BADJI MOKHTAR- ANNABA UNIVERSITY

جامعة باجي مختار- عنابة

Faculty: Engineering Sciences
Department: Electronics

Year: 2020

THESIS

Presented for the obtention of the Master's degree

Title:

Study and Analysis of Multiview Video Coding

Field: Telecommunications.

Specialty: Networks and telecommunications.

By:

Boumachta Meriem & Merzoug Fatima

Supervisor: Dr. NASRI Seif Allah El Mesloul

In Front of the JURY

President : Dr. Saliha HARIZE University of Annaba

EXAMINER: Prof. Noureddine DOGHMANE University of Annaba

We dedicate this work to our families, friends and everyone who supported us to complete this thesis.

Acknowledgements

All praises and thanks are for Almighty Allah the most gracious and merciful, for helping us in the completion of this study.

Our deep respect and appreciation to our family's members; their tender care and patience have not ceased even in our worst moments.

Our deepest and profound acknowledgments are to our supervisor Dr. NASRI Seif Allah El Mesloul for his continuous support, generous helps, and fruitful constructive suggestions.

We are also thankful to Dr. Saliha Harize and Prof. Noureddine Doghmane for kindly accepting to examine and review this Master thesis.

We are especially indebted to the outstanding staff of Electronic Department at Annaba University for their useful assistance, valuable advices, the delivered teaching materials and sessions during our Bachelor and master studies

Finally, we thank the countless people who contributed to this research and anyone who helped us in any way.

Abstract

This dissertation presents an analysis of coding multiview videos using latest video codecs. In fact, multiview video is a form of 3D imaging technology which offers viewers an enriched experience with depth perception and multiple viewing angles. MVV is generated when multiple arranged cameras are used to capturing the same scene. Compared to the traditional 2D video that only employs one camera, MVV technology uses two or more cameras with specific settings and arrangements. The output raw data of MVV capturing system is usually of a huge size and therefore needs particular coding techniques before storage and/or transmission. Coding techniques of MVV are based on exploiting similarities of the adjacent views during the coding processes in addition to spatial and temporal correlations that exist within each view. Video coding standards such as H.264 and H.265 have delivered extended profiles, MVC and MV-HEVC respectively, that take into consideration views dependencies. This research manuscript introduces multiview video technology with a special focusing on its coding techniques. In addition to the reported literature review about MVV coding, experiments were carried out to investigate recent codecs standards effect in terms of compression efficiency and complexity. As a first step, MV-HEVC was tested using different test conditions and distinct video sequences. Furthermore, versatile video coding (VVC) which has been recently released in July 2020, is reported, tested and compared to MV-HEVC in this dissertation. Bit rate and PSNR (peak signal to noise ratio) are used as objective evaluation metrics for assessing the compression efficiency. Encoding time is used to evaluate the codecs complexity.

Key word: **3D, 2D, MVV, H.264, H.265, MVC, MV-HEVC, HEVC, VVC.**

Résumé:

Cette thèse présente une analyse du codage de vidéos multi-vues en utilisant des codecs vidéo récents. En fait, la vidéo multi-vues est une forme de technologie d'imagerie 3D qui offre aux téléspectateurs une expérience enrichie avec une perception de la profondeur et des angles de vision multiples. MVV est généré lorsque plusieurs caméras disposées sont utilisées pour capturer la même scène. Par rapport à la vidéo 2D traditionnelle qui n'utilise qu'une seule caméra, la technologie MVV utilise deux caméras ou plus avec des paramètres et des arrangements spécifiques. Les données brutes de sortie du système de capture MVV sont généralement d'une taille énorme et nécessitent donc des techniques de codage particulières avant le stockage et / ou la transmission. Les techniques de codage de MVV sont basées sur l'exploitation des similitudes des vues adjacentes pendant les processus de codage en plus des corrélations spatiales et temporelles qui existent dans chaque vue. Les normes de codage vidéo telles que H.264 et H.265 ont fourni des profils étendus, MVC et MV-HEVC respectivement, qui prennent en compte les corrélations entre les vues. Ce manuscrit de recherche présente la technologie de vidéo multi-vue avec un accent particulier sur ses techniques de codage. En plus de l'étude théorique rapportée sur le codage MVV, des expériences ont été menées pour analyser les codecs vidéo récents en termes d'efficacité et de complexité de compression. Dans un premier temps, le MV-HEVC a été testé en utilisant différentes conditions de test et des séquences vidéo distinctes. En outre, le codec (VVC) qui a été récemment publié en juillet 2020 est rapporté, testé et comparé au MV-HEVC dans cette thèse. Le débit binaire et le PSNR sont utilisés comme des mesures d'évaluation objectives pour évaluer l'efficacité de la compression. Le temps de codage est utilisé pour évaluer la complexité des codecs.

Mot clé: 3D, 2D, MVV, H.264, H.265, MVC, MV-HEVC, HEVC, VVC.

ملخص

تقدم هذه الأطروحة تحليلًا لترميز مقاطع فيديو "Multiview" باستخدام أحدث برامج ترميز الفيديو. في الواقع، يعد الفيديو متعدد العروض شكلاً من أشكال تقنية التصوير ثلاثي الأبعاد التي تقدم للمشاهدين تجربة ثرية من حيث إدراك العمق وزوايا المشاهدة المتعددة. يتم إنتاج MVV عن طريق استخدام عدة كاميرات مرتبة لالتقاط نفس المشهد. مقارنة بالفيديو التقليدي ثنائي الأبعاد الذي يستخدم كاميرا واحدة فقط، تستخدم تقنية MVV كاميرتين أو أكثر مع إعدادات وترتيبات محددة. عادةً ما تكون البيانات الأولية الناتجة من نظام التقاط MVV ذات حجم ضخم وبالتالي تحتاج إلى تقنيات تشفير معينة قبل التخزين و / أو الارسال. تعتمد تقنيات الترميز الخاصة بـMVV على استغلال أوجه التشابه بين جهات التصوير المتجاورة أثناء عمليات التشفير بالإضافة إلى التجانسات المكانية والزمانية الموجودة في كل جهة على حدى. قدمت معايير ترميز الفيديو مثل H.264 و H.265 ملفات تعريف موسعة، MVC وMV-HEVC، والتي تأخذ في الاعتبار التشابه بين جهات التصوير. يقدم هذا البحث تقنية الفيديو متعدد العروض مع التركيز بشكل خاص على تقنيات الترميز الخاصة بها. بالإضافة إلى الدراسة النظرية على تكنولوجيا MVV، أجريت تجارب لتحليل برامج ترميز الفيديو الحديثة من حيث كفاءة الضغط والتعقيد. أولاً، تم اختبار MV-HEVC باستخدام ظروف اختبار مختلفة ولقطات فيديو متباينة. علاوة على ذلك، تم تقديم برنامج الترميز (VVC) الذي تم إصداره مؤخرًا في جويلية 2020 واختباره ومقارنته بـ MV-HEVC في هذه الأطروحة. تم استخدام معدل البث و PSNR كمقاييس موضوعية لتقييم فعالية الضغط. كما يستخدم وقت الذي تستغرقه عملية الترميز لتقييم مدى تعقيد برامج الترميز.

List of abbreviations

2D	Two- dimensions
3D	Three-dimensions
3DTV	Three-dimensional Television
ALF	Adaptive Loop Filter
AVC	Advanced Video Coding
AU	Access Unit
CABAC	Context-based Adaptive Binary Arithmetic Coding
CAVLC	Context-based Adaptive Variable Length Coding
DBF	Deblocking Filter
DC	Discrete cosine
DCT	Discrete cosine transform
DST	Discrete sine transform
FPS	Frame Per Second
FVV	Free Viewpoint Video
GOP	Group of pictures
GO-GOP	Group of Group of pictures
HD	High Definition
HDTV	High Definition television
HEVC	High Efficiency Video Coding
HVS	Human Visual System
HLS	High level serval
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union –Telecommunications Standardization Sector
JCT-VC	Joint Collaborative team on Video Coding
JMVC	Joint Multiview Video Coding
JMVM	Joint Multiview Video Model
JVT	Joint Video Team
LCD	Liquide Crystal Display
MB	Macroblock
MC	Motion Compensation
ME	Motion Estimation
MPEG	Moving Pictures Experts Group
MV	Motion vector
MVC	Multiview Video Coding
MVV	Multiview Video
NAL	Network Abstraction Layer
PSNR	Peak Signal to Noise Ratio
PU	Prédiction unit
QP	Quantization Parameter
QTMT	Quadtree with nested multi- type tree
SAO	Sample Adaptativ offset
TU	Transform unit
TLV	Television
URQ	Uniform reconstruction quantization
UHD	Ultra high Definition
RTP/IP	Real-time Transport Protocol/Internet protocol
TCP/IP	Transmission Control Protocol/Internet Protocol

List of figures

Chapter 1

Chapter 2

Chapter 3

List of tables:

Chapter 1

Chapter 2

Chapter 3

Contents:

General Introduction

General Introduction:

In the prehistoric past, there was no technology or modernity. Man depended on drawing and sculpture on rocks, caves, and other tools for communication and for delivering a meaningful content. Over time humanity searched for ways and methods to develop and improve their life quality.

Nowadays, the technology is developing in all fields generally and exponential advancement in communication and information specifically. A study presented by cisco stats that video traffic on the internet will take over 82% of all communications by the year 2021 [1]. Demand for more advanced and interactive video technologies has also grown rapidly. 3D video technology is an important type of advanced video technology, which is used in various fields such as: education, medicine ... etc.

This technology aims to mimic the real 3D world and provides angular details and depth perception .3D imaging [2] [3] [4] has many systems; each is characterized by its specific capture, coding and viewing techniques. 3D imaging systems [5] can be split into three main categories:

- **Stereoscopic**: (Also known as Stereoscopy / stereo imagery) is a technique that creates the depth illusion of an image using stereopsis for binocular vision [6][7].
- **Head-mounted system**: Head-mounted displays (HMDs) or Head-worn displays (HWDs) are designed as on-body devices and coupled with the human eyes to support mobile users.
- **Autostereoscopic**: Offers the viewers a glasses-free 3D experience and guarantees a sense of depth from different angles.

There are several types of 3D autostereoscopic systems from which we briefly mention the most leading technologies:

- **Multiview Video (MVV) technology** [2]: Multiview video is composed of a set of synchronized cameras capture the same scene from various positions.
- **3D Stereoscopic Display (3DS)**: Where glasses are required to feel the sensation of depth.
- **Autostereoscopic multi-view displays (AMD)**: Where the 3D effects are perceived with no glasses. AMD offers an enriched experience compared to 3D.

MVV technology is considered as one of the exceedingly important standards of 3D technology. Where applied in various fields such as games, cinema, education… etc.

Multiview videos needs an enormous amount of data. If no compression techniques are applied, storage or transmission of MVVs could be difficult or even impossible with the conventional storage devices and bandwidth capabilities. Thereby, multiview video codecs are proposed as fundamental techniques to efficiently compressing and adapting MVV for storage and transmission over different networks.

Recent video coding standards such as H264, H265 and, H266, works to reduce the amount of video data and maintains decent visual quality.

Video compression takes advantage of similarities that exist within and between video images. Since several cameras are used to capture the same scene from separate locations to produce multiview video, there will be similarity between adjacent broadcasts views. Multiview video encoding must meet a list of requirements to ensure a decent 3D service for MVV users. This thesis deals exclusively with MVV technology as a research case. Accordingly, the thesis presents a literature review of the fundamental concepts of 2D and multiview video coding by describing the following codecs: AVC/H.264, MVC/H.264, HEVC, MV-HEVC and VVC.

The dissertation is structured in the following way:

Chapter 01: Discusses the fundamentals of multi-view video system, where a considerable amount of attention is given to the coding part. It begins with introductory sections of 3D imaging history and concepts. In addition, it gives a brief overview of the entire multiview video system chain covering acquisition and display techniques as well as the basics of multi-view coding.

Chapter 02: Focuses on video coding concepts by providing an overview of the H.264 video coding standard, its architecture, and its key features. Chapter 2 also provides an overall overview of the concepts of HEVC coding while focusing on the MV-HEVC extended multi-view profiles and VCC standard.

Chapter 03: This chapter outlines the experimental work and discuss the obtained results of MV-HEVC and VVC coding standards in terms of compression performance (PSNR / bit rate) and encoding complexity (time).

Chapter 01

The basics of

MVV technology

1.1Introduction:

This chapter covers the main components of a multiview video system. It starts by defining 3D technology in section 1.2. Early history and development track of 3D imaging will be presented in section 1.3. Free viewpoint video (FVV) and three-dimensional television (3DTV), that are the two principal scenarios of MVV system, are presented in section 1.4. In addition, the last section details concepts of MVV (acquisition, visualisation and coding).

