of this research study is to evaluate the effects of different Bit-Error Rates (BERs) on the two NAL units parts encoded in the MV-HEVC bitstream. In video coding, the NAL units contain compressed video data that are transmitted across a communication channel. Due to the sensitivity of the NAL units, any bit or burst errors can cause a significant impact on the overall video quality. Therefore, identifying the most sensitive compressed data is crucial for improving video transmission reliability.

The experimental work is conducted in error-prone environments with randomly placed bit or burst errors to obtain the desired NAL unit error. The BERs are varied to investigate their impact on the VCL NAL and non-VCL NAL units. The VCL NAL unit contains the video coding views, which is responsible for compressing the video frames. On the other hand, the non-VCL NAL unit includes the metadata and control information necessary for decoding the compressed video frames .

To identify the most sensitive compressed data in MV-HEVC bitstream, the study focuses on the NAL VCL and non-VCL NAL units. The overall block diagram of the evaluation process is shown in Figure 4.4.

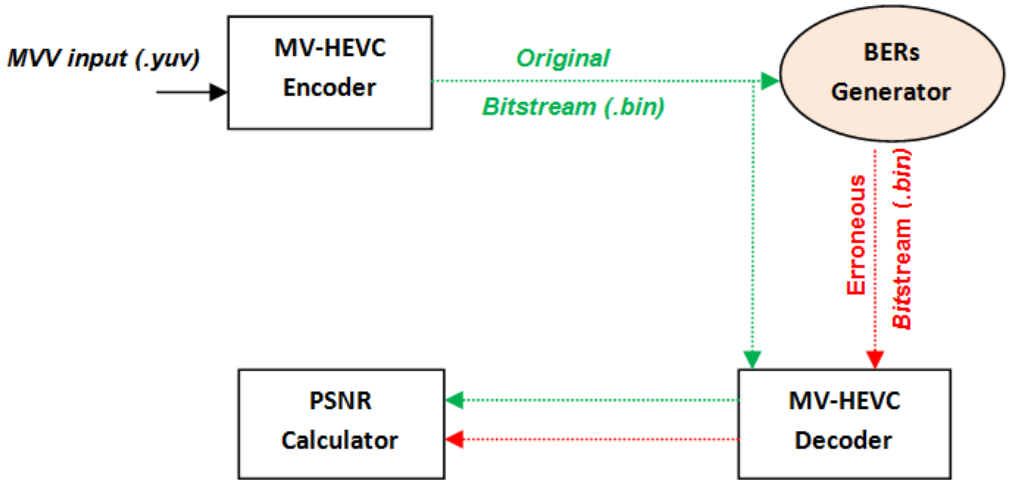

Figure 4.4: MV-HEVC Random Error Sensitivity Evaluation Diagram.

In our experimental research, we began by encoding MVVs in YUV format using the MV-HEVC HTM 16.6 benchmark software, which produced a binary format MV-HEVC bitstream. We then utilized our own BER generation algorithm to inject different levels of BERs into the encoded bitstream. We evaluated both the original and erroneous bitstreams after the MV-HEVC decoding stage, which included two views: the base view and the dependent view.

To ensure the highest possible quality results, we chose a random-access encoding configuration instead of using a low pass encoding configuration. Random-access encoding is a technique used in video compression that allows the decoder to access

any frame in the video sequence independently, without having to decode previous frames in the sequence. This results in higher quality video with less compression artifacts.

For this study, two video sequences are selected which are Kendo and Sharek, their details and encoding configuration settings are reported in Table 4.1 and Table 4.2, respectively. By using random-access encoding, we were able to achieve high quality results for both the original and erroneous bitstreams, which enabled us to accurately evaluate the impact of bit errors on the decoded video.

Sequences	Resolution	Frames	Frame rates (fps)	Cameras spacing(cm)
Kendo	1024x768	300	30	5
Sharek	920x1088	300	30	Stereo

Table 4.1: HD-MVV sequences

Setting	Value
Prediction Structure	IP
GOP length	8
Intra period	24
Profile	Main
Intra mode	1^{st} frame in each GOP
BERs	(0,2,4,6,8) $\times$ 10^{-4}

Table 4.2: MV-HEVC Encoder configurations

4.4.1 Experiment Results: Objective Evaluation

In this experiments, the objective comparisons results between the original and erroneous bitstreams of the MV-HEVC codec are obtained by using the objective

PSNR calculations. The obtained results are shown in Figure 4.5, and Figure 4.6.

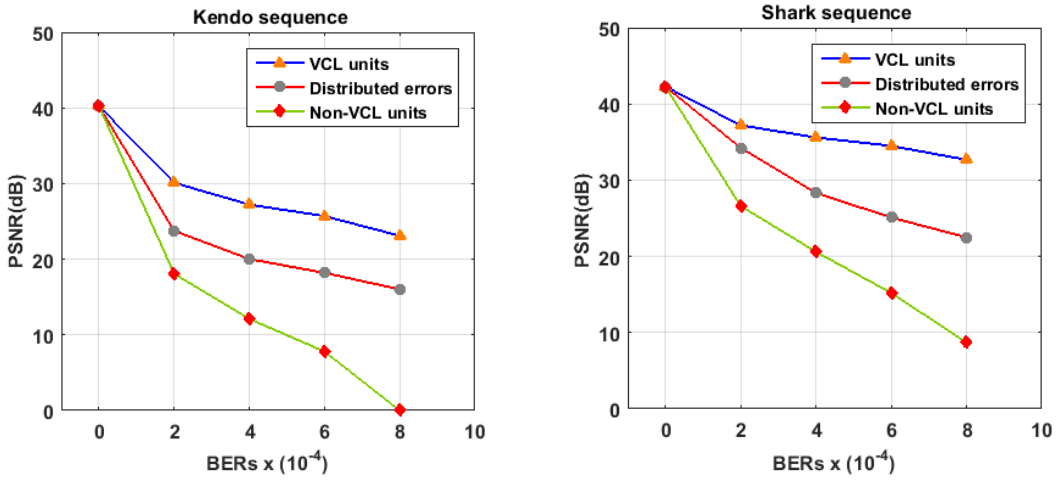

Figure 4.5: Encoding error sensitivity for non-VCL NAL and VCL-NAL units at various BERs.

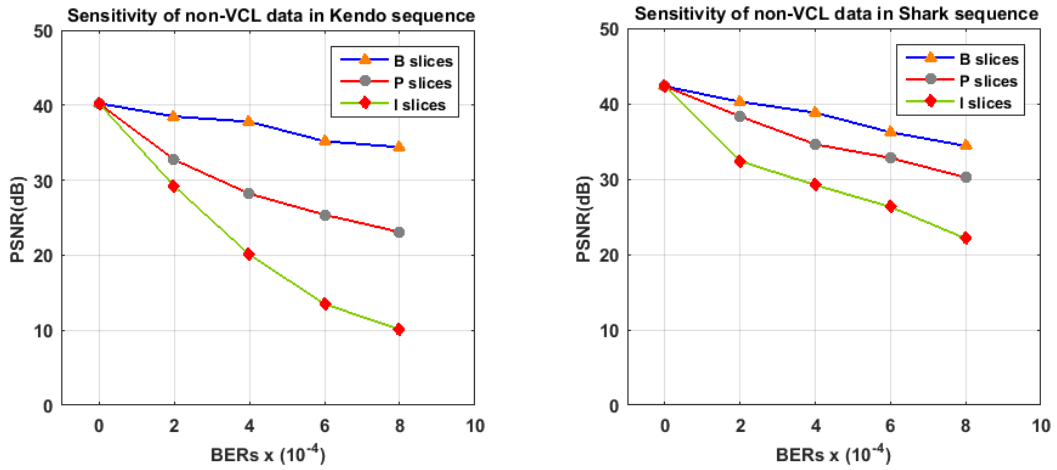

Figure 4.6: Encoding error sensitivity for Non-VCL NAL With different BERs.

