POLITECNICO DI TORINO

Master’s degree in cinema and media engineering

Master’s degree thesis

Evolution of video codecs: a comparative analysis of HEVC, VVC, AV1, and AVM

Supervisor:

Prof. Maurizio Martina

Dr. Alessandra Mosca

Dr. Giuseppe Nicastro

Candidate:

Alessia Casella

In collaboration with

April 2025

Si sapis, sis apis.

ABSTRACT

In recent decades, the consumption of video content has increased exponentially, accounting for over 80% of global Internet traffic. This growth has come with a continuous evolution of video compression technologies, necessary to guarantee the efficient transmission of content across the multiple digital platforms and devices available today. Video compression actually plays a crucial role in reducing the amount of data required to represent visual content, balancing perceived quality, encoding efficiency, and computational complexity.

This study's aim is to analyze and compare four of the latest video codecs: HEVC (High Efficiency Video Coding), VVC (Versatile Video Coding), AV1, and AV2 (AOMedia Video 1 and 2). These video coding technologies and standards represent the state of the art in video compression and are used in a wide range of applications, including online streaming, television broadcasting, video conferencing, and multimedia content storage. Particular attention is given to AVM, a newly introduced technology, designed to further enhance compression efficiency and address the ever-evolving market demands.

The study is based on an experimental approach, which involves encoding and decoding video sequences under controlled conditions, using dedicated software tools and predefined test parameters.
The comparative analysis focuses on objective quality metrics, including PSNR (Peak Signal-to-Noise Ratio) and VMAF (Video Multi-Method Assessment Fusion), to assess the impact of compression on perceived quality. Additionally, the comparison takes into consideration aspects such as bitrate efficiency or computational complexity, which may be considered as key factors in determining the adoption of a codec in different application contexts.

Index

Introduction

Today's society is becoming more automated, data-oriented and multimedia centric day by day. The amount of data sent among users every day is uncountable, but more than 80% [1] of the Internet traffic is occupied by videos as a media format. Video has become the most used source of data consumed globally [2], due to the fact that it is so important for human beings to shoot, record, store and view their video/images files. This rise in media consumption is also enabled by the over-the-top video streaming services such as Netflix, YouTube or Twitch.

The latest video coding standards, like the ones developed by the MPEG Group jointly with ITU-T, such as AVC, HEVC or VVC had a huge impact on the general concept of video coding. The aim was to make its fruition easier for everyday use and trying to figure out the best video compression, in conjunction with the great growth of video consumption.

Just to have a general view on how internet runs:

- Every second, 1 million minutes of video content pass through the internet. [3]
- 78% of people watch online videos every week, and 55% watch online videos every day. [4]
- YouTube is the second most visited website in the world. [5]
- More than 1 billion hours of video are watched on YouTube every day. [6]
- Video content accounts for almost 50% of all mobile network traffic. [7]
- It is estimated that by 2022 online video will account for more than 82 percent of all Internet traffic, almost 15 times more than in 2017.[3]

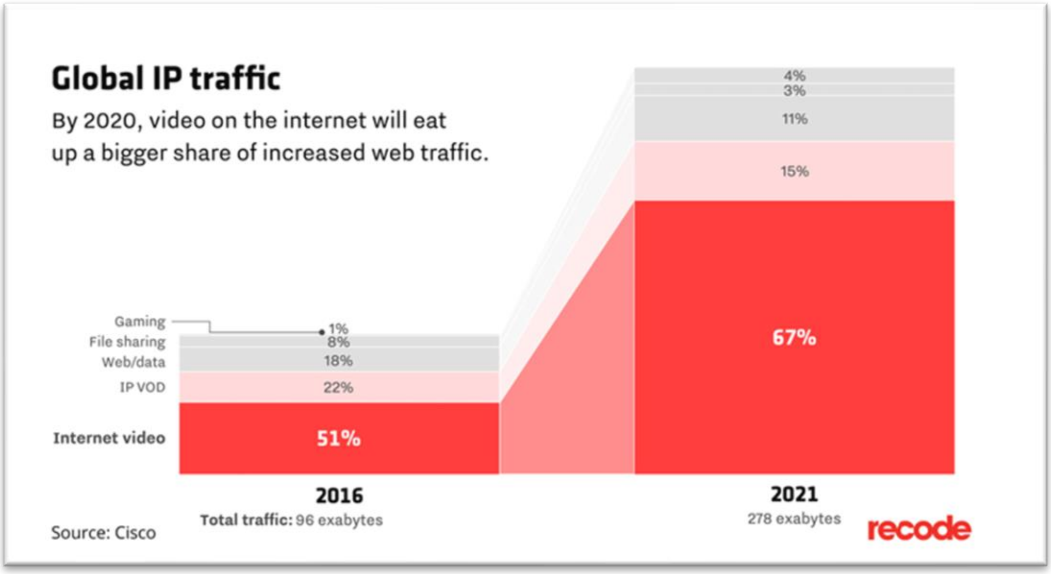


Figure 1: General view of the increase of video streaming in recent years.