1.2Definition of 3D:

Three dimensions are terms that characterize the space which surrounds us and perceived by our vision. 3D relief allows seeing 3D images through devices and technical methods that allow the human brain to simulate the depth perception through three axes: width (X-axis), height (Y-axis) and depth (Z-axis). The principle of 3D is therefore to "deceive" our brain and give it the illusion of the perception of two dissociated images, so that it reconstitutes the relief. This technique is applied in several fields such as entertainment, military simulations, medical applications, video game, ... etc.

1.3History of 3D video:

The 3D technology video was invented several years ago, when Charles Wheatstone (the father of 3D) has constructed the first stereoscope in 1832. His idea is based on two mirrors that capture the same image, which will be reflected at a 90-degree angle, and the result provides a 3D visualisation of the original image.

Figure 1.1: Wheatstone's stereoscope [8]

In 1838 the physicist David Brewster produced the structure of the stereoscope in 3D vision / image. In 1840, photography was invented. Consequently, drawings and painting were replaced by photography in existing stereoscopic cameras. During 1844, Brewster further improved the stereoscope by adding prismatic lenses to fuse and enlarge stereo images and improve quality. Later in 1903, the first film using 3D stereoscopic was released publicly technology by Lumiere brothers. After that the demonstration of the first stereoscopic television (TLV) happened 1928. The 50's (1950) were considered as the first gold age of the 3D cinematographic industry. The second golden age was triggered with the famous movie "The Matrix", released in 1999. it was a successful application of multiview system cameras.

A group of 120 precision-operated cameras was installed to produce virtual freezes and slow-moving effects.

Figure 1.2: Multiview system used in The Matrix film [2]

The world of 3D cinema has seen huge success after the Avatar movie in 2010, the film combined between real persons and computer graphic generated images 3D technology [2].

1.4 Multiview Video (MVV):

Multiview video (MVV) is an extension of the conventional stereo video with a higher number of views to capture the global action of the scene with different position. The utilization of more than one camera allows to take multiple simultaneous shots captures, are obtained in a single take without having to start and stop the action [2].

1.4.1 MVV System:

MVV system in mainly divided into two scenarios: FVV (free viewpoint video) and three-dimensional television (3DTV).

- **Free Viewpoint (FVV):**

The Free-viewpoint video (FVV), also known as free-viewpoint television, which allows users to freely move through the scene, navigating along an arbitrary trajectory as if there were a virtual camera positioned anywhere in the scene. This functionality can enhance the user experience in broadcasting of events, including sports or interactive video communication and more [46].

. **3DTV (Three-dimensional Television):**

This concept inspired by the human visual system, where the human eyes have a mean distance. Thus, we see a slightly different picture with the left eye in comparison to the right eye. Stereoscopic 3D is the basic application of the multiview system, where two images are transmitted to the viewers. Moreover, N views can be employed depending on capturing and displaying capabilities [2].

1.4.2 Multiview Acquisition:

The principle of this process depends on the cameras number and arrangement. The most used multiview cameras arrangements are shown in the figure below:

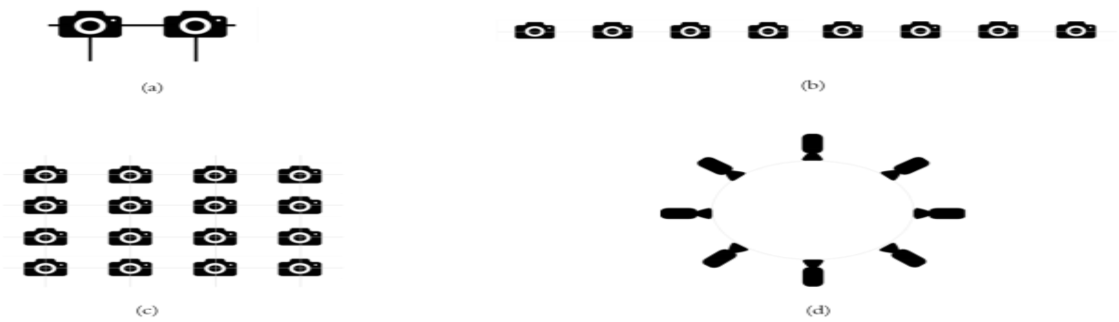

Figure 1.3: Multiview video acquisition arrangement [2]

(a) Binocular system, (b) Linear system , (c) Bidimensional arrays system ,

(d) Omnidirectional system

(a) Binocular System: This system contains two connected cameras which work in the same way of human visual system (HVS), this system is applied in 3D stereo visualisation and requires specific glasses for depth perception.

(b) Linear System: This system is a collection of connected cameras in a form of a horizontal array. This configuration provides only one plan for navigation from the point of view; it greatly facilitates the estimation of the scene depth. This system produces 3D content for autostereoscopic screens.

(c) Bidimensional Array System: A combination made up of cameras placed vertically and horizontally in order to create 2D linear arrays. It supports horizontal and vertical motion parallax.

(d) Omnidirectional or Global Systems: Multiples cameras cover to the center of the scene. It is mainly oriented for free video navigation.

All Multi-view acquisition systems must support intrinsic settings like ISO, shutter speed and aperture for all cameras in order to produce good video. Synchronizing accurately the multiple cameras and using similar frame rates facilitate the integration of multi-view video data [2].

1.4.3 MVV Display:

Multiview video data can be displayed on a wide range of 3D display technologies. In this part we will only consider autostereoscopic screens that can support MVV. Offer an instant 3D effect and does not depend on specific glasses, instead of exposing only one left and one right images. Multi-view displays support n > 2 views. Autostereoscopic display distributes several stereoscopic pairs to define the observation areas. Thus, the observers move horizontally and visualize several display windows based on the parallax effect has two optical methods used to design auto-stereoscopic multi-view display. Both techniques are cited below [2]:

- **Parallax barrier technique:**

Parallax Barrier is an optical barrier located in front of a source of image like an liquid crystal display (LCD) display or any other pixelated displays. To allow it to display a stereoscopic image without the need of the viewer using 3D eyewear [20] a vertical barrier separates the columns of images from the left and right eye alternately. Thus, light may only pass through the necessary viewing areas. On the other side, this technique suffers from a lighting problem [2], which causes a subdued brightness and if the screen is tilted, the 3D illusion disappears [21].

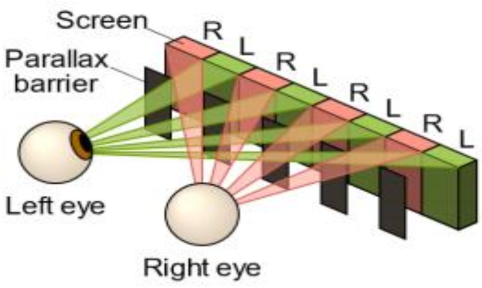

Figure 1.4: Autostereoscopic multiview display based on:" **Parallax Barrier Technique**" [22]

- **Lenticular technique:**

The lenticular technique is widely used in the manufacture of both printed and electronical displays. It is based on cylindrical lenses which a placed in front of the pixelated displays. These lenses work as tiny loupes, allowing each of the eyes to see a single point among the various mixed views in each image. Nowadays, the lenticular filter technology is the main technology used in the commercially available multi-view displays.

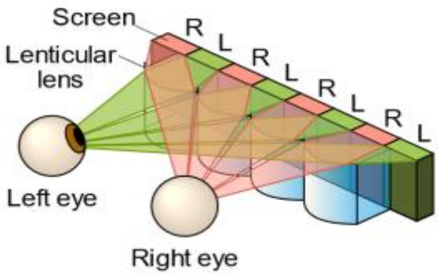

Figure 1.5: Autostereoscopic multiview display based on:" **Lenticular Technique**" [22]

1.5 Multiview Video Coding (MVC):

Multiview Video Coding (MVC, also known as MVC 3D) is a stereoscopic video coding standard for video compression that allows for the efficient encoding of video sequences captured simultaneously from multiple camera angles in a single video stream [23]. The compression is a necessary process to broadcast multi-view video content over transmission channels. To do so, you must use special coding techniques which consider reducing the volume of data. Fortunately, multi-view video has a high correlation between views. This correlation may be used to further improve compression. It is important to note that MVV sequences need to be simultaneously coded by the video codec itself. Figure 1.6 shows the three-dimensional correlation that can be found in multi-width video content [2].

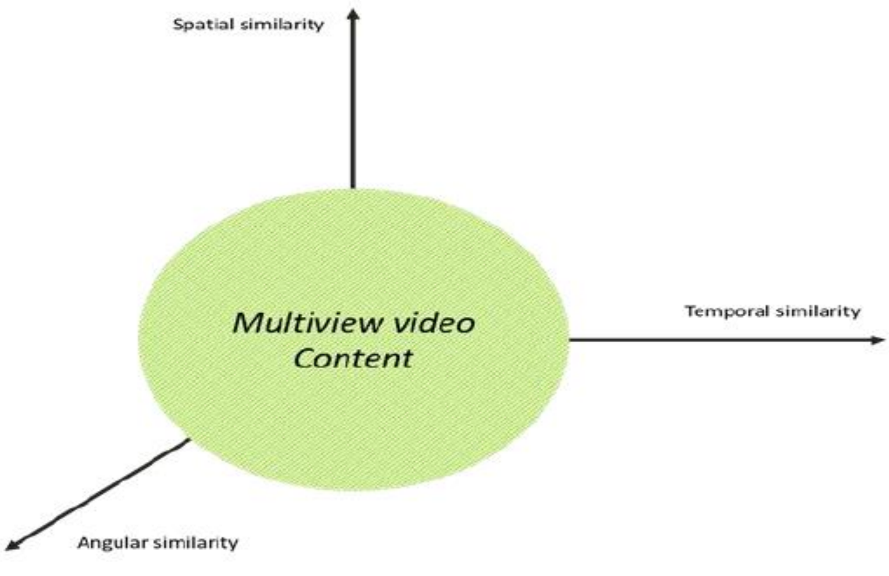

Figure 1.6: Multiview Video Correlation Types [2]

Table 1.1: The role of each type of correlation [11]

Type of correlation	**Description**
Spatial correlation	The similarity blocks region of the same frame
Temporal correlation	Represents the similarities between different frames in the same views of a video sequence
The interview correlation noted "angular similarity"	It refers to the similarity between frames of different views

1.5.1 Multiview Video Coding (MVC) History:

It is so hard to know the exact date when MVC was invented, Lukas made one of the first noted proposals in 1986. His research introduced the concept of interview prediction, other experiments, then followed, notably the work of Dinstein et al [44] in 1989. Perkins [45] described a mixed resolution coding structure as well as a transformation domain technique for disparity compensated prediction [2].

The first international standard that supports MVC was presented in 1996 and consisted of extending H.262/MPEG-2 to only support encoding of two views. In this first multiview standard, the left view was chosen as the base view which offers compatibility with the conventional H.262/MPEG-2 decoder. Following the progress in video compression technologies and multimedia services, the International Telecommunication Union – Telecommunications Standardization Sector (ITU-T) Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) joint forces and form a collaborative team namely Joint Video Team (JVT) in 2001. Meanwhile, the subgroup Three-dimensional audio visual (3DAV) of Moving Pictures Experts Group (MPEG) triggered the standardization process of MVC in 2005 after receiving evidential outputs of some proposed multiview video coding schemes. The MVC scheme based on the codec H264, with hierarchical structure [2]. A joint multi-view video model (JMVM) has been developed as an extension of the H.264/MPEG-4 AVC to support the development of the future MVC reference software [24].

1.5.2 MVC Requirement:

There is a checklist of requirements that must be respected during the development of any video coding scheme. Therefore, the main requirements of MVC are defined below:

- **Compression efficiency:**

MVC must achieve a high compression efficiency when compared to encoding each frame of the same content independently [24]. Compression efficiency can be evaluated in terms of PSNR (dB) relative to the bit rate of the compressed video [2].

- **Random access [25][2]:**

Low-delay Random Access is high on the list of video coding requirements. Random access ensures that any frame within the multi-vision video architecture can be accessed, coded, decoded, and rendered with a relatively low delay. Fast random access improves user interaction and navigation through multi-view video content. Especially for applications where view switching such as FVV is required, MVC diagrams must be carefully designed to reduce the number of decoded frames between views.