Analysis results show that all parts of the NAL units suffer from a steady decrease in PSNR as the BERs increase, whether for Kendo or Shark video sequences. What is striking in Figure 4.5 is the Non-VCL NAL data that suffer from a high degradation in video quality. For instance, the PSNR of Non-VCL NAL units at the highest BERs is reduced by 18 dB for Kendo and 12 dB for Shark compared to evenly distributed BER injection. At the same condition, the quality of VCL NAL units is reduced by 5 dB for Kendo and 7 dB for Shark when injecting BERs into all MVV encoded data.

This means that the most sensitive part of the MV-HEVC bitstream in error-prone conditions is the Non-VCL data. This is because Non-VCL NAL information shared for multiple views and slices has an effect on the quality of the reconstruction. On the other side, a VCL-NAL data contains a representation of video samples which only affects the NAL payload depending on the slice level I, P and B, as shown in Figure 4.6. In which, when various BERs are injected only in the I slice, the quality of the P and B slices are also affected by this modification. Moreover, we can clearly seen that the highest quality degradation observed is I-slice PSNR, which is much smaller than the other slices P and B. This explains the relationship between the MV-HEVC interview prediction and the sensitivity of VCL NAL data in error-prone environment. In which, if the spatial pixels are affected and compensated in inter or inter-view modes, the quality fluctuation between the intra frames and their adjacent frames will appear, which lead to incorrect pixels reconstruction and causes error propagation in all dependent views and slices (P or B).

In addition to VCL and non-VCL NAL units sensitivity in each sequence, the results also show that the degradation of the PSNR is more noticeable in the Kendo video compared to the Sharek video. This implies that the content of the video plays an essential role in determining the degree of PSNR degradation. In particular, videos that have high resolutions, complex scenes, fast motion, or frequent changes in camera viewpoints tend to have a more significant reduction in PSNR than videos with simpler scenes and slower motion. Therefore, it is likely that the Kendo sequence has more complicated scenes and faster motion than the Sharek video. The PSNR is a widely used metric to evaluate video quality, and its reduction can be caused by various factors such as compression, noise, or artifacts. In this case, the difference in content complexity between the two videos is a significant factor that affects the PSNR values. To find out which part is the most sensitive among the MV-HEVC VCL data, more visual results are discussed in the next section.

4.4.2 Experiment Results: Subjective Evaluation

In this section, we will present the results of our study on subjective quality assessments, which were conducted by comparing slices of video sequences. The Kendo

sequence was used for this purpose, and the slices were extracted from the base view and the enhancement view in the form of I-slices, P-slices, and B-slices. The purpose of this study was to compare the quality of the original slices with that of the decoded slices that contained errors. To achieve this, we used the HTM.16.6 reference software.

Figure 4.7 presents the comparison results that were achieved in our study. The right column of the figure shows the original slices, while the left column displays the slices that were decoded erroneously using the HTM.16.6 software. We compared the erroneous decoded slices with the original ones to assess the level of quality degradation that occurred during the decoding process.

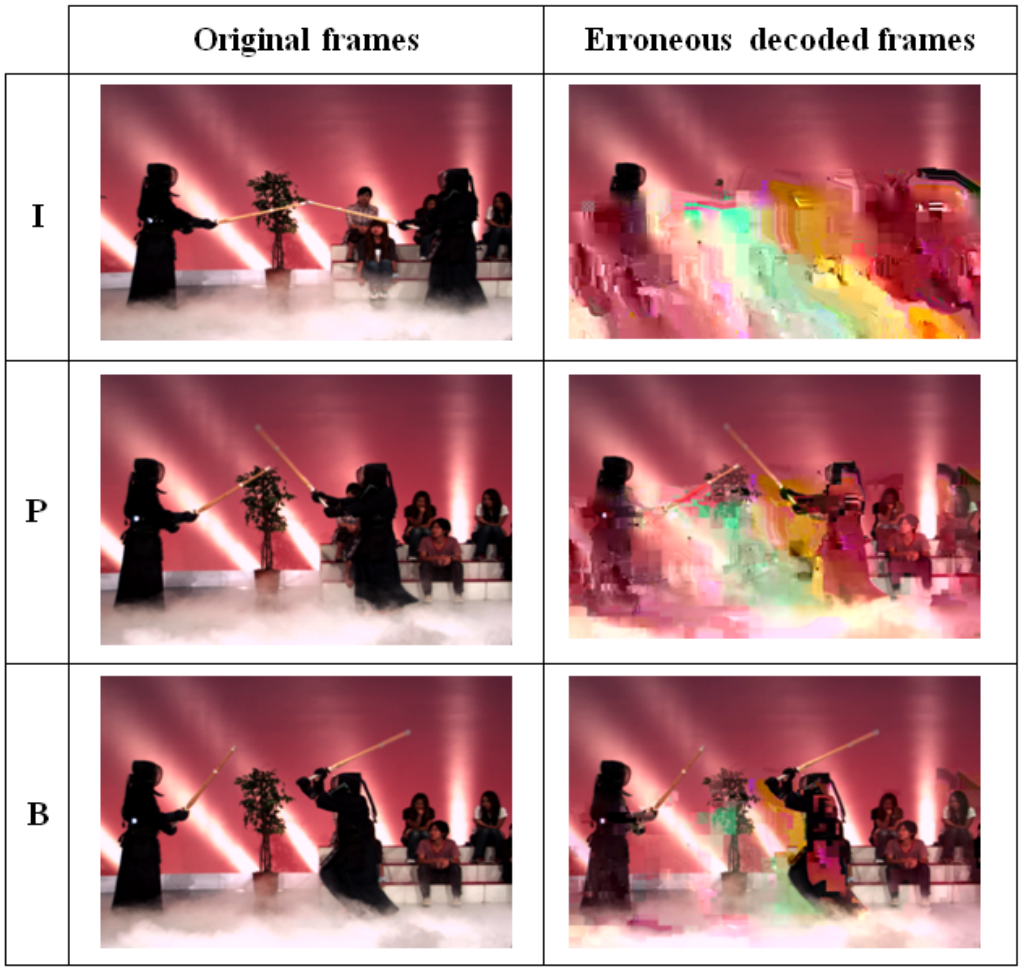

Figure 4.7: Perceived visual quality effect on Kendo video slices (I, P and B) at the BER ($8x10^{-4}$).

The results depicted in Fig.7 highlight that the reconstructed I-slice is the most vulnerable to degradation, particularly in the active areas which contains very little discernible information when compared to the original. This is due to the fact that I-slices are independent frames that do not depend on other frames in the video sequence for decoding. As a result, any loss or corruption of data in an I-slice leads to a significant loss of visual quality.