The switch from analog to digital carried out a lot of advantages: better content quality, more reliability in transmissions and more flexibility at application level. Unfortunately, at first it was almost impossible for digital to compete with the analogic field, since the volume of data generated by the digitization process was excessively large and incompatible with storage and transmission systems. That's why, starting from the 80s, the first video compression formats were implemented. So, video compression formats therefore play a key role in the distribution of video content in broadcasts. [8]

Video compression technologies and standards have played a pivotal role in forming the modern digital ecosystem, enabling the efficient transmission, storage, and consumption of video content. The main goal of video compression is to reduce the data required to represent visual content while maintaining an acceptable level of quality for viewers. This has always been achieved by exploiting features such as redundancies within video data, similarities between frames or patterns within a single frame, to guarantee the optimization of the storage and bandwidth utilization. Without any compression, storage and transmission of digital video content would have been unrealistic, especially with the growing demand for higher definition and immersive video experiences. These standards have definitely evolved to deal with these challenges, making sure that video content can be efficiently passed across multiple platforms, from the more traditional (like broadcasts) to modern streaming ones. These standards are developed thanks to the collaboration of several international organizations, such as ISO/IEC MPEG, ITU-T or Alliance for Open Media, which aim to establish globally identified solutions for video encoding.

Today, video compression can't be defined only as a technical necessity; it is the cornerstone of the digital world. It supports diverse applications, from the daily used ones, like online streaming and videoconferencing, to emerging technologies like virtual reality and machine-driven video analytics. Furthermore, it keeps on evolving, replying to technological advancements and user demands, keeping up with improving efficiency and enabling new use cases.
These continuing advances mean that video compression standards remain critical to the increasingly video-centric nature of modern communication and entertainment.

The primary object of this thesis is to conduct a comparative analysis of four prominent video codecs: AV1, AVM, HEVC, VVC. These codecs represent the forefront of video compression technology and they are developed by different organizations with distinct goals and philosophies.

This work aims to achieve the following specific goals:

1. **Understand the Video Coding Technologies and Standards**
   Analyze the technical and organizational processes behind the development of video coding standards, highlighting the contributions of MPEG, ISO/ITU-T, and the Alliance for Open Media (AOMedia) to the field.
2. **Evaluate Codec Performance**
   Benchmark the codecs based on key performance metrics such as compression efficiency, computational complexity, and visual quality across various scenarios and use cases.
3. **Identify Strengths and Weaknesses**
   Provide an in-depth comparison to identify each codec's strengths, weaknesses, and potential areas of application, considering both technical and economic factors.
4. **Forecast Future Trends**
   Explore the role of these codecs in shaping the future of media distribution,

including their adoption, interoperability, and impact on emerging technologies like streaming services and immersive media.

By utilizing Shell scripting and codec-specific tools, this thesis ensures a reproducible and efficient methodology for performance evaluation.
This approach not only facilitates accurate benchmarking but also highlights the practical challenges and considerations involved in real-world implementations of these codecs. Ultimately, the findings aim to contribute to a deeper understanding of the trade-offs involved in codec selection.

Video codecs hints

The representation of multimedia content in digital uncompressed format requires a large amount of data. Despite the continuous increase in storage and transmission capacities, without adopting appropriate compression strategies, the digitization process would end up generating a data volume that is excessively high and incompatible with most modern storage and transmission systems. Video encoding, or compression, is the process responsible for converting a video in digital format into a numerical stream of smaller size, suitable for storage and transmission.

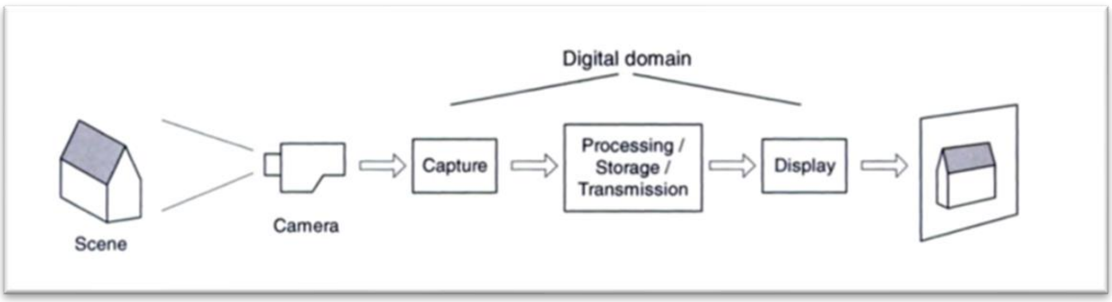


Figure 2: General video codecs' path.

Generally, the term *video codec* (enCOder/DECoder) is used to describe a software that implements both encoding (encoder) and decoding (decoder) of a digital sequence. The more efficient the codec is at compressing and transmitting the video sequence data, the better the received quality will be. On the receiving side, the codec decompresses the data and displays it on the screen.

An important part of modern digital multimedia systems is a powerful video codec. Its capability to compress video data with the least possible quality degradation is crucial in a world where high-resolution video content, such as 4K and 8K formats, is becoming increasingly popular. When a video codec has been developed well, this means that the video data can be delivered through a network with low bandwidth, with examples being the ones used in internet streaming or video conferencing, while at the same time, the end user gets an acceptable visual experience. In addition, it allows for effective storing of large video libraries, making it possible to save huge amounts of video content on modern storage systems. The problem of storage and transmission systems would become unsolvable in the absence of video codecs merely due to the vast amount of uncompressed video data. Without video codecs, the sheer size of uncompressed video data would pose insurmountable challenges for storage and transmission systems. Most current video codecs comply to international standards. This thesis will explore what a video coding standard is, why standardization is necessary, and the standardization process itself.

Chapter 1: Video coding standards

1.1 What a standard is

Everything mostly used in daily routines has been previously standardized.