- **Scalability:**

Scalability is an essential requirement for video coding models. It allows decoders to access part of the bit stream while being able to generate decent video and display it on the terminal. This technique allows any part of the video bit stream to be accessed by the decoder to produce suitable video quality. Scalability improves the interoperability of the same video stream across different networks and terminals. It offers multiple levels of resolution and different frame rates of the same video. It allows MVV content to be displayed on screens with a limited number of views.

- **Backward compatibility:**

At all times, the bitstream corresponding to a view must be conforming to AVC [24].

- **Low-delay coding:**

The MVC must provide support for low delay coding and decoding modes. Low delay mode is important for real-time applications such as streaming and broadcasting using MVV. [24]

- **Robustness:**

Robustness against errors, also called error resilience, must be supported. This allows the delivery of MVV contents over error-prone networks, including wireless and other networks.[24]

- **Parallel Processing:**

MVC should support the parallel processing within different frames or segments of MVV to facilitate the efficient implementation of encoders and decoders [24].

- **Resource consumption:**

MVC must be efficient in terms of resources consumption such as processing power, used memory and bandwidth occupation. The MVC should be able to exploit the similarity of the interviews without significantly increasing the complexity of the coding as this could interfere with the smooth 3D display [24].

- **System support requirements [24]:**

- **Synchronization:**

MVC should support exact time Synchronization between multiple views.

- **Display Generation:**

MVC must enable the robust and efficient generation in virtual or interpolated views.

- **Non-planar imaging and display systems**:

MVC should support efficient representation and encoding methods for 3D display, including integral photography and display systems for non-planar (e.g. dome) images.

- **Camera Parameters:**

MVC should support transmission of camera parameters.

1.6 Conclusion:

This chapter highlighted 3D video development history and its current status. A synthesis study of the 3D video production chain was then presented, including the three main parts of the 3D video production chain: 3D video acquisition, 3D display and 3D video encoding. More attention was given to the 3D video coding part, and more precisely to multi-view video coding, its concepts, requirements, and the compression technology underlying it.

Chapter 02

Video coding technologies

2.1 Introduction:

In April 2005, the H265 standard was initiated by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG), who are working together in a partnership well known as the Joint Collaborative Team on Video Coding (JCT-VC). JCT-VC is a working group to prepare for the future generation of standard encoders. Although the H.264/AVC video coding standard was efficient in terms of compression rate and visual quality, it was not efficient enough to keep up with the evolution of visual technology. However, the growing popularity of HD video, the emergence of technologies beyond high definition (Ultra HD: 4k, 8k format), 3D or "multiview", and the increased desire to use high resolutions with excellent visual quality, especially in mobile applications, impose strict constraints for encoding that exceed the capabilities of the H.264/AVC standard. This standard requires high-resolution images, which means better image definition by improving color, contrast, and frame rate. With its improved coding structure and several modifications that significantly increase the coding choices of the competition. High Efficiency Video Coding (HEVC) is the most powerful encoder available. Featuring (H265) up to 50% throughput savings over H264. As a result, the second version of the HEVC with the multiview extensibility extension (MV-HEVC) was completed in 2014 and released in early 2015.

2.2 Brief History of video coding:

In November 1992, the Moving Pictures Expert Group completed its first standard for video and audio, called MPEG-1[14].

In 1994, MPEG 02 [47] was released as a development of MPEG1; which is also known as ITU.2, it supported interlaced video coding. The development of video coding for telecommunication applications evolved with the development of the ITU-T video coding standards (H.261, H.262 (MPEG-2) and H.263).

MPEG-4 Visual (MPEG-4 Part 2) has also started to emerge in some areas of application of previous coding standards. The first design proposal of this new standard was adopted in October 1999. In December 2001, VCEG and the Moving Picture Experts Group (MPEG) ISO formed a Joint Video Team (JVT), with a charter to finalize the draft of the new video coding standard for submission of formal approval as H.264 / AVC in March 2003.

After that a new standard comes high efficiency video coding HEVC with better compression efficiency compared to the previous one (H264). The most recent standard H.266 was published in July 2020 with more features and enhanced coding capabilities [41]. Main aforementioned standards will be detailed in the following sections.

2.3 Basics of H.264/AVC:

MVC standard is an extended profile of H. 264 video codec. In this section, we provide a brief description of the H.264 key features, which are also used with slight differences in MVC core. Developed in 1998-2003, H.264 is a standard approved by ITU-T recommendation as ISO/IEC international standard 14496-10 (MPEG-4 part 10) advanced video coding (AVC) [17].

2.3.1 Structure of H.264:

Compression efficiency and integration with transport protocols impact on the global performance, consequently H.264 is organized into two conceptual (layers**):** Video Coding Layer (VCL) and Network Adaption Layer (NAL) [17].

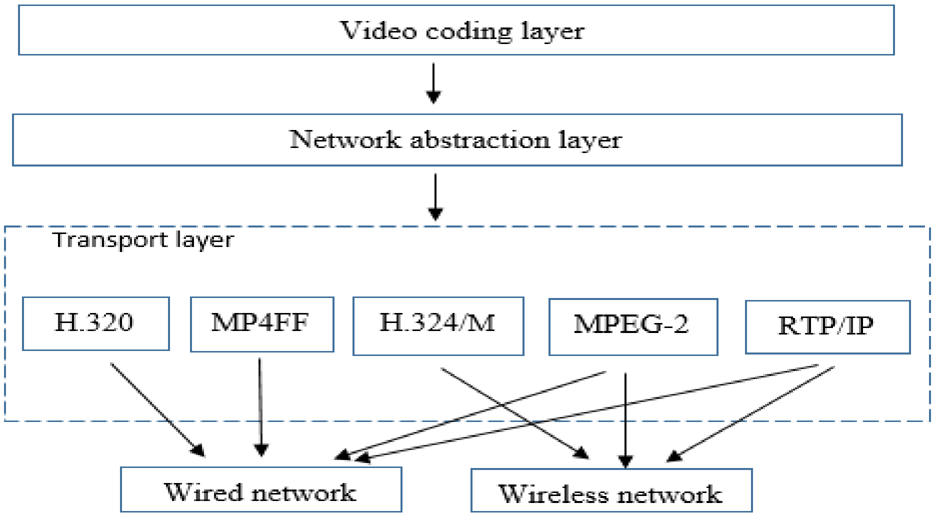

Figure 2.1: H.264 structure

- **VCL**: Offers performing compression tools as the intra-prediction, variable-size motion compensation /motion estimation (ME/MC), in-loop filtering, context-based adaptive binary arithmetic coding (CABAC)… etc.
- **NAL**: Allows the adaptation to different transport types as packet switched transport (RTP/IP, TCP/IP, . . .) and circuit switched transport (MPEG-2, H.320, . . .) [17].

The **VCL** layer efficiently represents the content of the video data. The **NAL** layer formats data and provides header information to adapting transmission over various communication channels or media storage [16].

2.3.2 H.264 encoder:

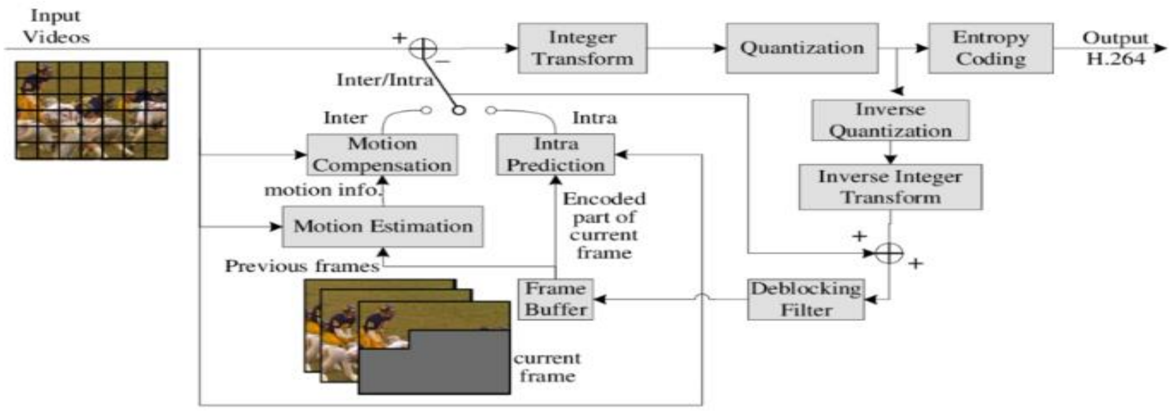

Figure 2.2 H.264 encoder structure [33]

The Input Sequence is a succession of images. Each image is slashed into "slices".A slice is a part of the image or the whole image depending on the input parameters.

These slices are divided into macroblock (MB) 16×16 block size (MB is basic frame unit). Each MB is encoded using intra or inter modes. Each of these codings generates several texture residualsto be compared later with a decision mode. The residual that gives the best coding result, in terms of a rate-distortion is then selected. This residual is decorrelated with the discrete cosine transform (DCT). The transformed residual is quantized and the generated coefficients are sent to the lossless entropy encoder which produces the bitstream. The encoder performs decoding and creates the necessary references for next predictions. Consequently, the transformed and quantized residuals are dequantified and will undergo an inverse transform inside the encoder. The inverse prediction is applied to the resulting blocks. This operation consists in adding the predictor selected in the decision mode (the best intra or inter predictor). Next, an anti-blocking filter is applied to the reconstructed image to eliminate certain degradations produced by the quantization module. Finally, the decoded MBs and slices are stored in memory. The decoded blocks of the current image, stored in this module, are used for the computation of the intra predictors. Similarly, previously decoded and unlocked (smoothed) images are used for the inter coding process [19].

The three main functions described above (temporal, spatial and entropy coding) are used in the majority of video codecs [26].

2.3.3 The Macroblock (MB):

Macroblock (MB) is the basic frame unit adopted in H.264 coding. Each MB is composed of 16 × 16 luma samples and two blocks of 8 × 8 chroma samples [2].

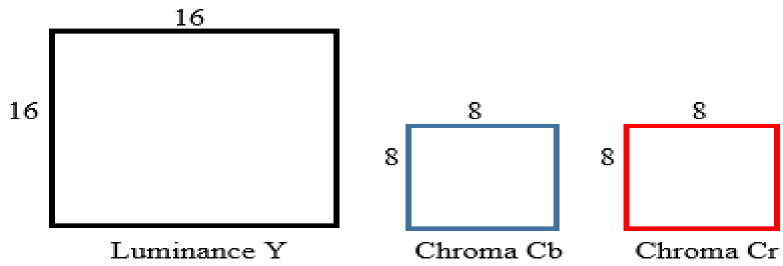

Figure 2.3: Macroblock structure in H.264

2.3.4 The slices:

In H.264/AVC, a slice corresponds to a part of an image. The MB are grouped into slices; otherwise, a slice is a set of macroblocks in raster scan order [17]. There are three types of slice are supported by H.264/AVC:

- **Slice I:** Use only intra prediction.
- **Slice P:** Uses both intra prediction and inter prediction for coding its macroblocks.

  However, only one direction is allowed for inter prediction
- **Slice B:** Uses intra prediction and inter prediction from two directions to coding it macroblocks.

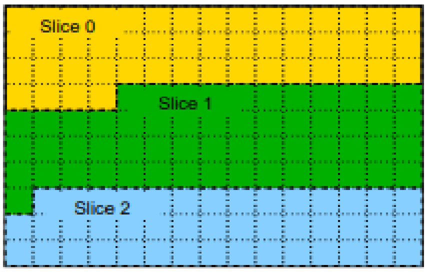

Figure 2.4: Spliting frame over slices in H.264 structure[17]

2.3.5 Intra-frame prediction :

The first image of a sequence (image I) is necessarily intra encoded because we do not yet have a reference image. Intra-frame prediction in H.264/AVC works by interpolating reconstructed adjacent pixels according to a predefined direction. As the blocks run from left to right and from top to bottom (raster scan), the reconstructed neighboring pixels belong to the previously coded blocks, i.e. they are located above and to the left of the current block. Some subtleties differentiate the prediction modes according to the partitioning used. For the luma component, the standard describes three partitioning possibilities for a MB: sixteen blocks of size 4×4, four blocks of size 8×8 or one block of size 16×16 pixels[19].