On the other hand, the P-frame, which is a predictive frame, offers better resistance to degradation than I-slices. Although P-frames use previously decoded frames for decoding, they can still exhibit some visual artifacts or broken parts in the video. This is because P-frames only store the differences between the current frame and the reference frame, resulting in some loss of information.

Lastly, the B-frame, which is a bi-directional predictive frame, offers the best quality among the three frame types analyzed in our research. B-frames use both past and future frames for decoding, resulting in better visual quality than I and P frames. However, B-frames are more computationally intensive to encode and decode, which can lead to higher complexity and longer processing times.

Furthermore, our research found that error propagation in the corrupted I-slice is persistent, leading to the same level of visual quality degradation in the subsequent views (P and B). This means that the inaccuracies in motion and disparity estimations in an I-slice can lead to cascading effect on subsequent P and B frames, resulting in a consistent degradation of visual quality across multiple views.

The reason behind this degradation lies in the fact that motion and disparity compensations are the main contributors to error propagation in MVVs. Motion compensation involves estimating the motion between frames, while disparity compensation involves estimating the difference in perspective between the views. Any errors in these estimations can lead to errors in subsequent frames, making the relationship between VCL NAL in the compressed MV-HEVC bitstream more complex.

4.5 Exploring the Impact of Packet Loss Rates on NAL Units in MV-HEVC Bitstream

After evaluating the effects of random errors on the MV-HEVC bitstream, we turned our attention to investigating the potential impact of packet loss on MV-HEVC NAL units (both VCL and non-VCL). To explore this, we introduced varying levels of packet loss into the MV-HEVC bitstream, as illustrated in the overall diagram depicted in Figure 4.8

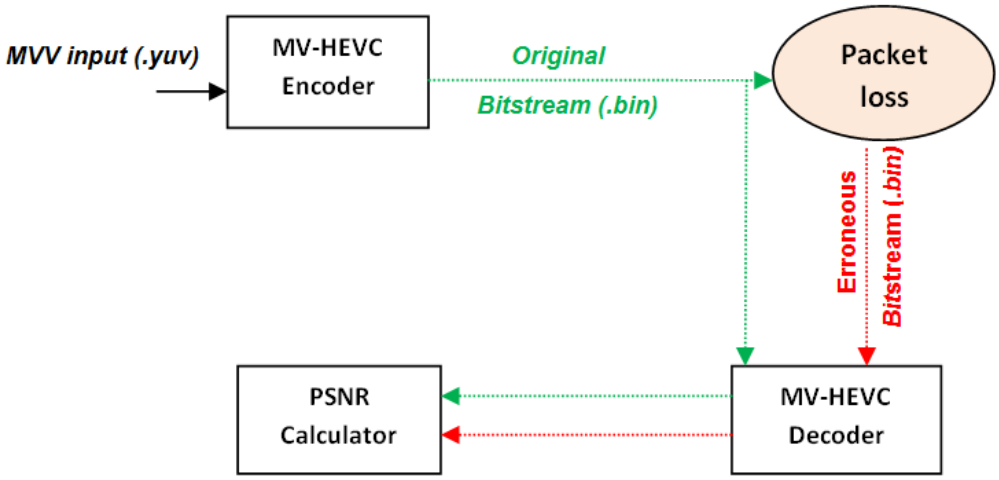

Figure 4.8: MV-HEVC Packet Loss Sensitivity Evaluation Diagram.

This study aims to investigate the impact of Packet Loss Rates (PLR) on non-VCL NAL units, such as VPS, SPS, and PPS slices. The study further examines the effects of packet loss on VCL NAL units, specifically the different frame types, including I, P, and B frames, to determine which frame type is the most sensitive to packet loss.

To ensure consistency, the study maintains the same experimental conditions and configurations as presented in Table 4.1 and Table 4.2, with the exception of varying packet loss rates (1%, 2%, and 4%) . Objective and subjective evaluations are conducted to further explore the effects of packet loss on the different NAL units.

4.5.1 Experimental Results: Assessing Packet Loss Effects on MV-HEVC non-VCL NAL Units

During the experiment, when varying packet loss rates of 1% ,2% , and 4% were applied to non-VCL NAL units (VPS, SPS, or PPS) for both Kendo and Sharek sequences, we observed that the MV-HEVC decoder stopped decoding.

This occurred because the non-VCL NAL units contain critical information necessary for decoding, and any loss of this information can result in the decoder being unable to continue. As a result, the decoder halts decoding, and the video playback is interrupted. These findings highlight the importance of maintaining the integrity of non-VCL NAL units in the MV-HEVC bitstream to ensure uninterrupted decoding and playback of video content.

4.5.2 Experimental Results: Assessing Packet Loss Effects on MV-HEVC VCL NAL Units

This section describes the experimental study conducted to examine the impact of varying PLR on the quality of MV-HEVC VCL NAL units. The primary objective of the study was to analyze the effect of packet loss on the quality of different types of frames (I, P, and B) and to investigate how the rate of packet loss influenced the overall video quality. We also analyzed the sensitivity of different frame types to the packet loss rates and identified which frame type was the most sensitive to packet loss.

4.5.2.1 Packet Loss Effect on I Frame

If a packet loss occurs in an I-frame, the entire frame will be lost and the MV-HEVC decoder was unable to properly decode other frames or views.

More precisely, the I frame is crucial for proper decoding and synchronization, losing even a single I-frame can negatively impact the quality of subsequent frames and adjacent views. This is due to the fact that I-frames are independently encoded and do not depend on other frames for reconstruction. Consequently, the decoder will not have access to the necessary information to reconstruct that frame,

and subsequent frames may not be decoded correctly, resulting in visible artifacts, distortion, or even complete video destruction.

4.5.2.2 Packet Loss Effect on P Frame

Table 4.3 shows the PSNR results of PLR occurrence in P frame after the MV-HEVC decoding process.

Sequences	PLR(%)	Frames					
		Base View (V1)			Enhancement View (V2)		
		I	B8	B4	P	B8	B4
		PSNR (dB)					
Kendo	0	45.61	44.40	44.15	44.30	43.60	43.04
	1	45.61	44.40	44.15	24.43	24.48	24.61
	2	45.61	44.40	44.15	18.09	16.21	17.15
	4	45.61	44.40	44.15	17	17.2	17.24
Sharek	0	44.77	42.19	42.35	44.13	41.63	41.84
	1	44.77	42.19	42.35	30.36	28.38	29.61
	2	44.77	42.19	42.35	27.01	26.11	25.67
	4	44.77	42.19	42.35	21.30	20.65	21.04

Table 4.3: Comparison of PSNR Results for P-Frames with Different Packet Loss Rates (PLR).

The results indicated that the PSNR of the base view frames remained constant for all PLRs. However, for the enhancement view, the PSNR decreased as the PLR increased for both kendo and sharek sequences.

This highlights that packet loss in P frame can significantly impact the decoding process of the enhancement view and of subsequent frames within the same view, in which even missing packets in P frame at a low rate of 1% can reduce the accuracy of the MV-HEVC decoding process, resulting in lower quality video output.