How can be best described what a standard is? In very simple words:

"*Think of them as a formula that describes the best way of doing something. It could be about making a product, managing a process, delivering a service or supplying materials – standards cover a huge range of activities.*" [9]

When a standard is interfaced, several documents are generally handled, which were formerly roughed out and authorized according to processes by a precise entity.

More in details, a video coding standard specifies the structure of compressed video data (the bitstream) and how to decode it. Standardization ensures interoperability between video encoders and decoders from different vendors across various applications. While providing flexibility and encouraging innovation, standards focus on the core decoding process without specifying all implementation details. It outlines the output an encoder should generate, not how the encoder itself works. Standards typically define a set of compression tools a decoder must implement to be compliant with the specific standard.

The widely used H.26x and MPEG-x series of standards are developed and published by international standardization bodies like the ITU-T and ISO/IEC.

A standardization body can be both a legal or administrative entity with determined tasks and composition. There is not a single way to describe a standardization body, as per previous given definition, since they can be an organization, an industrial consortium, an authority, a company or even a foundation.

Figures 1.1 and 1.2: Example of organizations that can be annotated as standardization bodies.

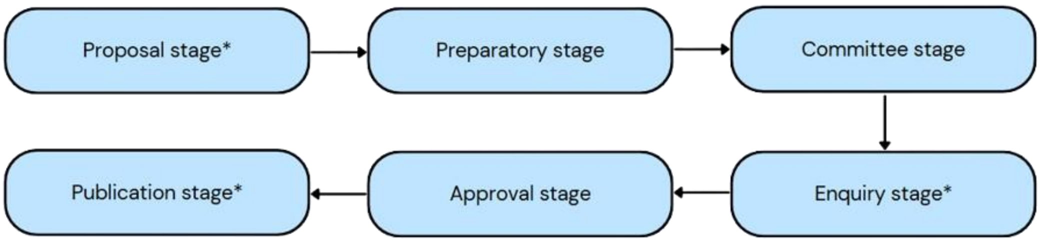
1.2 International Organization for Standardization

One of the most relevant international bodies for standardization is ISO (International Organization for Standardization), which was first created in 1947 in London, and it's currently located in Genève. ISO is an independent and non-governmental organization made up of members belonging to the standardization bodies of 172 countries.
ISO members are the leading standards organizations in their respective countries, with one member per country. Each member represents ISO within its own nation.
There are three member categories, each with varying levels of access and influence. This structure helps ensure inclusivity while also acknowledging the different needs and capabilities of each national standards body:

- **Full members (or member bodies)**: set the course for ISO standards development and strategy by participating and voting in ISO technical and policy meetings, sell and adopt ISO standards nationally.
- **Correspondent members**: attend ISO technical and policy meetings as observers, sell and adopt ISO standards nationally/within their membership territory.
- **Subscribers' members**: keep up to date with ISO's work but cannot participate in it, do not sell or adopt ISO standards nationally. [10]

Due to the difficulties that come from the process of standardization, the development track
(as called by ISO [11]) is estimated to be in the range of hundreds of millions of dollars, in a period of 18, 24 or 36 months*[1].

[1] *Rough estimate given by Leonardo Chiariglione, former convenor of ISO MPEG, now convenor of MPAI.*



Scheme 1.1: The ISO development track.[2]*

As got into details in [11], there are six phases in the standardization activity, articulated in:

1. **Proposal stage**: the first step that confirms that an International Standard for a specific topic is actually needed.
2. **Preparatory stage**: despite this step is not obligatory, this one is where the working is set up and together with the Project leader the Working draft is prepared.
3. **Committee stage**: this stage is considered optional. Here the working group draft is shared among the members of the parent committee.
4. **Enquiry stage**: The Draft International Standard (DIS) is sent to ISO Central Secretariat by the Committee Manager. Then the members have twelve weeks to make their considerations and vote.
5. **Approval stage**: this stage is not mandatory (and so automatically skipped) if the DIS has been approved. On the other hand, if the draft improves some technical changes the Final Draft International Standard (FDIS) becomes an obligatory step.
6. **Publication Stage**: the final step is the publication of the standard. No more corrections are here done to the FDIS, excepting the editorial ones.

During the Preparatory stage experts look after the issue around patents, copyright and conformity assessment. The patent policy adopted by ISO is a reasonable and non-discriminatory licensing of the technologies, aiming to be royalty free or fair.

[2] *NOTE: "*" stages are mandatory.*

1.2.1 International Telecommunication Union

ITU-T (International Telecommunication Union) is a specialized agency of the United Nations responsible of themes related to technologies, information and communication, so it mainly works on standards that guarantee interoperability, quality, reliability and security.

The main activity of the organization may be divided like so:

- Radiocommunication: related to the global use and satellite;
- Standardization: related to the development of telecommunication standards and to broad its application on the net;
- Development: in relation to the work on implementing a new form of global connectivity in line with the Sustainable Development Goals (SDGs). [47]

The study group of ITU's Telecommunication Standardization Sector (ITU-T) assemble experts coming from all over the globe which common aim is to develop international standards, also known as ITU-T Recommendations which can define elements in the global infrastructure of information and communication technologies.