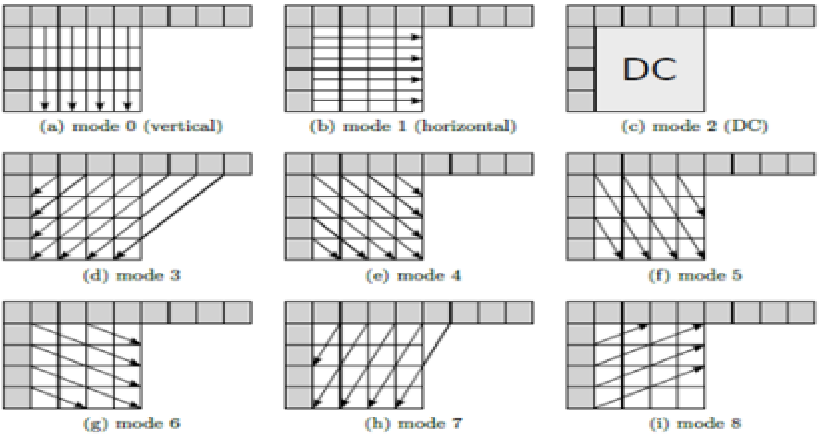

Figure 2.5: Nine intra-image prediction modes for 4×4 pixel luma blocks. In gray, the pixels of the adjacent blocks previously encoded.[19]

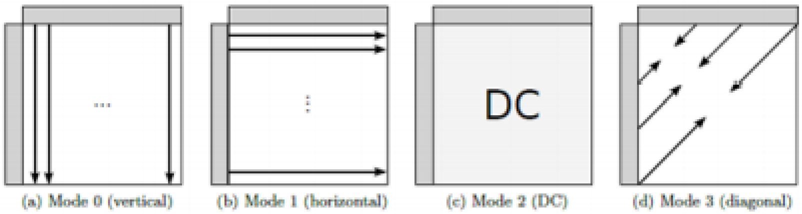

Figure 2.6: Four intra-image prediction modes for 16×16-pixel luma macroblocks. In gray, the pixels of adjacent blocks previously encoded.[19]

2.3.6 Inter-frame prediction:

Inter prediction is used for images that have a reference image. It is applied to P (prediction) and B (bi-prediction) images to eliminate redundancies between successive images. Inter prediction identifies the position of a MB in the current image relative to its position in the reference image. H.264/AVC supports a wide range of block sizes from 16×16 to 4×4. The component of the luminance of each macroblock can be split in four ways as shown in (Figure 2.7)

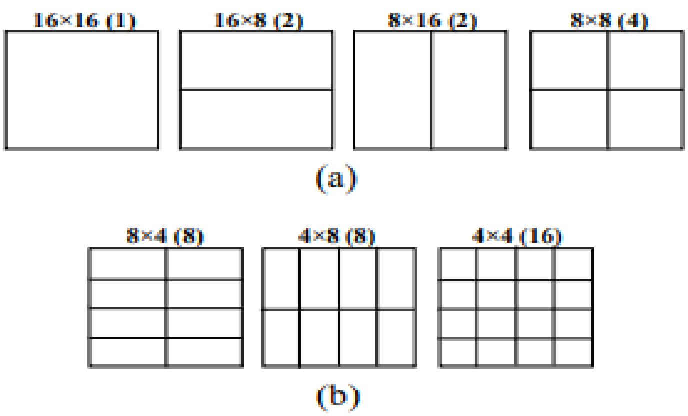

Figure 2.7: Macroblock decomposition

2.4 High Efficiency Video Coding (HEVC):

2.4.1 Definition of HEVC Standard:

High Efficiency Video Coding (HEVC) as a successor of the H264 standard was created to address and respond to all existing applications in AVC with more effective features, mainly, like compression performance. HEVC supports video with higher resolutions and improves parallel processing. The first edition of HEVC was released in 2013. This standard covers the distribution of high definition video e.g. UHD, 4K and 8K videos. HEVC offers more compression rates, better visual quality and lower bandwidth requirements compared to H.264.

In order to achieve these improvements, HEVC has adopted innovative tools such as: accurate intra/inter predictions, an adaptive sample loop offset filter and block partitioning based on a tree grid.

2.4.2 HEVC profiles: [18]

HEVC standard has 3 profiles:

- **The " Main " profile:**

Encoded video in 8 bits with chrominance sampling 4:2:0.

- **The profile " Maine 10 ":**

This profile allows a depth of 10 bits under 4:2:0 chrominance sub-sampling.

- **The "Main Still Picture" profile:**

This profile is dedicated to still picture coding and uses the same tools as those used to encode "intra" video images.

2.4.3 Detailed description of the HEVC standard[19]:

HEVC is a hybrid encoder which shares nearly similar architecture of the H264 standard.

Main differences of the compared encodes are regrouped in the table below:

Table 2.1: Principal tools H.264/AVC and HEVC

	H264/AVC	**H265/HEVC**
Coding unit	Macroblock 16×16	LCU 8×8 to 64×64
Partitioning	Sub-Blocks up to 4×4	Quarternary –Tree
Transformed	Full DCT 8×8 and 4×4	TU From 32×32 to 4×4 square and Rectangular
Prediction Intra	9 predictors max	35 maximum predictors
Prediction Inter	Direct mode	Merge mode
Reference	Multiple	Multiple
Entropy Encoder	CAVLC, CABAC	CABAC
Filters	Anti-block filter	Anti-block filter Adaptive Loop Filter (ALF) Filter Sample Adaptive Offset(SAO)
Specification	Resolution limited to 4K (4,096x2,304) Frequency level: 59.94 fps.	Resolution level up to 8K UHDTV (8192x4320) Frame rate up to 300 fps

2.4.4 HEVC global codec structure:

HEVC standard functions in the same way as H.264/AVC with the addition of new enhancements these enhancements include [26]:

- More flexible partitioning of the video frame.
- Greater flexibility in the prediction and partial block transformation modes.
- More sophisticated interpolation and filtering.
- Integration of parallel processing in the high-level profile.

All these improvements have resulted in a video coding standard that allows better compression, with a cost of higher coding and decoding complexity.
Steps performed via a Video Encoder are:

- Partition each frame into several units.
- Each unit predicted using intra/inter prediction.
- Transformation and quantification of residual block.
- Transformed and Quantized Coefficient entropy coding.
- Control the prediction data and filter it.

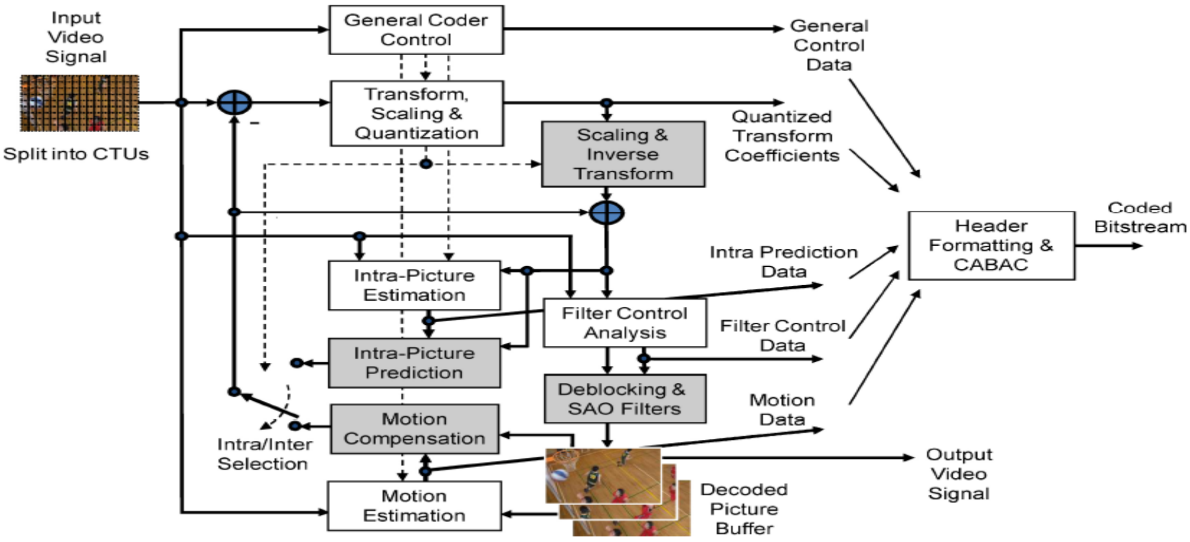

Figure 2.8: HEVC Structure Encoder [27]

The decoder inverts all the functions applied in the encoder to recover the original video.

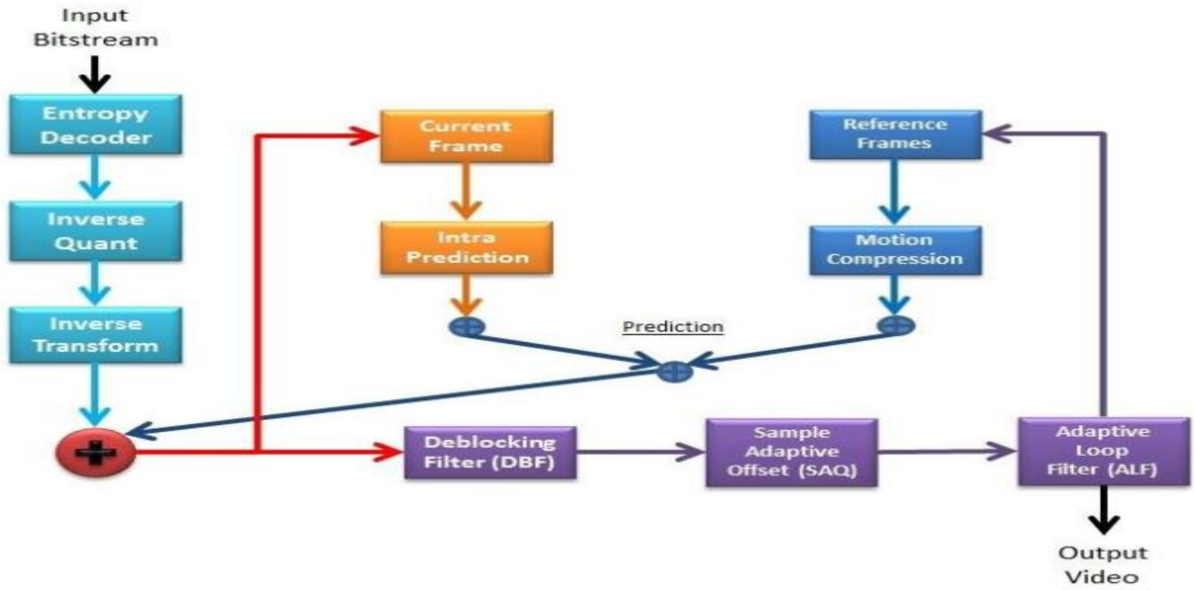

Figure 2.9: HEVC Structure Decoder [32]

2.4.5 HEVC Structure:

HEVC is a hybrid codec with temporal and spatial prediction. Each video is partitioned into frames and each frame is divided into small units called coding tree units (CTUs). Each CTU contains the CU coding units which can be divided into prediction units (PU) and (TU) transformation units. In addition, each CTU contains luma coding tree blocks (CTB) and chroma coding tree blocks (The luma component in any N × N rectangular area is covered by a luma CTB and chroma CTBs cover each area of N / 2 × N / 2 of each of the two chroma components. The values of N size vary between 64, 32 and 16 as specified by the syntax parameter encoded in the sequence parameter sets (SPS). Depending on the availability of encoder resources (processor speed and memory storage), the size of the CTBs will vary in

HEVC. As with H.264, the macroblock size is fixed at 16 × 16. The CTU form by a CTB luma and CTB chroma. Luma and chroma CTBs can be divided into multiple coding blocks (CBs). The tree structure is used to partition CTB units. This partitioning is done simultaneously for luma and chroma. [28]

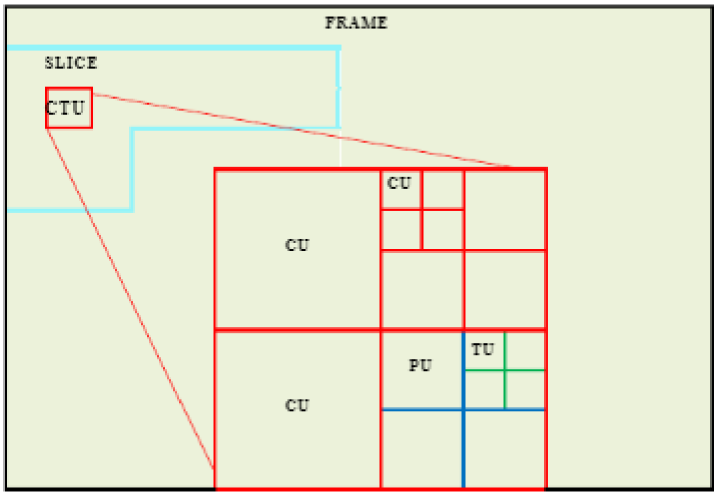

Figure 2.10: Frame partitioning [28]

- **Prediction Unit:**

Prediction Unit is a core component in coding unit (CU), that defines the size of possible partitioning for each level of the tree decomposition and changes according to the type of prediction (inter / intra) [19].