Furthermore, it has observed that a PLR =4 % can cause decoding errors that negatively impact the overall accuracy of the entire view, as demonstrated by the

subjective evaluation results presented in Figure 4.9 for the Kendo sequence.

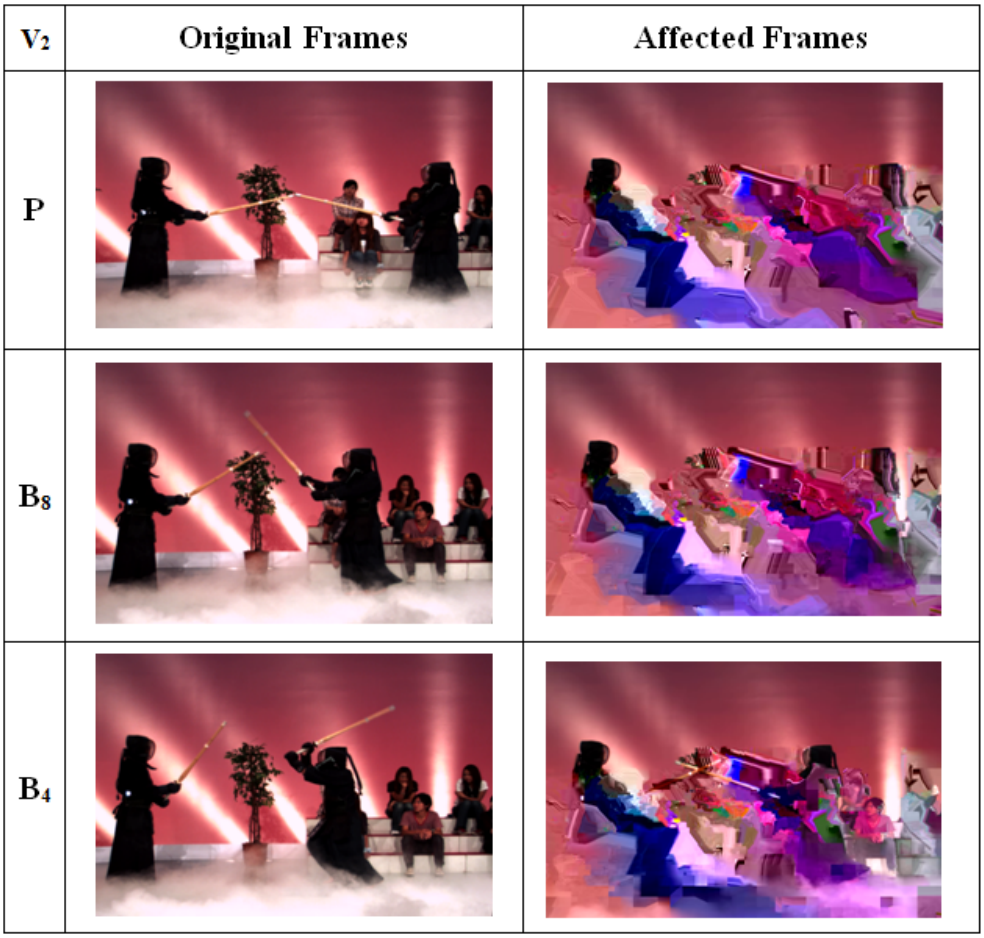

Figure 4.9: Packet loss effect on P frame at the PLR (4%) for Kendo sequence.

In addition, the impact on video quality will depend on the type of prediction used. In P-frames, only the differences between the current frame and the previously decoded frame are coded. If the prediction used is inter-prediction or inter-view prediction, then the loss of a P-frame will result in errors propagating to subsequent frames. If the prediction used is intra-prediction , then the impact of packet loss will be less severe, as the lost frame can be reconstructed from neighboring frames.

4.5.2.3 Packet Loss Effect on B Frames

When packet loss happens within a B-frame, the resulting impact on MVV quality can vary depending on factors such as the type of prediction employed and the position of the lost frame within the GOP hierarchy.

Analysis of the PSNR results presented in Table 4.4 indicates that when a packet loss occurs in the B8 frame of both views (V1 and V2), it has a discernible effect on the quality of the B4 frame for both views. This is because the prediction of B4 relies on the information provided by B8. However, the loss of a packet does not affect the quality of either the I-frame (V1) or P-frame (V2) since these frames have independent predictions.

Sequences	PLR(%)	Frames					
		Base View (V1)			Enhancement View (V2)		
		I	B8	B4	P	B8	B4
		PSNR (dB)					
Kendo	0	45.61	44.40	44.15	44.30	43.60	42.1
	1	45.61	44.40	44.15	44.30	35.20	33.17
	2	45.61	44.40	44.15	44.30	32.03	30.8
	4	45.61	44.40	44.15	44.30	29.6	29.76
Sharek	0	44.77	42.19	42.35	44.13	41.63	41.84
	1	44.77	42.19	42.35	44.13	40.38	40.20
	2	44.77	42.19	42.35	44.13	38.9	39.17
	4	44.77	42.19	42.35	44.13	34.8	35

Table 4.4: Comparison of PSNR Results for B-Frames with Different Packet Loss Rates (PLR).

On overall, B-frames are predicted using both previous and future frames, so if a B-frame is lost, it may affect subsequent frames that use it as a reference. However, if the lost B-frame is not used as a reference for subsequent frames, then the impact of the loss will be minimal, as shown in Figure 4.10 and Figure 4.11, which demonstrate the comparative effects of lost B-frames on the overall video quality. By conducting a subjective evaluation, it becomes possible to measure and understand the perceptual implications of such losses. This assessment helps in determining the error sensitivity of the MV-HEVC coding standard and assists in developing strategies to mitigate the potential negative consequences of lost B-frames on subsequent frame references.

V_1	Original Frames	Affected Frames
I		
B_8		
B_4		

Figure 4.10: Packet loss effect on B frame at the PLR (4%) for Kendo sequence (The effect on Base View).

Figure 4.11: Packet loss effect on B frame at the PLR (4%) for Kendo sequence (The effect on Enhancement View).

4.6 Conclusion

In this chapter, we have evaluated the sensitivity of the MV-HEVC bitstream in an error-prone environment, an area that has received little attention in previous studies. Our primary objective was to investigate the impact of two NAL units, VCL NAL and Non-VCL NAL, on the decoded perceived visual quality of MV-HEVC. We applied two types of errors, random bit errors, and erasure errors (packet loss), to perform our analysis.

Our analysis indicate that the effects of both random bit errors and packet loss on MV-HEVC are severe, with damaged video slices significantly affecting overall

multi-view video quality. Nonetheless, we also discovered that packet loss on Non-VCL and VCL NAL units of MV-HEVC bitstream has a more severe impact than random bit errors. This is due to the predictive coding structure and inter-view dependencies of MV-HEVC.

This study further shows that the non-VCL NAL and non-VCL active area are the most sensitive encoded data in the MV-HEVC bitstream. Therefore, these areas must be protected to minimize error propagation in multi-view videos.

Our research is a significant contribution to the literature as it highlights the importance of error analysis in the MV-HEVC standard and the need for measures to enhance the quality of multi-view videos in error-prone environments.