Its current main location is in Genève and it has 193 member states. [48]

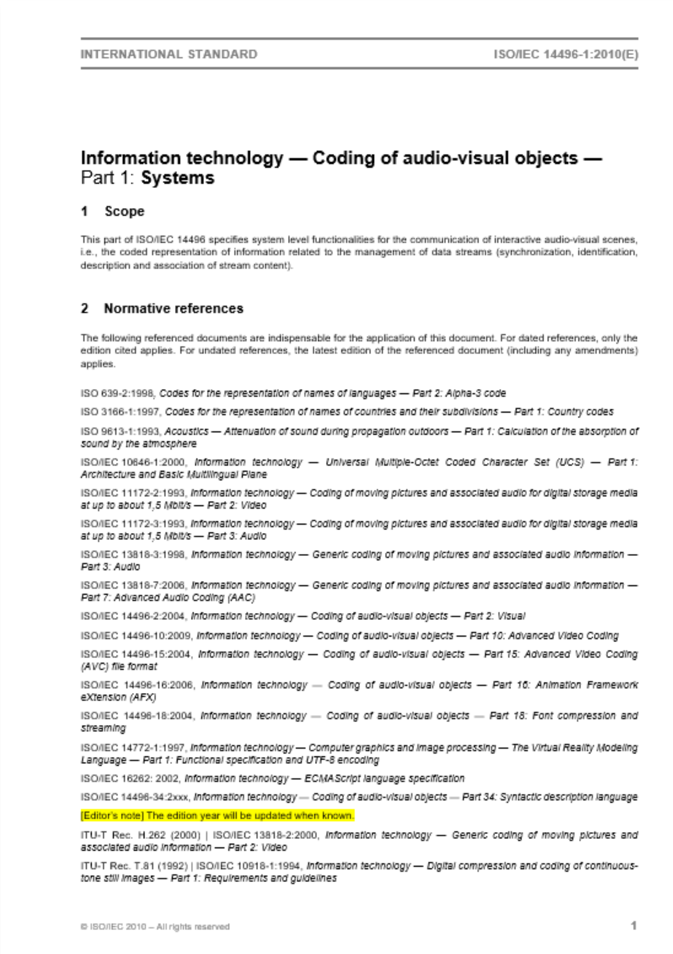
ITU's members are so divided:

- **Member States**: ITU boasts of having as many as 194 member countries, spread all over the globe. [12]
- **Sector members**: ITU welcomes different kinds of companies from all parts of the world. Some of the biggest industries are ITU's members, such as IBM, Huawei, Intel or Microsoft. [13]
- **Associates**: associate membership allows organizations to participate in the work of a specific ITU Sector, without becoming full members. [14]
- **Academia**: it refers to the involvement of academic institutions, that are 160 in the ITU's organization. [15]
- **Regional and International Organizations**: these are organizations that operate across multiple countries, either within a specific, they may include intergovernmental organizations, non-governmental organizations (NGOs), and regional groups focused on telecommunications and ICT development. [16]

1.3 Moving Picture Expert group

Conscientiously, all regarding the internet, image, video and audio processing must be driven by laws too. Inside the giant ISO group, there are several smaller organizations that focus on various topics. The explored one in this thesis is MPEG (Moving Picture Expert Group), formal designation ISO/IEC JTC 1/SC 29/WG 11, which is a technical committee made by the international organizations ISO and IEC (International Electrotechnical Commission) in 1988. The committee was established by Leonardo Chiariglione and Hiroshi Yasuda and the first gathering was in May 1988.

The primary function of MPEG is to develop standards for compression, encoding and transmission of multimedia data in order to improve their quality, combined with a dimension reduction, so that the streaming and the distribution of these can be more efficient. Every MPEG standard is subdivided into different parts, and every part is divided into layers. A part is a document, which concerns a specific topic (e.g., system, visual or audio).



INTERNATIONAL STANDARD ISO/IEC 14496-1:2010(E)

Information technology — Coding of audio-visual objects —
Part 1: **Systems**

1 Scope

This part of ISO/IEC 14496 specifies system level functionalities for the communication of interactive audio-visual scenes, i.e., the coded representation of information related to the management of data streams (synchronization, identification, description and association of stream content).

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 639-2:1998, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

ISO 3166-1:1997, *Codes for the representation of names of countries and their subdivisions — Part 1: Country codes*

ISO 9613-1:1993, *Acoustics — Attenuation of sound during propagation outdoors — Part 1: Calculation of the absorption of sound by the atmosphere*

ISO/IEC 10646-1:2000, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

ISO/IEC 11172-2:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 2: Video*

ISO/IEC 11172-3:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio*

ISO/IEC 13818-3:1998, *Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio*

ISO/IEC 13818-7:2006, *Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC)*

ISO/IEC 14496-2:2004, *Information technology — Coding of audio-visual objects — Part 2: Visual*

ISO/IEC 14496-10:2009, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-15:2004, *Information technology — Coding of audio-visual objects — Part 15: Advanced Video Coding (AVC) file format*

ISO/IEC 14496-16:2006, *Information technology — Coding of audio-visual objects — Part 16: Animation Framework eXtension (AFX)*

ISO/IEC 14496-18:2004, *Information technology — Coding of audio-visual objects — Part 18: Font compression and streaming*

ISO/IEC 14772-1:1997, *Information technology — Computer graphics and image processing — The Virtual Reality Modeling Language — Part 1: Functional specification and UTF-8 encoding*

ISO/IEC 16262: 2002, *Information technology — ECMAScript language specification*

ISO/IEC 14496-34:2xxx, *Information technology — Coding of audio-visual objects — Part 34: Syntactic description language*

[Editor's note] The edition year will be updated when known.