CU can be split into one, two or four PUs depending on the PU split type. The HEVC defines two forms of splitting for intra-coded Cu and 8 splitting shapes for inter-coded CU. In contrast to PU, there can only be one split PU splitting type is specified differently as shown in figure 2.11.

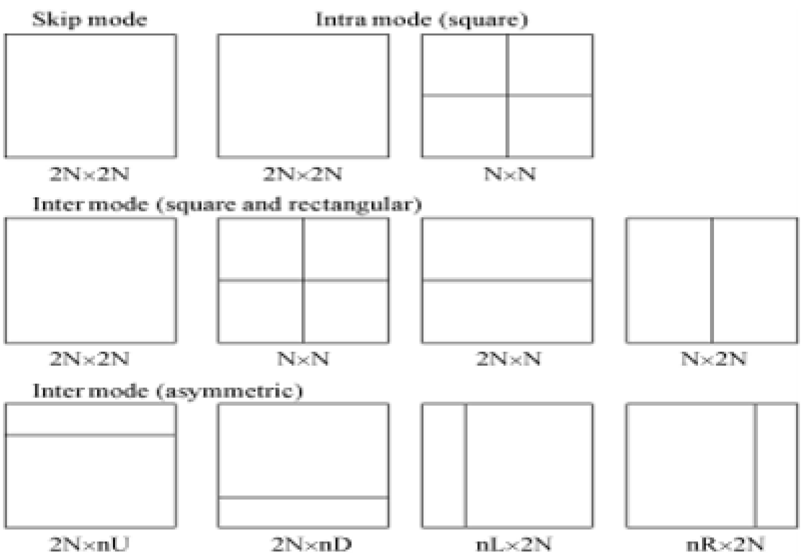

Figure 2.11: Illustration of PU splitting types in HEVC [30]

- **PU Splitting Type**:

Each CU in HEVC can be classified into three categories: skipped CU, inter coded CU, and intra coded CU. An inter-coded CU uses motion compensation scheme for the prediction of the current block, while an intra-coded CU uses neighboring reconstructed samples for the prediction. A skipped CU is a special form of inter-coded CU where both the motion vector difference and the residual energy are equal to zero. For each category, PU splitting type is specified differently as shown in Fig. 2.11 ; when CU size is equal to (2N×2N). Only part (−2N×2N) PU splitting type is allowed for the skipped CU. For the intra coded CU, two possible PU splitting types of (part –2N×2N) and (part −N×N) are supported.
Finally, total eight PU splitting types are defined as two square shapes (part −2N×2N), (part −N×N), two rectangular shapes (part −2N×N and part −N×2N), and four asymmetric shapes (part −2N×nU, part −2N×nD, part −nL×2N, and part −nR×2N) for inter coded CU [30].

- **Transformation Unit:**

The Transform Unit (TU) defines the size for the transform and quantization applied to a prediction unit. Three levels of decompositions are possible at most for this TU which takes sizes ranging from 4 × 4 to 32 × 32. For encoding a sequence, it is necessary to define the size of the largest LCU (Largest Coding Unit) and the partitioning depth of the CUs and TUs. The n partition sizes of the three units CU, PU and TU are then determined in a recursive during encoding [19].

2.4.6 HEVC Coding Tools [31]:

- **Intra prediction:**

Intra prediction has 35 direction modes (33 spatial directions in addition DC mode and planar mode), differently to H264 which only uses 8 directions.

All of these modes (35 modes) are not systematically enabled for the different sizes of PU which have 3 predictive indicators for 64 x 64 PU, 17 for 4 x 4 PU and 34 for the other sizes by PU. This multi-mode operation frequently captures repetitions of neighboring pixels at the expense of signal overload [19].

- **Inter prediction:**

HEVC standard defines CU as the basic encoding unit that replaces the MB used in H.264 / AVC, CU size ranges from 8 × 8 up to 64 × 64. Indeed, the fact of using wide CUs reduces the data necessary to define the movement, for high resolution videos for example this makes the video compression more efficient without having to lose its quality. A CU can be partitioned into several CUs following the partitioning in tree structure, the different levels of partitioning of the tree can go up to 4 levels, the CU at level 0 is called LCU. Finally, the prediction is made for CUs of sizes: 2N × 2N, 2N × N, N × 2N and N × N which constitutes the unit of prediction. Inter prediction revolves around two aspects, the computation of the prediction and the encoding of the motion vector. These two points have been improved with a particular effort on the encoding of the motion vector [19].

- **Quantization control**:

It is similar to H.264 / MPEG-4 AVC, where uniform reconstruction quantization (URQ) is used in HEVC, with quantization scaling matrices for different transform block sizes is supported.

- **Transform and entropy coding:**

There are 4 possible sizes of transformation: 32 x 32, 16 x16, 8 x 8 and 4 x 4, with separable transformation becoming from the integer coefficient (DCT). The codec (CABAC) has been improved and developed by HEVC and is more efficient in its compression compared to H264 CABAC.

- **Deblocking filter**:

The process of loop release filtering has been improved by simplifying the design. The simplification assists in decision support and filter processes, which makes it user friendly for parallel processing.
A sample of adaptive offset (SAO) is added in the inter-frame prediction loop after the deblock filter. This is a non-linear amplitude mapping scheme, which helps to reconstruct the original signal amplitudes using a look-up table [29].

2.4.7 HEVC stream:

HEVC bitstream contains a unit named "Network Abstraction Layer" (NAL), which is composed of a payload and header. Header section (NAL) consists of 5 bits type of NAL unit, 6 bits layer identifier known as ("nuh_layer_id") and 3 bits time sublayer identifier.

A new structure of the "Video Parameter Set "(VPS) has been added to HEVC meta data presentation to enable compatibility for extension from the standard and to include the dependency between time sublayers. The VPS shares all information and data needed for HEVC decoding. [2]

2.5 VVC Standard:

Versatile Video Coding (VVC) standard named H.266 is the latest emerging video coding standard by ITU-T and ISO/IEC. H.266 is developed by the Joint Video Experts Team (JVET) and published in July 2020.

The VVC was designed to achieve significantly improved compression capability over previous standards such as: HEVC.

VVC is estimated to deliver 30% to 50% of compression efficiency for the same video quality when compared to HEVC. Some key application fields for using VVC include in particular ultra-high-definition video (4K to 16K resolution), video with high dynamic range and wide colour gamut. It also works with 360° video omnidirectional content. VVC supports all the YUV formats, 4:4:4, 4:2:2 and 4:2:0 with 10 to 16 bits for each component. The important design considerations for VVC were low computational complexity on the decoder side and ease of use for parallelization at various algorithmic levels. It is expected that VVC will be available on the market very soon.

The new standard is expected to enable the provision of ultra-high definition (UHD) services at bit rates that are currently used to transport high definition TV (HDTV). Otherwise, the use of VVC would make it possible to store twice as much video content on a server or to send via a streaming service [34][35].

- **VVC codec structure:**

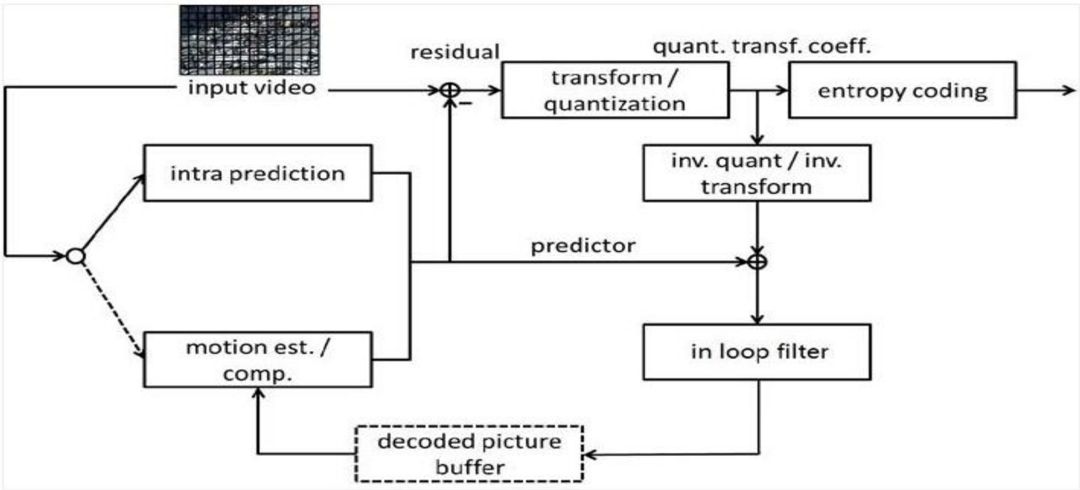

Figure 2.12: VVC codec structure [37]

VVC introduces serval improvements and new tools compared to the previous codecs HEVC:

- **Block Partitioning:**

VVC extends the concept of the CTU in HEVC. CTU can measure up to 128 × 128 pixels and is partitioned with a quad with nested multi-type tree diagram (QTMT). This allows a block to be divided into square, binary or horizontal and vertical ternary sub-units. This structure unifies the concepts of coding unit (CU), prediction unit (PU) and transform unit (TU) coding into CU. Such flexibility allows detailed modeling of video content.

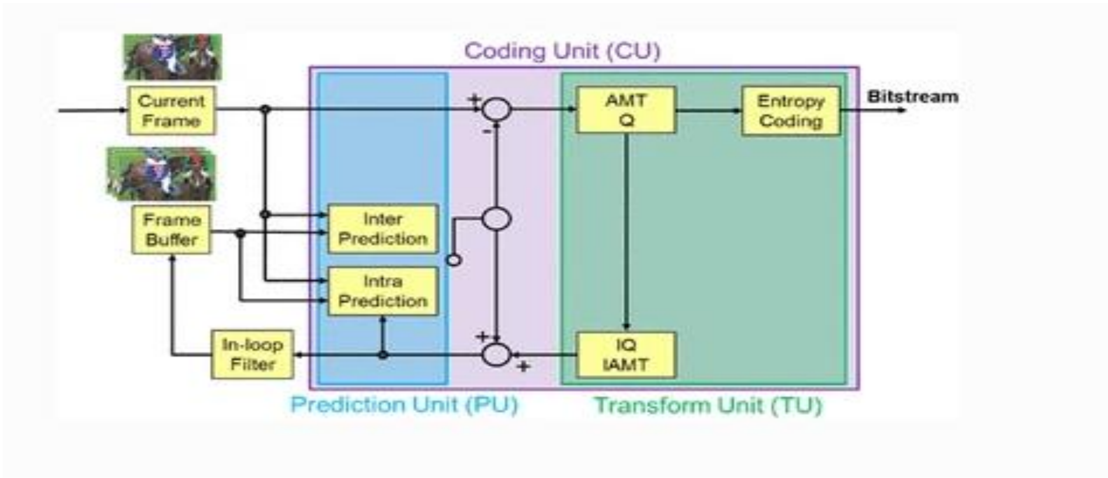

Figure 2.13: Coding structure of VVC [38]

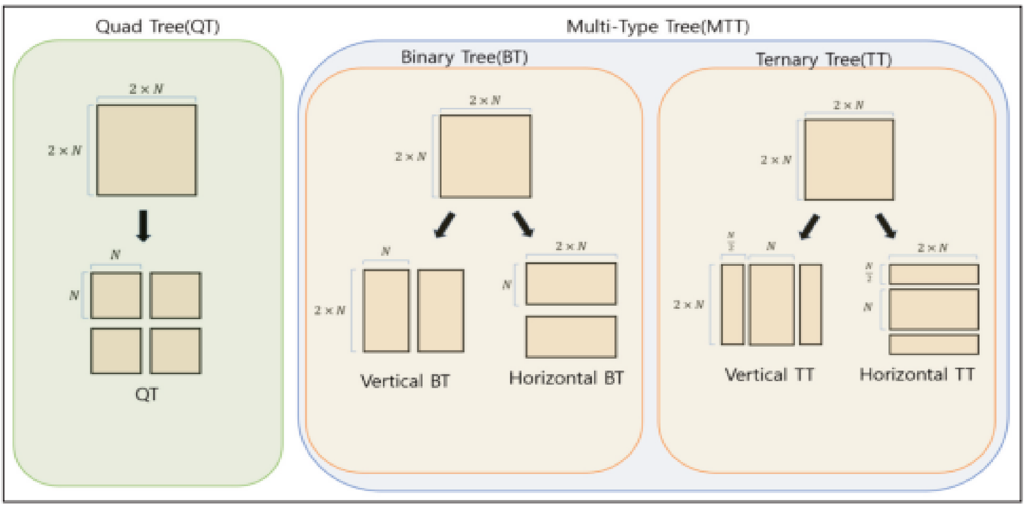

Figure 2.14: Division the bloc in quad-tree and multi-type-tree [39]