As future work, the analysis of the M-HEVC error sensitivity study can be used to optimize the encoding and transmission of compressed video data. Specifically, the study can help to identify the most error-sensitive parts of the data, which can then be encoded using more robust error correction techniques and transmitted over higher protected channels. On the other hand, less error-sensitive parts of the data can be encoded using less robust techniques and transmitted over less protected channels.

Moreover, the study suggests that encoding the less error-sensitive parts of the video data in enhancement views can be an effective way to reduce the overall bit rate while maintaining video quality. This approach can be further explored and evaluated for its practical applications in various video coding scenarios. Overall, the M-HEVC error sensitivity study provides valuable insights for improving the efficiency and reliability of compressed video transmission and storage, and can be further developed to enhance the performance of video coding systems.

General Conclusion

The development of Multi-View Videos (MVVs) has made it possible to create realistic scenes from different angles, making them ideal for use in various 3D applications. However, the high-quality videos and virtual navigation around the scenes have brought new challenges to video compression technologies. To address this, the H.264/AVC standard has been extended to support multi-view video coding with the introduction of MVC, which has significantly improved compression efficiency expressed as a trade-off between bit rate and video quality

The introduction of Ultra High Definitions (UHD) technologies has slowed the success of AVC solutions, but the High Efficiency Video Coding (HEVC) standard has been developed with advanced coding structures and new parallel processing features. HEVC has achieved twice the compression level of AVC while maintaining the same quality, making it suitable for all existing applications and a wide range of video resolutions. HEVC has been extended for 3D video coding with the Multi-view extension (MV-HEVC) and 3D video extension (3D-HEVC), which enable new coding tools to improve the coding of dependent video views and associated depth data for each view of the MVV needed to generate 3D contents. However, the computational complexity required for HEVC is high, making real-time implementation challenging.

This thesis focuses on reducing the complexity of MV-HEVC while ensuring a satisfactory trade-off between bit rate and video quality. We based this complexity reduction on exploiting the similarity between different views of the MVV and the shift resulting from camera spacing. Furthermore, we analyzed the sensitivity of the MV-HEVC bitstream under error-prone conditions. In summary, our research work resulted in the following:

- We proposed the previous Earlier Disparity Estimation (EDE) algorithm to reduce the computation time of the disparity estimation process. The EDE uses similarity analysis between views to evaluate the rate of predicted units (UP) only from shifted parts. Our research work aims to visually adjust dependent views by eliminating shifts between views. These parts contain no information in the base view reference frames during inter-view prediction.

- The encoding process takes more computation time for a large search window. Therefore, we reduced the DE search window based on the similarities between views and levels of coding unit partitioning. This approach reduces the number of search points in a limited window.

- We improved motion estimation (ME) for dependent views by introducing Early Decision for Coding Unit Splitting (EDCS).

- We also explored the sensitivity of the MV-HEVC bitstream under error-prone conditions. This study aims to evaluate the effects of different coded components in the MV-HEVC bitstream, such as types of network abstraction layer (NAL) units, which include both the video coding layer (VCL) and the non-video coding layer (non-VCL). We subjected coded samples to various error rates to identify the most sensitive compressed data that needs to be encoded in better-protected channels, and the less error-sensitive encoded data that can be encoded in dependent views.

The evaluation of the approach aimed at reducing the complexity of MV-HEVC is carried out through several scenarios and using three prediction structures: IP, PIP and IBP. We focused on this evaluation on the distance between the basic view and the dependent views, especially for the IP prediction structure. We have used distance as the primary factor to evaluate our approach, due to its significant influence on inter-view disparity. In our experiments, we eliminated disparity for each run to analyze the trade-off between encoding time savings and video quality/bit rate. We compared our results with those obtained from the HTM reference software. Our experimental results indicate that our proposed method provides promising encoding

time savings by utilizing between-view visual adjustment, with negligible effect on RD performance. Additionally, our experiments demonstrated that $\Delta time$ increases as the inter-camera distance, represented by inter-view disparity, increases.

The proposed method is also evaluated using three-view prediction structures (one base view and two dependent views). We thus used in our experimentation the type of prediction structures: PIP and IBP. As the IP prediction structure, we eliminated while encoding the offset parts from the three views of the PIP structure. The parts removed depend on the position of each view in the PIP prediction structure. The experimental results show that the gain in encoding time is greater than for the IP prediction structure. This is justified by the increase in the number of parts eliminated during encoding, which also affects RD performance. The best gain in encoding time exceeded 46% for the "Panatomim" video using a QP equal to 40.

The impact of the proposed approach for the bidirectional prediction of B views is also analyzed using the IBP prediction structure. The mechanism for eliminating the offset parts is a bit different than the PIP prediction structure, in particular for view B. This is due to the use of two reference views by this type B view. This constraint has increased the gain in encoding time along with the bit rate with respect to the PIP structure.

Due to the scarcity of error analysis for the MV-HEVC standard, we evaluated in this thesis the sensitivity of the MV-HEVC bitstream in an error-prone environment. We focused in this study mainly on the two NAL units which are VCL NAL and Non VCL NAL coded in MV-HEVC while analyzing the decoded perceived visual quality of decoded MV-HEVC video. To achieve this, we introduced two types of errors: random bit errors and erasure errors (packet loss), and analyzed their effects.

The results of our analysis indicate that both random bit errors and packet loss have a considerable impact on the quality of multi-view video. However, we also discovered that packet loss on Non-VCL NAL and VCL NAL units of the MV-HEVC bitstream has a more severe effect compared to random bit errors. This phenomenon can be attributed to the predictive coding structure and inter-view dependencies inherent in MV-HEVC.

Furthermore, our study demonstrates that the non-VCL NAL and non-VCL active area are the most sensitive encoded data components in the MV-HEVC bitstream. Specifically, we have found that error propagation within the corrupted I slice exhibits a persistent nature, resulting in consistent degradation of visual quality across subsequent frames (P and B).

This behavior can be attributed to the prominent role played by motion and disparity compensations in the propagation of errors within the MVVs. Motion compensation entails the estimation of motion between frames, while disparity compensation involves estimating the perspective differences among the views. In which, any inaccuracies in these estimation processes can give rise to errors that propagate through subsequent frames. Consequently, the relationship between the VCL NAL within the compressed MV-HEVC bitstream becomes considerably intricate.

Moreover, this analysis study not only highlighted the sensitivity of both VCL and non-VCL NAL units in each video sequence but also shed light on the relationship between content characteristics and the degradation of PSNR in video sequences. The analysis demonstrated that videos with high resolutions, complex scenes, fast motion, or frequent changes in camera viewpoints exhibit a more substantial reduction in PSNR compared to videos with simpler scenes and slower motion. This implies that content complexity plays a crucial role in determining the extent of PSNR degradation.

4.7 Perspective

The proposed techniques in this thesis have improved the performance of MV-HEVC, making it more efficient and robust. However, there is still room for improvement. Future research can focus on developing more sophisticated algorithms to further reduce computation time, increase encoding speed, and improve motion estimation. Additionally, analyzing error transmission under more diverse conditions, such as different types of noise and transmission environments, can provide more insight into the performance of MV-HEVC under real-world scenarios.

Overall, this thesis has contributed significantly to the field of video compression

and error transmission analysis by proposing innovative techniques and evaluating the sensitivity of MV-HEVC under error-prone conditions. We hope that our work will inspire further research and development in this field, leading to more efficient and robust video compression techniques.