ITU-T Rec. H.262 (2000) | ISO/IEC 13818-2:2000, *Information technology — Generic coding of moving pictures and associated audio information — Part 2: Video*

ITU-T Rec. T.81 (1992) | ISO/IEC 10918-1:1994, *Information technology — Digital compression and coding of continuous-tone still images — Part 1: Requirements and guidelines*

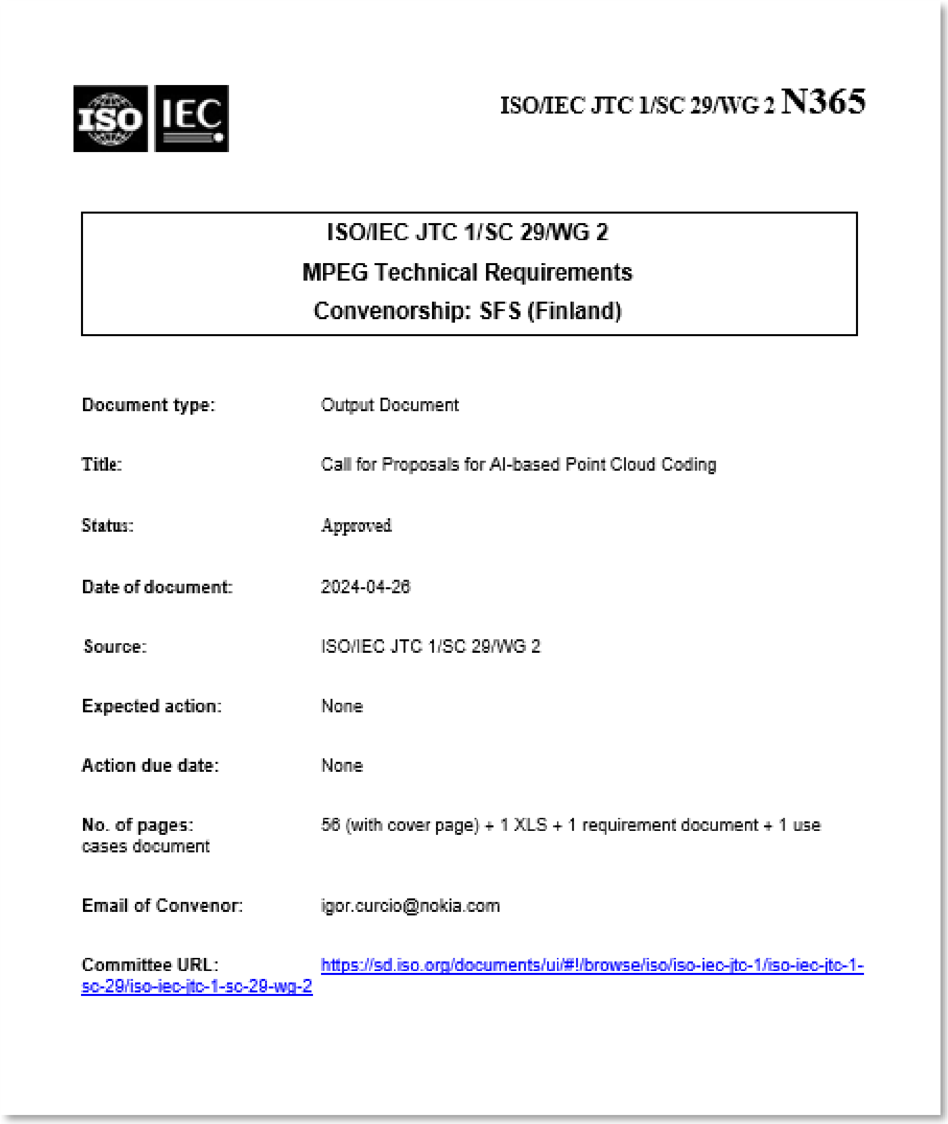
© ISO/IEC 2010 – All rights reserved 1

Figure 1.3: Example of a document concerning MPEG-2 part 1: systems. [17]

This huge work is done by smaller or bigger industries, spread all around the world, which can join MPEG to help developing services and interoperable products.

Technologies involved are pretty sophisticated, that's why MPEG needs all the industries owning good technologies to work with them to make standards that are, every time, more innovative and viable. [18]

To join the MPEG works, experts or companies must be part of one of the nation members of ISO, so that industries can join the international meetings where their proposal is presented for the first time.



ISO/IEC JTC 1/SC 29/WG 2 N365

ISO/IEC JTC 1/SC 29/WG 2
MPEG Technical Requirements
Convenorship: SFS (Finland)

Document type:	Output Document
Title:	Call for Proposals for AI-based Point Cloud Coding
Status:	Approved
Date of document:	2024-04-26
Source:	ISO/IEC JTC 1/SC 29/WG 2
Expected action:	None
Action due date:	None
No. of pages:	56 (with cover page) + 1 XLS + 1 requirement document + 1 use cases document
Email of Convenor:	igor.curcio@nokia.com
Committee URL:	https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-2

Figure 1.4: Example of the first page of an MPEG Call for proposal. [19]

To arrive at the final standard there are several stages before, as depicted in Scheme 1.2, as it was done for the ISO group in 1.1.1:

1. **Use cases/contexts and requirements**: it all starts from the MPEG members, which bring new ideas. Once it's clarified the use context, the use cases and the requirements to support the use cases must be pointed, in order to go on with the work.

2. **Call for evidence**: if the industries think they might have the right technologies and tools to achieve the goal asked by MPEG, they will respond with a Cfe response where they attach their final result.
3. **Call for proposal**: if Cfe produces effective results, MPEG moves on into publishing the Call for proposal, where industries must attach to their Cfp responses all regarding the performance, test data and to demonstrate which technology has been involved to reach that achievement.