- **Intra prediction:**

VVC uses 67 intra modes, including 65 directional, DC and planar modes. Intra prediction is performed for CPUs 4 × 4 to 64 × 64 pixels to further reduce inter-component redundancy, VVC uses linear mode pattern prediction which predicts chrominance samples as a function of reconstructed luma samples. An additional improvement is that VVC extends the reference samples by allowing the use of multiple reference lines, which improves the quality of prediction.[36]

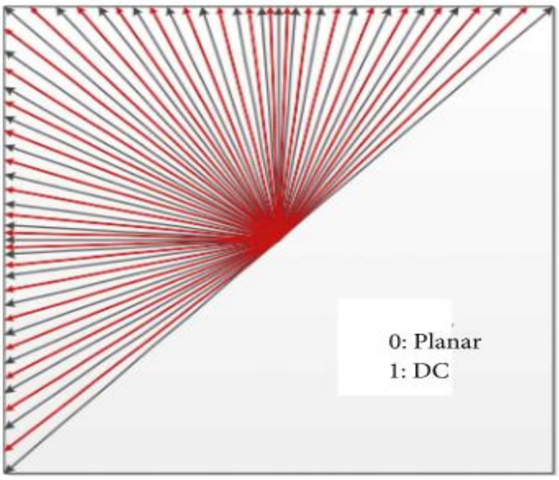

Figure 2.15: 67 intra prediction of VVC [40]

- **Inter prediction:**

VVC motion can be signaled by explicit transmission of motion parameters, via a skip mode or a merge mode which includes the derivation of motion parameters from spatial and temporal candidates. The merge mode uses spatial, temporal and zero motion vector (MV) candidates such as HEVC. A new schema called motion vector difference merge mode (MMVD) has been introduced that refines the motion derived via a motion vector difference (MVD). VVC uses affine motion compensation prediction. The motion is indicated by motion information at two or three control points from the corners of the block. At the decoder side, the motion for each slice is derived based on this information. [36]

- **Transformation unit:**

VVC uses up to 64 × 64 for luminance and 32 × 32 for chrominance samples which are appropriate for higher resolution samples. A multi-transform selection schema allows to choose best horizontal and vertical transform cores from different discrete cosine transform (DCT) and discrete sine transform (DST) cores. To exploit spatial redundancy further, the secondary transform is introduced that uses 4 × 4 and 8 × 8 inseparable transforms. The new scalar dependent quantization is used in which the set of allowable reconstruction values depends on previously reconstructed coefficients. [36]

- **Entropy encoding:**

VVC uses context enhanced binary arithmetic coding (CABAC). A new context model initialization is introduced depending on QP. [36]

- **In-Loop filters:**

The VVC system includes three looped filters. While the Release filter (DBF) and the adaptive sample offset (SAO) and the adaptive loop filter (ALF). ALF uses a block classification schema to select among 25 different filters, based on the direction and level of local gradients. ALF is applied after DBF and SAO, on blocks of 4 × 4 pixels [36].

2.6 Multiview video coding:

Multiview video coding (MVC), standardized within the Joint Video Team (JVT) of both the Video Coding Experts Group (VCEG) and the ISO / IEC Moving Picture Experts Group (MPEG), is considered the most important process in the technology of multiview and is available in most important video apps such as 3D TV and free video[42]. Simultaneous recording from a moving scene with a serval camera generates enormous amounts data that require efficient compression. The simplest solution is to code independently each video of each camera with a conventional video codec, e.g. H.264 / AVC, HEVC. However, this solution does not offer optimal compression efficiency because in MVV there is significant spatio-temporal correlation between the various views. In order to achieve a better compression efficiency, exploitation of these inter-view dependencies is indispensable, and this is the case of MVC codecs [43].

2.6.1 MVC/H264:

The MVC group of JVT chose the MVC method based on H.264 / AVC as the MVC reference model, because this method showed better coding efficiency than simultaneous H.264 / AVC coding and the other methods. Motion compensated video coding, which has several new features that significantly improve its performance and bit rate distortion. Main MVC features are mentioned below:

- Motion-compensated prediction of variable block size with block size up to 4 x 4 pixels;
- Motion vector accuracy of a quarter pixel;
- Multiple reference image for motion compensation;
- Bi-directional predicted image as reference for motion prediction.
- Intra-image prediction in the space domain;
- Adaptive release filter in the motion-compensated prediction loop;
- Small block size transformation (4x4 block transform);
- Improved entropy coding methods.[13]

The new MVC design for the H.264 standard is shown in Figure 2.16

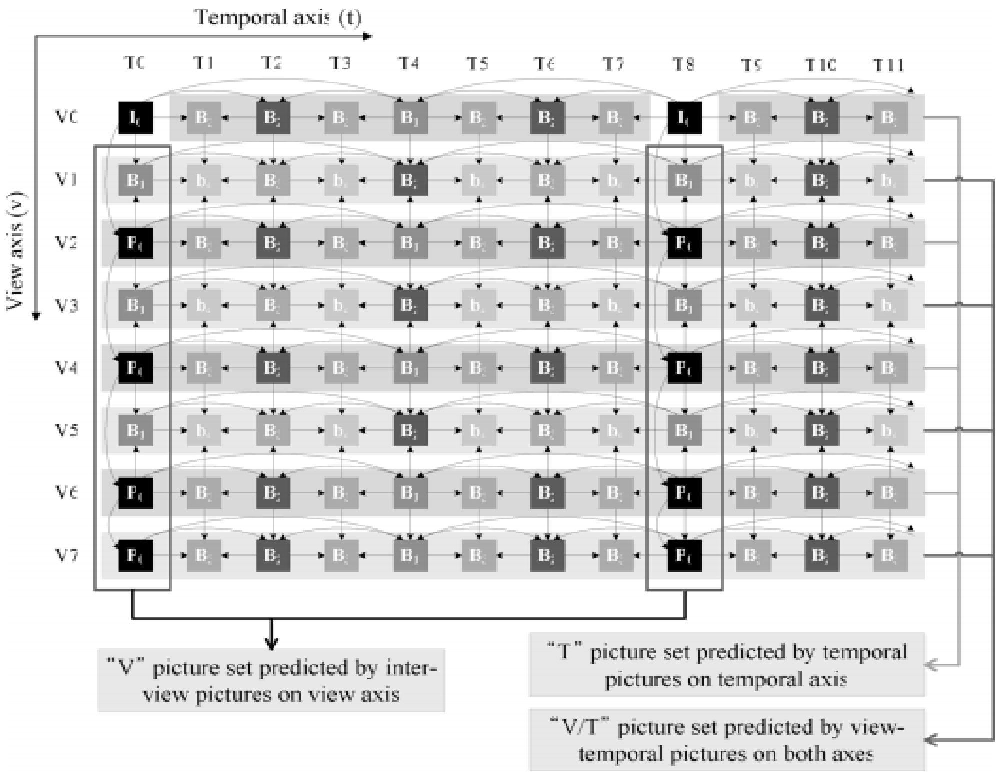

Figure 2.16: Structure of the H.264/MVC method [12]

Fig. 2.16 depicts an example of theH.264/AVC-based MVC structure, in which there are eight parallel views. This structure utilizes the hierarchical B pictures, which not only improves the coding efficiency, but also provides temporal scalability. This structure can be divided into three kinds of picture sets: the picture set predicted by the inter-view pictures on the view axis, the picture set predicted by the temporal pictures on the temporal axis, and the picture set predicted by the view-temporal (spatio-temporal) pictures on the view and temporal axes. In this structure, the I pictures are only used at random access points. The pictures on the temporal axis T0 are predicted spatially, the pictures on the view axes V0, V2, V4, and V6 are predicted temporally, and the pictures on the view axes V1, V3, V5, and V7 are predicted temporally and spatially.

There are other multi-view prediction structures in the literature based of modifications of the H264 AVC, such as structure design of the GOP group of groups of pictures (GGOP) prediction group using the shared reference picture memory, the prediction coder based on view interpolation and the structure using a multi-view layer depth image [43].

2.6.2 MV-HEVC:

A Call For Proposals [48] was launched in 2011 to solicit expert contributions for the development of new 3D video coding technology. The responses were good enough to facilitate the establishment of JCT-3V in july 2012. The main objective of JCT-3V was to develop a 3D video coding technology more advanced than ITU-T H.264 multi-view video coding extension (MVC). The team completed this work in June 2016, after having analysed and defined 3D and multi-view encoding extensions for ITU-T H.265 HEVC. The 3D-HEVC contains a specific design to compress 3D video with depth layers while MV-HEVC processes several textures layers of the 3D video format. MV-HEVC [49] at first was integrated to the second edition of HEVC [50] standard and finalized later in february 2015.

MV-HEVC is based on HEVC coding architecture, with new high level syntax (HLS) specific features that use multi-view and stereoscopic presentation. MV-HEVC and 3D-HEVC, use a multi-layer concept which are created from the inter-layer structure to obtain compression performance between different layers. Each layer can define the texture, format, depth and other auxiliary information related to a particular camera view. While all layers belonging to the same camera perspective are referred as a view, layers carrying the same type of information (for example, texture or depth) are generally also known as components for 3D video [2].

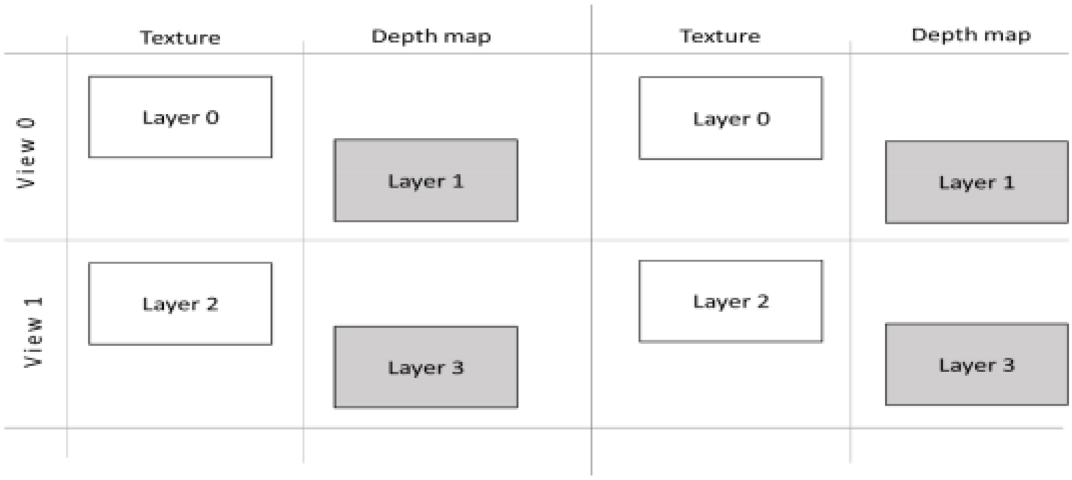

Figure 2.17: Layers division in MV-HEVC [2]

MV-HEVC includes high level syntax additions (HLS) and it can be implemented using existing single-layer 2D decoding cores. Moreover, MV-HEVC also shares the same HLS with all multilayer HEVC extensions. HLS permits the extraction of a unique texture base view from the MV-HEVC bitstream which is decodable by the main profile HEVC decoder [2].

2.7 Conclusion:

This chapter focuses on video coding concepts by providing an overview of the H.264 video coding standard, its architecture, and its key features. An overview of the concepts of HEVC coding was presented in this chapter with a special focusing on the MV-HEVC extended multi-view profiles. In addition, the latest video codec VVC / H.266 was detailed.

Chapter 03

Experimental evaluation

3.1 Introduction:

In this chapter, the performance of both MV-HEVC and VVC is compared and evaluated in terms of PSNR (dB) and bitrate (kbps) over several QP values. Three different video sequences have been used in the experiments.

3.2 Evaluation metrics, Platform, and test conditions:

Table 3.1 describes the used multiview video sequences and their parameters. In addition, samples of the tested sequences are shown in Figure 3.1.

Table 3.1: MVV sequences for compression frequency evaluation.

Database	Video sequences	Frame rate	Image resolution
MERL	Vassar	25	640 × 480
Fujii Lab	Kendo	30	1024×768
Fujii Lab	Balloon	30	1024×768

Figure 3.1: Samples of the used sequences

Table 3.2 shows the common initial configuration that was used to provide a fair comparison. A total of 8 successive images are coded for each sequence used. Because of the complexity of

VVC codec, as when we have tested for the first time it needed about 5 to 6 days to encode 250 frames for one QP.