Bibliography

[1] D. Save and A.N. Mohari. Different video compression standards. *International Conference On Emanations in Modern Technology and Engineering*, 5:83–85, 2017. https://ijritcc.org/download/conferences/ICEMTE-2017.

[2] S.C Lim, D. Y. Kim, and J. Kang. Simplification on cross-component linear model in versatile video coding. *Electronics*, 9:83–85, 2020. https://doi.org/10.3390/electronics9111885.

[3] M. Claypool, K. Claypool, and F. Damaa. The effects of frame rate and resolution on users playing first person shooter games. *Proceedings. SPIE, Multimedia Computing and Networking*, 6071:1–11, 2006. https://doi.org/10.1117/12.648609.

[4] M. Abomhara, O.O. Khalifa, O. Zakaria, A.A. Zaidan, B.B. Zaidan, and A. Rame. Video compression techniques: An overview. *Journal of Applied Sciences*, 10:1834–1840, 2010. https://scialert.net/abstract/?doi=jas.2010.1834.1840.

[5] R. V. Bidwe, S. Mishra, Sh. Patil, K. Shaw, D. R. Vora, K. Kotecha, and B. Zope. Deep learning approaches for video compression: A bibliometric analysis. *Big Data and Cognitive Computing*, 6:1–40, 2022. https://doi.org/10.3390/bdcc6020044.

[6] B. Bross, J. Chen, J.R. Ohm, G.J. Sullivan, and Y.K. Wang. Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc). *Proceedings of the IEEE*, 109:1463–1493, 2021. https://doi.org/10.1109/JPROC.2020.3043399.

[7] J. Vanne. *Design and Implementation of Configurable Motion Estimation Architecture for Video Encoding.* PhD thesis, Tampereen teknillinen yliopisto, Tampere University of Technology, 2011. https://doi.org/10.3390/bdcc6020044.

[8] I. Park and D. W. Capson. Improved motion estimation time using a combination of dynamic reference frame selection and residue-based mode decision. *Signal, Image and Video Processing*, 6:25–39, 2012. https://link.springer.com/article/10.1007/s11760-010-0169-5.

[9] J. Lainema, F. Bossen, W. Han, J. Min, and K. Ugur. Intra coding of the hevc standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1792–1801, 2012. https://doi.org/10.1109/TCSVT.2012.2221525.

[10] T. Zhang, H. Chen, M.T. Sun, D. Zhao, and W. Gao. Hybrid angular intra/template matching prediction for hevc intra coding. In *Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2015. https://doi.org/10.1109/VCIP.2015.7457833.

[11] J.L Lin, Y.W. Chen, Y.P. Tsai, Y.W. Huang, and Sh. Lei. Motion vector coding techniques for hevc. In *13th International Workshop on Multimedia Signal Processing*, pages 1–6. IEEE, 2011. https://doi.org/10.1109/MMSP.2011.6093817.

[12] J. Sang, Q. Liu, and C.L Song. Robust video watermarking using a hybrid dct-dwt approach author links open overlay panel. *Journal of Electronic Science and Technology*, 18:1–20, 2020. https://doi.org/10.1016/j.jnlest.2020.100052.

[13] C. Lan, J. Xu, W. Zeng, G. Shi, and F. Wu. Variable block-sized signal-dependent transform for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 8:1920 – 1933, 2018. https://ieeexplore.ieee.org/document/7888910.

[14] M. Servais and G. Jager. Video compression using the three dimensional discrete cosine transform (3d-dct). In *Proceedings of the 1997 South African Symposium on Communications and Signal Processing. COMSIG '97*, pages 27–32. IEEE, 1997. https://doi.org/10.1109/COMSIG.1997.629976.

[15] H.Ch. Chang, L.G. Chen, M.Y. Hsu, and Y.Ch. Chang. Performance analysis and architecture evaluation of mpeg-4 video codec system. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 449–452. IEEE, 2000. https://doi.org/10.1109/ISCAS.2000.856361.

[16] T.C. Wang, Y.W. Huang, H.C. Fang, and L.G. Chen. Parallel 4/spl times/4 2d transform and inverse transform architecture for mpeg-4 avc/h.264. In *2003 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 800–803. IEEE, 2003. https://doi.org/10.1109/ISCAS.2003.1206095.

[17] G. J. Sullivan, J.R. Ohm, W.J Han, and T. Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1649 – 1668, 2012. https://doi.org/10.1109/TCSVT.2012.2221191.

[18] J.R Ohm, G.J. Sullivan, H. Schwarz, T.T. Keng, and T. Wiegand. Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc). *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1669 – 1684, 2012. https://doi.org/10.1109/TCSVT.2012.2221192.

[19] T. T. Alfaqheri. *Error Control Strategies in H.265/HEVC Video Transmission.* PhD thesis, TSchool of Engineering Design and Physical Sciences Brunel University London, 2011. https://bura.brunel.ac.uk/handle/2438/20149.

[20] Sujatha C. and K. U. Surekha. Video compression standards: A survey. *International Journal of Creative Research Thoughts (IJCRT)*, 6:84 – 88, 2018. https://doi.org/10.1109/TCSVT.2012.2221191.

[21] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi. Video coding with h.264/avc: tools, performance, and complexity. *IEEE Circuits and Systems Magazine*, 4:7 – 28, 2004. https://ieeexplore.ieee.org/abstract/document/1286980.

[22] X. Xu and Sh. Liu. Recent advances in video coding beyond the hevc standard. *APSIPA Transactions on Signal and Information Processing*, 8, 2019. https://doi.org/10.1017/ATSIP.2019.11.

[23] A.B. Dehkordi, M.A. Mahsa, T. Pourazad, and P. Nasiopoulos. Compression of high dynamic range video using the hevc and h.264/avc standards. In *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. IEEE, 2014.

[24] S. Kwon, A. Tamhankar, and K.R. Rao. Overview of h.264/mpeg-4 part 10. *Journal of Visual Communication and Image Representation*, 17:186–216, April 2006. https://doi.org/10.1016/j.jvcir.2005.05.010.

[25] D. Marpe, T. Wiegand, and G.J. Sullivan. The h.264/mpeg4 advanced video coding standard and its applications. *IEEE Communications Magazine*, 44:134 – 143, August 2006. https://doi.org/10.1109/MCOM.2006.1678121.

[26] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the h.264/avc standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17:1103 – 1120, 2007. https://ieeexplore.ieee.org/abstract/document/4317636.

[27] T.Y. Kuo and Ch. H. Chan. Fast variable block size motion estimation for h.264 using likelihood and correlation of motion field. *IEEE Transactions on Circuits and Systems for Video Technology*, 16:1185 – 1195, October 2006. https://https://doi.org/10.1109/TCSVT.2006.883512.

[28] P. Jayakrishnan and H. M. Kittur. 2-dimensional systolic architecture for h.264/avc variable block size motion estimation. In *International Conference on Circuits, Communication, Control and Computing*. IEEE, 2014.

[29] N.A. Khan, S. Masud, and A. Ahmad. A variable block size motion estimation algorithm for real-time h.264 video encoding. *Signal Processing: Image Communication*, 21:306–315, 2006. https://doi.org/10.1016/j.image.2005.11.004.