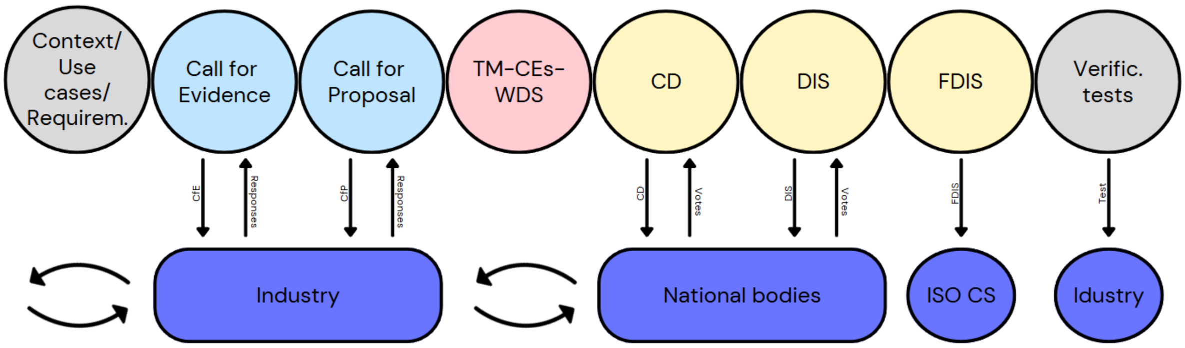
4. **Test models**: here the build of an initial Test Model is asked. MPEG experts can define if there are critical points in the TM, making Core Experiments (CE) to improve the ™, and after all the revisions produce a working draft (WD).

5. **Committee draft**: a document called Committee Draft is published so that, after the submission to the national standard organizations, can create a new document with all the comments done by the National bodies experts.

6. **Draft International Standard**: this document is sent again to the NBs in order to receive positive votes (made only as comments) again.

7. **Final Draft International Standard:** this is the document that will be officially published as an International Standard, after a process of revision made by ISO.

8. **Verification tests**: this might be considered the final stage of the creation of an MPEG standard, where the verification tests are done in order to give precise indications on what to expect from a compression standard. What is needed now are: specification of tests, collection of appropriate test material, execution of reference or proprietary software, execution of subjective tests, test analysis and reporting. [20]



Scheme 1.2: ISO's stages.

Obviously, all this iter includes every part of the standard, so the experts must be able to take into consideration all the innovative technologies regarding video, image, audio etc.

1.4 MPEG and the joint video standards

What video coding does is transforming the video, given as an input, into a compact binary code for more economic and light-weighted storage and transmission. [21]
The most relevant video coding standards over the last 20 years, were published by the union of different groups, coming from the two standardization entities ISO/IEC MPEG, and the ITU-T, which were already broadly discussed.

Figures 1.5 and 1.6: ISO's official logo and ITU's official logo.

All the joint standards generally share the same approach based on:

1. lossless compression based on spatial redundancy (the correlation among sequent pixels in the block), temporal redundancy (temporal redundancy: time correlation between block and half blocks) and entropy coding;
2. compression by deleting what is considered irrelevant, to better explain: the information that is not reconstructable by the codec anymore, but that is perceptible by the human vision system;
3. lossy compression related to the quantization process.

[22]
The Joint Standards are[3]:

- **H.262/MPEG-2**, dated 1995;
- **H.264/AVC** (*Advanced Video Coding*), dated 2003;
- **H.265/HEVC** (*High Efficiency Video Coding*), dated 2013;
- **H.266/VVC** (*Versatile Video Coding*), dated 2020.

1.4.1 H.262/MPEG-2

This codec was first standardized by ITU-T Study Group 16 Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), to work with satellite broadcasting, video conferencing or digital video recording. It also offered a more sophisticated mechanism for the handling of interlaced images and, of course, a better compression in comparison to the previous standard (MPEG-1). The MPEG-2 standard was capable of coding standard definition television at bitrates from about 4-9 Mbits/s and high-definition television at 15-25 Mbit/s, which was a great result at the time. [23] Thanks to it, it became a cornerstone for media distribution formats like DVDs and digital television, solidifying its role as a fundamental technology in multimedia systems.

1.4.2 H.264/AVC

[3] *for all the standards described below, the H.xxx is the ITU name, the other one is the Joint group name.*

AVC is the video coding standard of the ITU-T VCEG group and the ISO/IEC MPEG group.
The primary objectives of the H.264/AVC standardization effort were to improve compression performance and create a "network-friendly" video format suitable for both "conversational" applications (like video telephony) and "non-conversational" uses (such as storage, broadcasting, or streaming). H.264/AVC has successfully enhanced rate-distortion efficiency compared to previous standards. [24] AVC was even the first one capable of combining both online and offline distribution in only one standard. Its usage spaces out from compression, distribution or recording of files on different streaming devices to broadcasting and Blu-Ray discs whenever the distribution is offline.
The broad variety of applications, such as real-time communication or interactive digital applications made AVC the most complete video coding standard ever made at the time. Moreover, includes the possibility to have access to the material present on the World Wide Web. Despite that, its primary usage was for the transmission of standard definition (SD) and high definition (HD) TV signals over satellite, cable, and terrestrial emission and the storage of high-quality SD video signals onto DVDs. [24]

1.4.3 H.265/HEVC

HEVC (MPEG-H part 2) is one of the most used by the streaming platforms, because it supports the transmission of 4K videos. It was developed by the MPEG group and the VCEG group which jointly formed the JCT-VC (Joint Collaborative Team on Video Coding). HEVC was developed to encompass nearly all current applications of H.264/MPEG-4 AVC, with a particular emphasis on two main aspects: supporting higher video resolutions and optimizing performance for parallel processing architectures. [25]
The HEVC codec offers significantly improved compression compared to previous codecs. This codec's aim is to double the video quality for the same bitrate compared to other codecs, making it an ideal choice for live streaming high-resolution video over the Internet, where bandwidth is often limited.