The GOP size is eight with insertion of the intra-coded frames (I) at the end of each GOP. Four QP values were chosen are described in table 3.2.

Table 3.2: Initial common encoding configuration.

Frame to be encoded	8
GOP size	8
Intra period	8
Quantization parameter	[25,30,35,40]

The tests were conducted using the HM 16.2 codecs, which includes a multiview profile extension, that was used for MV-HEVC. VTM 2.0 codec was used as software platform for the VVC standard.

It should be mentioned that the software models used are developed using the C++ programming language and are designed for research purposes and not for commercial applications. All simulations were performed on a PC with intel core i5 2.60 GHz CPU and 4 GB RAM.

3.3. Results and discussion:

The figure 3.2 and figure 3.3 represent the performance of the MV-HEVC codec over 250 frames using two video sequences for each test. The figures show that the video quality increases with the rise of bit rate values.

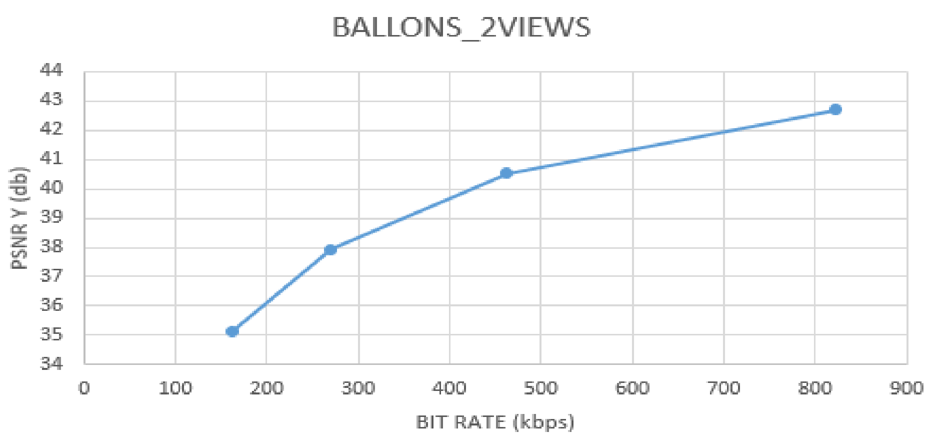

Figure 3.2 Compression performance using multiview video sequences

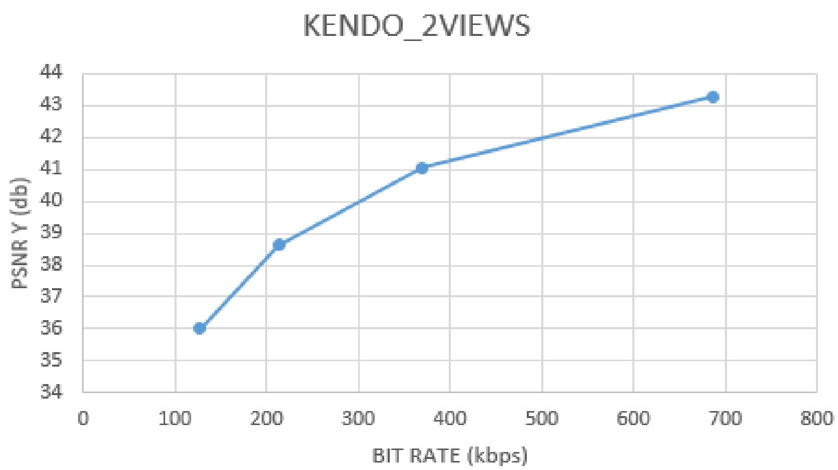

Figure 3.3 Compression performance using multiview video sequences

As it was expected, results in Figure 3.4, Figure 3.5 and Figure 3.6 obviously show that VVC exceeds MV-HEVC in terms of bitrate saving and video quality. This outperformance ultimately covers all the carried-out simulations through the different QP values and the various MVV sequences. The rate distortion curves of the HD MVV sequences, illustrated in Figure 3.4 and Figure 3.5, prove that VVC codec provides much better compression performance compared to MV-HEVC over the entire bitrate range. For example, when QP=35, the bitrate saving gain of VVC exceeds 18 % and 34 % for Balloon and Vassar sequences, respectively. Moreover, Figure 3.6 reveals that further bitrate saving gains were achieved by VVC for the standard definition MV-HEVC sequences, whereby a gain of 17% is marked for Kendo sequence.

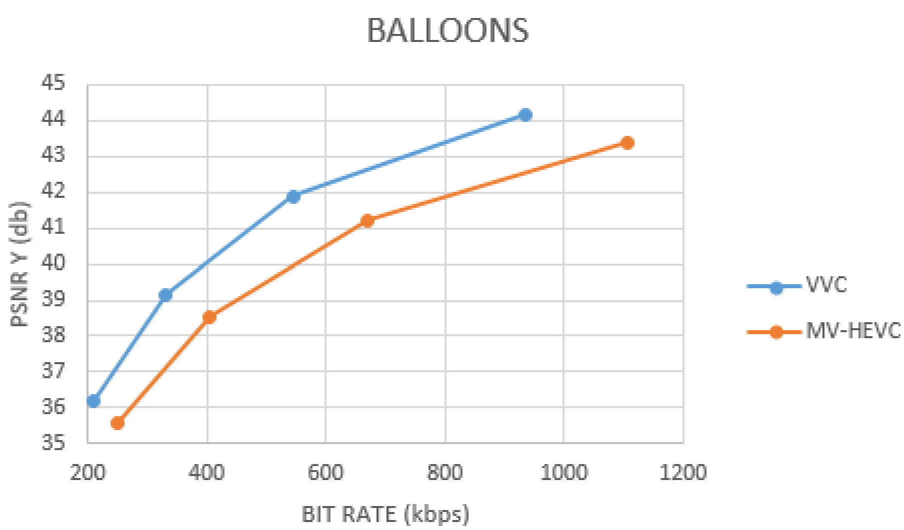

Figure 3.4 Compression performance using multiview video sequences

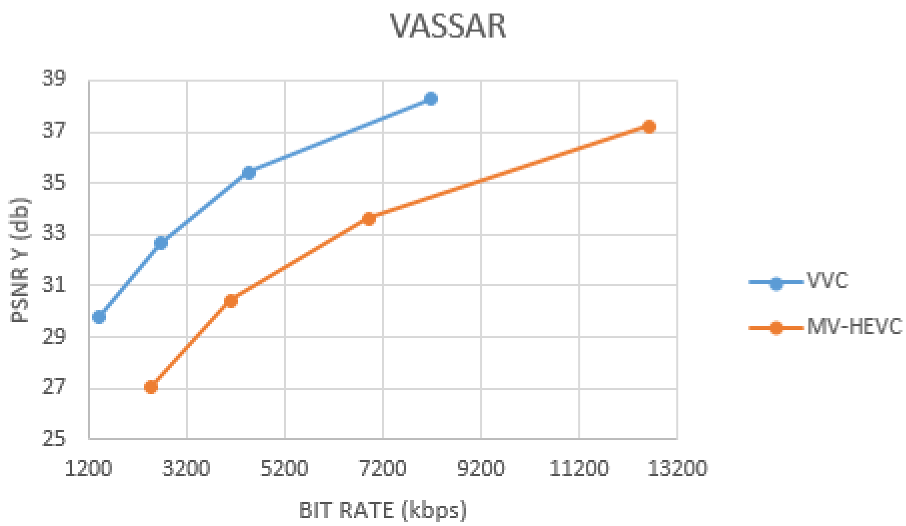

Figure 3.5: Compression performance using multiview video sequences

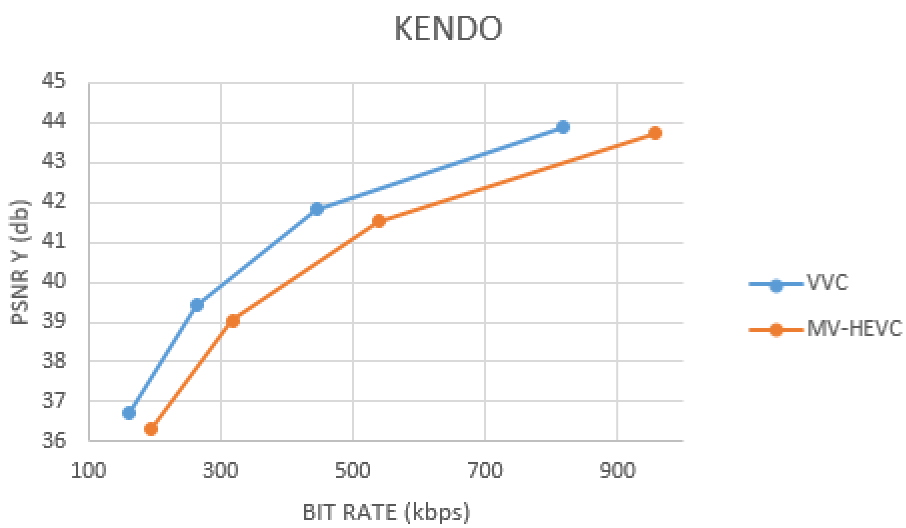

Figure 3.6: Compression performance using multiview video sequences.

Tables 3.3 and 3.4 represent the time needed to encode 8 frames from Balloon, Kendo and Vassar sequences, respectively.

Table 3.3: Time comparison for Balloon

QP	VCC	MV-HEVC
25	8611.879 (sec)	2731.621 (sec)
30	4575.547 (sec)	2458.636 (sec)
35	2733.795 (sec)	2272.836 (sec)
40	2170.572 (sec)	1781.689 (sec)

Table 3.4: Time comparison for Kendo

QP	VCC	MV-HEVC
25	10210.24 (sec)	7029.055 (sec)
30	5826.535 (sec)	3880.963 (sec)
35	3832.184 (sec)	3685.884 (sec)
40	2772.167 (sec)	2577.063 (sec)

Table 3.5: Time comparison for Vassar

QP	VCC	MV-HEVC
25	39581.7 (sec)	7730.82 (sec)
30	29760.63 (sec)	6982.465 (sec)
35	25083.54 (sec)	6063.337 (sec)
40	19613.94 (sec)	5094.833 (sec)

Figure 3.7, Figure 3.8 and Figure 3.9 reveals that MV-HEVC exceeds VVC in term of time saving. For example, when QP=30 the time saving of MV-HEVC exceeds 53 % and 23 % for Balloon and Vassar sequences, respectively. This performance is due to the complexity of the VVC codec that needs more time to encode the video and provide higher quality and less bit rate.