[30] R. Sjoberg, Y. Chen, A. Fujibayashi, M. M. Hannuksela, J. Samuelsson, and Th. K. Tan. Overview of hevc high-level syntax and reference picture management. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1858 – 1870, 2012. https://doi.org/10.1109/TCSVT.2012.2223052.

[31] K.E Psannis and Y. Ishibashi. Efficient error resilient algorithm for h.264/avc: mobility management in wireless video streaming. *Springer Telecommunication Systems*, 41:65 – 76, 2009. https://doi.org/10.1007/s11235-009-9151-3.

[32] M. Sima, Y. Zhou, and W. Zhang. An efficient architecture for adaptive deblocking filter of h.264/avc video coding. *IEEE Transactions on Consumer Electronics*, 50:292 – 296, 2004. https://ieeexplore.ieee.org/abstract/document/1277876.

[33] J. Li and K.N. Ngan. Adaptive partition size temporal error concealment for h.264 using weighted double-sided ebme minimization. *Signal Processing: Image Communication*, 23:451–462, 2008. https://www.sciencedirect.com/science/article/abs/pii/S0923596508000398.

[34] Kh.Q. Dinh, H. Shim, and B. Jeon. Deblocking filter for artifact reduction in distributed compressive video sensing. In *IEEE Transactions on Circuits and Systems for Video Technology*, pages 305 – 319. IEEE, 2014.

[35] Ch. Ram and S. Panwar. Performance comparison of high efficiency video coding (hevc) with h.264 avc. In *13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*. IEEE, 2018.

[36] G.J. Sullivan, J.M. Boyce, Y. Chen, J.R. Ohm, C. A. Segall, and A. Vetro. Standardized extensions of high efficiency video coding (hevc). *IEEE Journal of Selected Topics in Signal Processing*, 7:1001 – 1016, 2013. https://doi.org/10.1109/JSTSP.2013.2283657.

[37] I.K Kim, J. Min, T. Lee, W.J. Han, and J. Park. Block partitioning structure in the hevc standard. *IEEE Journal of Selected Topics in Signal Processing*, 22:1697 – 1706, 2012. https://ieeexplore.ieee.org/abstract/document/6324412.

[38] L. Zhao, X. Guo, Sh. Lei, S. Ma, and D. Zhao. Simplified amvp for high efficiency video coding. In *2012 Visual Communications and Image Processing*, pages 1 – 4. IEEE, 2012. https://doi.org/10.1109/VCIP.2012.6410747.

[39] S.H Tsang, W. Kuang, Y.L Chan, and W.Ch Siu. Decoder side merge mode and amvp in hevc screen content coding. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 260 – 264. IEEE, 2018. https://doi.org/10.1109/ICIP.2017.8296283.

[40] J. Young Lee. Novel skip motion estimation for efficient inter coding in hevc. *Multimedia Tools and Applications*, 80:4493 – 4505, 2021. https://doi.org/10.1007/s11042-020-09945-9.

[41] T. Zhang, T. H. Chen, M.T Sun, A. Saxena, D. Zhao, and W. Gao. Adaptive transform with hevc intra coding. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1 – 6. IEEE, 2015.

[42] Ch.Ch. Chi, M. A. Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl. Parallel scalability and efficiency of hevc parallelization approaches. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1827 – 1838, 2012. https://doi.org/10.1109/TCSVT.2012.2223056.

[43] K. Misra, A. Segall, M. Horowitz, Sh. Xu, A. Fuldseth, and M. Zhou. An overview of tiles in hevc. *IEEE Journal of Selected Topics in Signal Processing*, 7:969 – 977, 2013. https://ieeexplore.ieee.org/abstract/document/6547985.

[44] P. Habermann, B. Juurlink, C. C. Chi, and M. Alvarez-Mesa. Efficient wavefront parallel processing for hevc cabac decoding. In *28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pages 339 – 343. IEEE, 2020. https://ieeexplore.ieee.org/document/9092425.

[45] R. Sjoberg, Y. Chen, A. Fujibayashi, M.M. Hannuksela, J. Samuelsson, and T.K. Tan. Overview of hevc high-level syntax and reference picture management.

IEEE Transactions on Circuits and Systems for Video Technology, 22:1858 – 1870, 2012. https://ieeexplore.ieee.org/document/6324417.

[46] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, Ch. Rosewarne, K. Sharman, J. Sole, and J. Xu. Overview of the range extensions for the hevc standard: Tools, profiles, and performance. *IEEE Transactions on Circuits and Systems for Video Technology*, 26:4 – 19, 2016. https://ieeexplore.ieee.org/abstract/document/7265015.

[47] H. Jo, S. Park, and K. Ryoo. High-throughput architecture of hevc cabac binary arithmetic encoder. *International Journal of Control and Automation*, 10:199 – 208, 2017. http://dx.doi.org/10.14257/ijca.2017.10.5.19.

[48] S. Khattak, R. Hamzaoui, S. Ahmad, and P. Frossard. Fast encoding techniques for multiview video coding. *International Journal of Control and Automation*, 28:569–580, 2013. https://doi.org/10.1016/j.image.2012.12.010.

[49] Y. Chen, M. M. Hannuksela, L. Zhu, A. Hallapuro, M. Gabbouj, and H. Li. Coding techniques in multiview video coding and joint multiview video model. In *2009 Picture Coding Symposium*, pages 339 – 343. IEEE, 2009. https://doi.org/10.1109/PCS.2009.5167472.

[50] C. Jiang and S. Nooshabadi. Decision zone-based parallel fast motion and disparity estimation scheme for multiview coding. In *Data Compression Conference (DCC)*, pages 608 – 609. IEEE, 2016. https://doi.org/10.1109/DCC.2016.20.

[51] A. Belbel, A. Bekhouch, N. Doghmane, S. Harize, and N. Kouadria. Exploring the use of early inter-view offset estimation for low complexity mv-hevc. In *23rd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1 – 6. IEEE, 2020. https://doi.org/10.1109/WPMC50192.2020.9309453.

[52] A. Ait-Beni-Ifit, O. Fdili, P. Corlay, F. Coudoux, and M.E. Hassouni. Overview on hevc inter frame video coding's impact on the energy consumption for next

generation wvsns. In *International Conference on Model and Data Engineering*, page 135–145. Springer, 2019. https://doi.org/10.1007/978-3-030-32213-7$_1$0.

[53] K.Y Kim, H. Y. Kim, J.S. Choi, and G.H. Park. Mc complexity reduction for generalized p and b pictures in hevc. *IEEE Transactions on Circuits and Systems for Video Technology*, 24:1723 – 1728, 2014. https://doi.org/10.1007/s11554-019-00923-5.

[54] V.D. Maksimovic, J.M. Todorovic, P.L. Spalevic, M.B. Petrovic, and B.S. Jaksic. The impact of successive b frames on video using h.264 and h.265 compression techniques. In *26th Telecommunications Forum (TELFOR)*, pages 1–4. IEEE, 2018. https://doi.org/10.1109/TELFOR.2018.8611854.

[55] Y. Chen, L. Zhang, V. Seregin, and Y. Wang. Motion hooks for the multiview extension of hevc. *IEEE Transactions on Circuits and Systems for Video Technology*, 24:2090 – 2098, 2014. https://doi.org/10.1007/s11554-019-00923-5.