Thanks to HEVC, companies can deliver high-quality live streams. Since it uses less bandwidth, more people can access the stream, increasing the number of viewers. Timing issues are also less problematic. With smaller file sizes, image data is transmitted more quickly. [26] While achieving high compression efficiency is crucial, it is equally important to consider the computational complexity of video codecs. In this context, HEVC delivers superior compression efficiency but requires a significantly higher computational cost compared to H.264. [27]

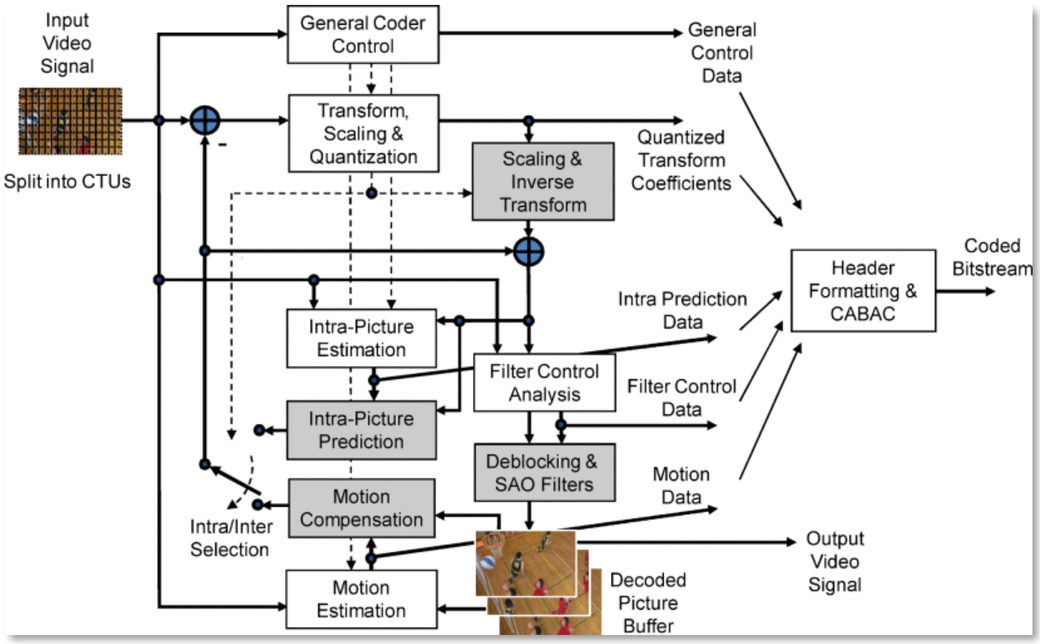


Figure 1.7: Typical HEVC codec's path. [25]

1.4.4 H.266/VVC

The latest one is VVC, born in 2020 (MPEG-I part 3), was developed by the Joint Video Experts Team (JVET) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) to serve an ever-growing need for improved video compression as well as to support a larger variety of today's media content and emerging applications. [28] In general, this type of codec is used for some specific cases such as video conferencing and OTT[4] streaming platforms because of its capacity to reduce the data and the bandwidth for high-definition videos. [29]

[4] *OTT: over the top platforms*

Besides the 4K streaming resolution, which already made VVC the most innovative and useful standard ever (in particular for the OTT and gaming solutions), the real revolutionary skill is brought by the support that it gives to transmit 360 videos, or rather 360 streaming (also called immersive video streaming).