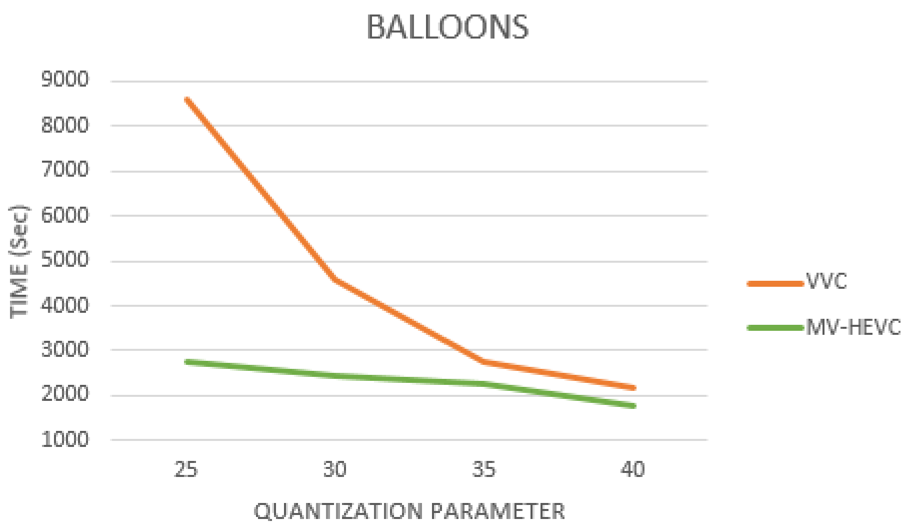

Figure 3.7 Compression performance using multiview video sequences

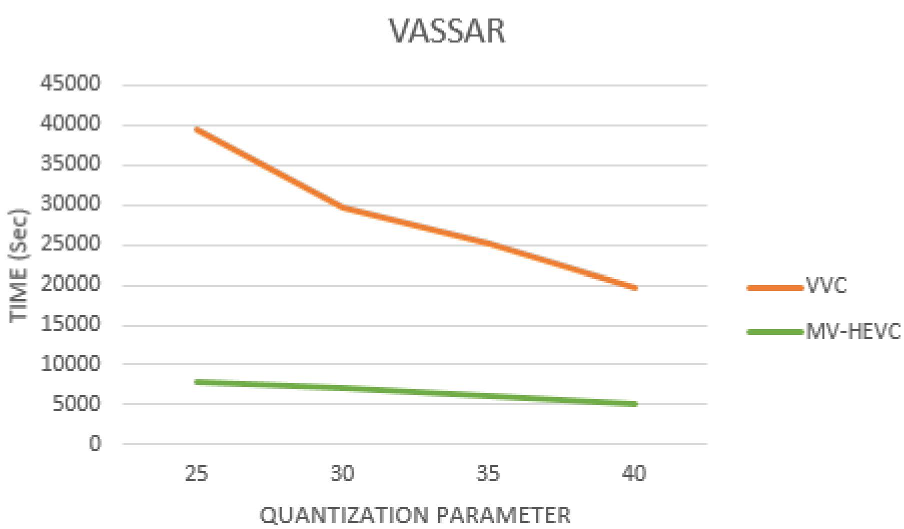

Figure 3.8 Compression performance using multiview video sequences

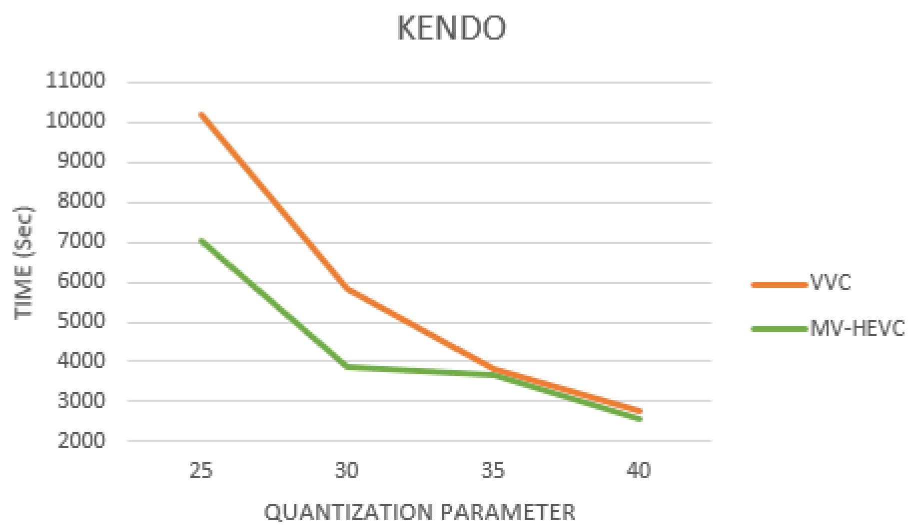

Figure 3.9 Compression performance using multiview video sequences

3.4. Conclusion:

The two codecs were simulated and evaluated using common test conditions and different MVV sequences. The results showed that the expected performance of VVC compared to MV-HEVC was superior in terms of compression efficiency. Substantial gains in bit-rate savings started from 17% for Kendo sequences and reached 34% for Balloons sequences. In addition, in terms of time MV-HEVC exceeds VVC with 53 % for Balloon when QP=30.

Conclusion

and Future perspectives

Conclusion and Future perspectives:

Researchers and institutions are working together to improve video coding techniques for the rising and divers visual quality. Multiview video technology represents a rich visual experience that offers viewers depth perception as well as free viewpoint navigation for certain applications. This Master dissertation reviewed MVV technology and its coding theory and concepts, focusing on MVC and MV-HEVC coding standards. Although no multilayers extensions have been published yet, VVC coding was also presented and tested in thesis.

The MV-HEVC uses tools of HEVC codec such as the innovative block partitioning to improve the rate distortion capability. MV-HEVC has been implemented and evaluated through different datasets and common test conditions. The used test video sequences were two texture views without depth map of SD and HD resolutions. VVC is based on a coding architecture similar to HEVC. It was developed by bringing enhancements to (HEVC) coding tools, and by adding coding tools for the sake of increasing the compression performance for a variety of video contents. VVC software codec was tested in this research project and compared to MV-HEVC. Despite the fact that MV-HEVC employs interview prediction coding, VVC outperforms MV-HEVC in terms of compression performance in all the reported cases. However, obtained results reveals the higher complexity produced by VVC compared to MV-HEVC.

Main propositions for future perspective are described as follows:

- Only two video resolutions have been tested in this dissertation (640x480 and 1024x768). We suggest addressing different video sequences with different resolution including 4K UHD and 8K UHD multiview video sequences, and 360 video sequences.

- Error resilience is an important feature to consider in future research. We propose to evaluate the resulted bitstreams of both MV-HEVC and VCC coding over error-prone networks of different topologies and conditions.

- Machine learning classification methods are suggested to be employed for future Multiview based coding schemes to improve further the compression performance and decreases the encoding complexity.

Bibliography

[1] Cisco, Visual Networking Index: Forecast and Methodology, 2016–2021,2017.

[2] Seif Allah El Mesloul Nasri "Multiview coding and compression for 3D video", Doctorate dissertation ,University Badji Mokhtar Annaba ,2018.

[3] T. Okoshi, "Three-Dimensional Imaging Techniques", Academic Press, UK, 1976.

[4] Y. Kim et al., "Three-dimensional display technologies of recent interest: principles, status, and issues," Applied optics, 50(34), H87–115, 2011.

[5] Urey. H et al., "State of the art in stereoscopic and autostereoscopic displays," Proceedings of the IEEE, 99(4):540–555, 2011

[6] Web site : https://en.wikipedia.org/wiki/Stereoscopy

[7] Urey, Hakan, et al. "State of the art in stereoscopic and autostereoscopic displays." Proceedings of the IEEE 99.4 (2011): 540-555.

[8] SPOTLIGHT, "Charles Wheatstone: the father of 3D and virtual reality technology", King's College London, 2016. [Online]. Available: spotlight.kcl.ac.uk.

[9] Web site : https://hevc.hhi.fraunhofer.de/mvhevc

[10] Web site : https://jvet.hhi.fraunhofer.de/svn/svn_VVCSoftware_VTM/

[11] Zatt, Bruno, et al. 3D Video Coding for Embedded Devices. Springer, 2013.

[12] Hur, Jae-Ho, Sukhee Cho, and Yung-Lyul Lee. "Adaptive local illumination change compensation method for H. 264/AVC-based multiview video coding." IEEE Transactions on Circuits and Systems for Video Technology 17.11 (2007): 1496-1505.

[13] Nemčić, Ognjen, Snježana Rimac-Drlje, and Mario Vranješ. "Multiview video coding extension of the H. 264/AVC standard." Proceedings ELMAR-2010. IEEE, 2010.

[14] Jarina, Roman, et al. "MPEG-1 bitstreams processing for audio content analysis." (2002).

[15] "Difference between 2D and 3D" ,15 February 2016.

[16] Adélaïde Blavier "Impact des images en 2D ou 3D sur les processus cognitifs impliqués dans le traitement visuel et dans le contrôle de l'action :le cas de la chirurgie minimale invasive "(Doctoral Dissertation , University of Liège),2006.

[17] Marco Cagnazzo, "The H.264 and HEVC standards", institut Mines-Telecom,.n.d

[18] web site https://fr.wikipedia.org/wiki/H.265/HEVC

[19] Yahi Amira, "Implantation de la chaîne inter-prédiction utilisée dans l'encodeur H.264/AVC sur une plateforme multi-composants",(Doctoral Dissertation , University of BADJI Moukhtar ANNABA),2016.

[20] "The Register - Sharp's 3D LCD: how's that work, then?". www.theregister.co.uk. 2004. Retrieved 2008-01-25.

[21] Hardesty, L. "Better glasses-free 3-D: a fundamentally new approach to glasses-free 3-D displays could save power, widen the viewing angle and make 3-D illusions more realistic [internet]. Massachusetts: MIT News Office; 2011.[cited 2017 Feb 20]."

[22] Urey. H et al., “State of the art in stereoscopic and autostereoscopic displays,” Proceedings of the IEEE, 99(4):540–555, 2011

[23] Web site : https://en.wikipedia.org/wiki/Multiview_Video_Coding

[24] Ho, Yo-Sung, and Kwan-Jung Oh. "Overview of multi-view video coding." 2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. IEEE, 2007.

[25] Anthony Vetro, Thomas Wiegand and Gary J. Sullivan ” Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard” Fellow IEEE PROCEEDINGS OF THE IEEE 2011.

[26]Web site https://fr.slideshare.net/MARSSALAH/master-thesis-72331022

[27] Elrowayati, Ali ,Abdullah, Mohammad Faiz Liew” Survey and Analysis of Digital Watermarking Algorithms on HEVC Video Coding Standard” 2014/12/25

[28] Saleh, Mohammed A., et al. "Review for high efficiency video coding (HEVC)." 2014 IEEE Conference on Systems, Process and Control (ICSPC 2014). IEEE, 2014.

[29] KASHYAP KAMMACHI SREEDHAR, MULTIVIEW VIDEO CODING FOR VIRTUAL REALITY;(Master Dissertation , Tampere University of Technology), April 2016.

[30] Kim, Il-Koo et al. “Block Partitioning Structure in the HEVC Standard.” IEEE Transactions on Circuits and Systems for Video Technology 22 (2012): 1697-1706.

[31] G. J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Circuits and Systems Magazine. 22(12), 1649 – 1668, Dec. 2012.

[32] Psannis, Kostas E., Christos Stergiou, and Brij Bhooshan Gupta. "Advanced media-based smart big data on intelligent cloud systems." IEEE Transactions on Sustainable Computing 4.1 (2018): 77-87.

[33] Tang, Qiang. Computationally efficient techniques for H. 264/AVC transcoding applications. Diss. University of British Columbia, 2010.

[34] “Versatile Video Coding – Algorithms and Specification” by speakers Mathias Wien, Benjamin Bross.

[35] Subbarayappa, Shreyanka, and K. R. Rao. "Overview and Extensions of the High Efficiency Video Coding (HEVC) and Beyond (Versatile Video Coding)."

[36] Pakdaman, Farhad, et al. "Complexity Analysis Of Next-Generation VVC Encoding and Decoding." arXiv preprint arXiv:2005.10801 (2020).

[37] Web site : https://www.programmersought.com/article/25173662549/ .

[38] Lin, Ting-Lan, et al. "Fast intra coding unit partition decision in H. 266/FVC based on spatial features." Journal of Real-Time Image Processing 17.3 (2020): 493-510.

[39] 남다윤, and 한종기. "Versatile Video Codec 의 Picture 분할구조 및 부호화 단위." 방송과 미디어 24.4 (2019): 26-38.

[40] Zhao, Jinchao, Yihan Wang, and Qiuwen Zhang. "Adaptive CU Split Decision Based on Deep Learning and Multifeature Fusion for H. 266/VVC." Scientific Programming 2020 (2020).

[41] Wiegand, Thomas, et al. "Overview of the H. 264/AVC video coding standard." IEEE Transactions on circuits and systems for video technology 13.7 (2003): 560-576.

[42] Müller, Karsten, et al. "3D high-efficiency video coding for multi-view video and depth data." IEEE transactions on image processing 22.9 (2013): 3366-3378.

[43] Book :George A.Tsihrintzis and Lakhmi C.jain eds. "Multimedia Services in intelligent Environment"Advanced Tools and Systemes , springer-verlag Berlin Heidelberger,2008.

[44] Dinstein et al., "On the com-pression of stereo images: Preliminary results", Signal Processing: Image Communications, (17)4, 373-382, Aug. 1989.

[45] M. G. Perkins, "Data compression of stereo pairs", IEEE Transactions on Communications, 40(4), 684-696, April 1992.

[46] Carballeira, Pablo, et al. "FVV Live: A real-time free-viewpoint video system with consumer electronics hardware." arXiv preprint arXiv:2007.00558 (2020).

[47] “ Overview: image and Video Coding Standards “ Bernd Girod: EE398B Image Communication II Video Coding Standards: MPEG-1,2,4 no. 1

[48] MPEG Video and Requirements Group, Call for Proposals on 3D Video Coding Technology, document N12036, Geneva, Switzerland, Mar. 2011

[49] G. Tech et al., “MV-HEVC DraftText9,” document JCT3V-I1002, Sapporo, Japan, Jul. 2014.

[50] High Efficiency Video Coding, document Rec. ITU-T H.265, Oct. 2014.