[56] W. Ahmad, M. Ghafoor, S.A. Tariq, A. Hassan, M. Sjostrom, and R. Olsson. Computationally efficient light field image compression using a multiview hevc framework. *IEEE Access*, 7:143002 – 143013, 2019. e http://creativecommons.org/licenses/by/4.0/.

[57] V.A. Borges, M.R. Perleberg, V. Afonso, M.S. Porto, and L. V. Agostini. Hardware and coding efficiency assessment of 3d-hevc dis tool using alternative similarity criteria. *Springer Nature : Analog Integr Circuits Signal Process*, 108:555–568, 2021. https://doi: 10.1007/s10470-021-01911-1.

[58] J. Bégaint, F. Galpin, Ph. Guillotel, and Ch. Guillemot. Deep frame interpolation for video compression. In *Data Compression Conference (DCC)*, pages 1–4. IEEE, 2019. https://doi.org/10.1109/DCC.2019.00068.

[59] Zhaoqing Pan, Jianjun Lei, Yun Zhang, Xingming Sun, and Sam Kwong. Fast motion estimation based on content property for low-complexity h. 265/hevc encoder. *IEEE Transactions on Broadcasting*, 62(3):675–684, 2016.

[60] Sangsoo Ahn, Bumshik Lee, and Munchurl Kim. A novel fast cu encoding scheme based on spatiotemporal encoding parameters for hevc inter coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):422–435, 2014.

[61] Gangyi Jiang, Baozhen Du, Shuqing Fang, Mei Yu, Feng Shao, Zongju Peng, and Fen Chen. Fast inter-frame prediction in multi-view video coding based on perceptual distortion threshold model. *Signal Processing: Image Communication*, 70:199–209, 2019.

[62] Peicheng Wang, Xingang Liu, and Binfei Shao. A fast cu decision algorithm for mv-hevc. In *2015 IEEE International Conference on Smart City/SocialCom/-SustainCom (SmartCity)*, pages 217–221. IEEE, 2015.

[63] Zhaoqing Pan, Xiaokai Yi, and Liming Chen. Motion and disparity vectors early determination for texture video in 3d-hevc. *Multimedia Tools and Applications*, 79(7):4297–4314, 2020.

[64] Yue Wang, Tingting Jiang, Siwei Ma, and Wen Gao. Novel spatio-temporal structural information based video quality metric. *IEEE transactions on circuits and systems for video technology*, 22(7):989–998, 2012.

[65] S. N. Khan, K. Khan, N. Muhammad, and Z. Mahmmod. Efficient prediction mode decisions for low complexity mv-hevc. *IEEE Access*, 9:150234 – 150251, 2021. https://ieeexplore.ieee.org/document/9605631.

[66] B. Girod and N. Farber. Feedback and error protection strategies for wireless progressive video transmission. *Proc. IEEE*, 87:465 – 482, 1999. https://ieeexplore.ieee.org/document/6324417.

[67] T. Stockhammer, H. Jenkac, and C. Weiss. Feedback-based error control for mobile video transmission. *IEEE Transactions on Circuits and Systems for Video Technology*, 12, 2002. https://ieeexplore.ieee.org/document/6324417.

[68] S. Ghahremani and M. GhanbariK. Error resilient video transmission in ad hoc networks using layered and multiple description coding. *Multimed Tools Appl*, 76:9033–9049, 2017. https://ieeexplore.ieee.org/document/6324417.

[69] S. Lawan and A. H. Sadka. Robust adaptive intra refresh for multiview video. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, 4, 2014. https://ieeexplore.ieee.org/document/6324417.

[70] S. Lawan, A. A Adams, and R. Ladodo. Efficient adaptive intra refresh error resilience for 3d video communication. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, 8.

[71] W.Y. Kung, C.S. Kim, and C.C.J. Kuo. Spatial and temporal error concealment techniques for video transmission over noisy channels. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 16:789 – 803, 2006. https://ieeexplore.ieee.org/document/6324417.

[72] Y. Li and R. Chen. Motion vector recovery for video error concealment based on the plane fitting. *Multimed Tools Appl*, 76:14993–15006, 2017. https://ieeexplore.ieee.org/document/6324417.

[73] K. Song, T. Chung, Y. Ohb, and Ch. SuKim. Error concealment of multi-view video sequences using inter-view and intra-view correlations,. *J. Vis. Commun. Image R.*, 20:281–292, 2009. https://ieeexplore.ieee.org/document/6324417.

[74] H. Schwarz, D. Marpe, and T. Wiegand. Analysis of hierarchical b pictures and mctf. *IEEE International Conference on Multimedia and Expo,Toronto,*, page 1929–1932, 2006. https://ieeexplore.ieee.org/document/6324417.

[75] G. Pan and W. Xiang. Rate-distortion optimized mode switching for error-resilient multi-view video plus depth based 3-d video coding. *IEEE Transactions on Multimedia*, 16:1797 – 1808, 2014. https://ieeexplore.ieee.org/document/6324417.

[76] G. Pan, P. Qiang, and W. Qionghua. Error-resilient multi-view video coding based on end-to-end rate-distortion optimization. *Chinese Journal of Electronics*, 25:277–283, 2016. https://ieeexplore.ieee.org/document/6324417.

[77] M.B. Dissanayake, D.V.S.X. De-Silva, S.T. Worrall, and W.A.C. Fernando. Error resilient technique for mvc using redundant disparity vectors. *Proc. of IEEE International Conference on Multimedia and Expo*, page 1712–1717, 2010. https://ieeexplore.ieee.org/document/6324417.

[78] A.B Ibrahim and A.H Sadka. Effects of gop on multiview video coding over error prone channels. *Third International Conference on Advanced Information Technologies and Applications,*, page 135–150, 2014. https://ieeexplore.ieee.org/document/6324417.

[79] A. Belbel, A. Bekhouch, N. Doghmane, S. Harize, and N. Kouadria. Improved inter-view correlations for low complexity mv-hevc. *Journal of Visual Communication and Image Representation*, 86:1–10, 2022. https://doi.org/10.1016/j.jvcir.2022.103525.

[80] Shahid Nawaz Khan, Khurram Khan, Nazeer Muhammad, and Zahid Mahmood. Efficient prediction mode decisions for low complexity mv-hevc. *IEEE Access*, 9:150234–150251, 2021.

[81] Siham Bakkouri, Abderrahmane Elyousfi, and Hamza Hamout. Fast cu size and mode decision algorithm for 3d-hevc intercoding. *Multimedia Tools and Applications*, 79(11):6987–7004, 2020.

[82] K. M¨uller and A. Vetro. Common test conditions of 3dv core experiments. *JCT-3V document JCT3V-G1100*, 2014.

[83] H.P. Thi, H.H Van, and T. Miyoshi. Correlating objective factors with video quality experienced by end users on p2ptv. *International Journal of Computer Networks and Communications*, 17:59–75, 2015. https://www.academia.edu/download/37831813/5.pdf.

[84] T. T. Alfaqheri and A.H. Sadka. Low delay error resilience algorithm for h.265|hevc video transmission. *Journal of Real-Time Image Processing*, 17:2047–2063, 2020. https://doi.org/10.1007/s11554-019-00923-5.