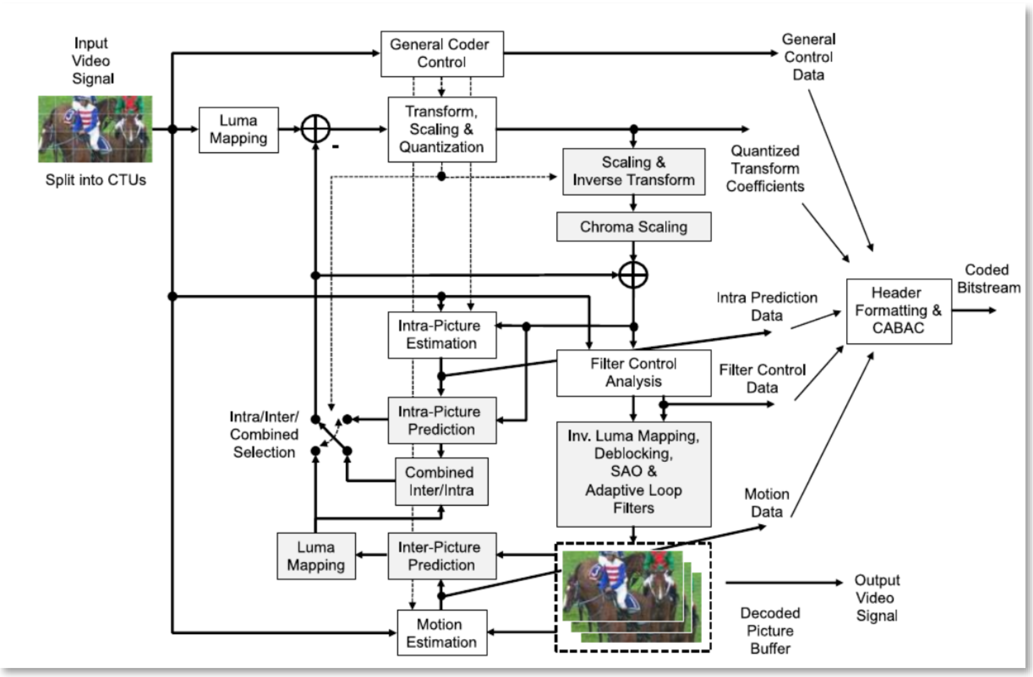
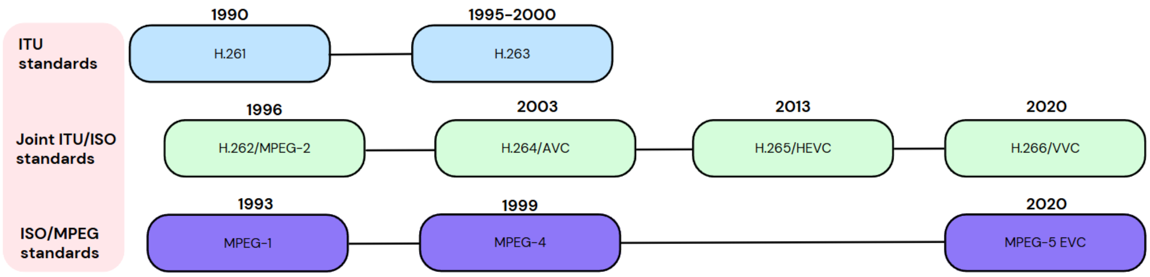


Figure 1.8: Typical VVC codec's path. [28]

In fact, from the outset, VVC was designed not only to deliver significant bitrate savings compared to its predecessor (HEVC), but also to offer broad adaptability, addressing both current and future media demands. This includes video content beyond standard and high definition with SDR, as well as higher resolutions (up to 8K or more), HDR, and WCG. It also covers computer-generated or screen content, such as in screen sharing and gaming, 360° video for immersive and augmented reality, and applications requiring ultra-low latency, like wireless displays and online gaming. [28]



Scheme 1.3: MPEG timeline.

1.5 Alliance for Open Media

AOMedia (Alliance for Open Media) is a technology consortium founded in September 2015, currently based in Wakefield (Massachusetts), with the goal of developing open, royalty-free, and highly efficient standards for multimedia compression. The creation of this consortium was a direct response to several challenges emerging in the global technology landscape, including the increasing technical complexity of compression formats, the costs associated with proprietary licenses, and the supposed fragmentation of standards across the industry. At the foundation of the development of proprietary standards there's a business model structured as follows: the patent holders allow the use of their products in exchange for the payment of significant royalties, which, in turn, are used to fund the development of new technologies for the next generation of standards. To fix the issues introduced by patents, several large companies have undertaken the development of new royalty-free video compression formats.

The main goal is providing a universal solution for video and multimedia compression that can be used without any particular restrictions, while offering benefits to both companies seeking to reduce operational costs and end-users who gain access to high-quality content with lower bandwidth requirements.

This open approach aims to democratize access to advanced technologies, foster innovation, and ensure a more sustainable and inclusive multimedia ecosystem.

Figure 1.7: Alliance For Open Media's official logo.

The founding members of AOMedia include prominent names such as Google, Microsoft, Amazon, Netflix, Cisco, Intel, and Mozilla, companies that play a pivotal role in shaping global network, hardware, and software technologies. Over time, the consortium has grown even stronger with the addition of other industry giants like Apple, Facebook, and hardware manufacturers, underscoring the growing consensus around the need for open multimedia standards.

The company's members are divided into two categories:

- **Board-level members**: they include the main companies such as Amazon, Apple and Google as said before. This one does have a primary level in the alliance, in fact they can vain positions in the AOMedia's governance and strategic direction.

- **Promoter members**: these ones are the ones with a tighter level of participation in the alliance, even due to the privilege and importance that the single company has in the global market.

Figure 1.8: AOMedia's founding and promoter members.

1.6 AOMedia standards

1.6.1 AV1

All the members worked together for the adoption of a new joint compression format called AOMedia Video 1 (AV1), the first video coding standard by AOMedia. AV1, which was launched in 2018, achieved a compression gain of about 30 percent over its predecessor VP9. It was developed with the aim of being royalty-free, scalable, flexible and optimized and it offered a better compression (considering the same perceived quality) and a support for 4K UHD, HDR and WCG in real-time videos. [30]

Figure 1.9: AV1's official logo.

1.6.2 AOMedia Video Model

AVM (AOMedia Video Model) is the latest AOMedia codec,born in 2024. AVM, also known as AOMedia Video 2, is a cutting-edge video codec being developed by the Alliance for Open Media (AOM). Building on the foundation of its predecessor, AV1, it offers improved video quality and greater compression efficiency in comparison to its previous one, delivering a significant performance upgrade. As an open standard, AVM tries to promote innovation and accessibility in video technology, trying to satisfy a broad range of applications, from streaming platforms to virtual reality and real-time communications. [31]

Figure 1.10: AV2's official logo.

Chapter 2: Experimental environment

The performance of the main Video Codecs was evaluated in a desktop environment.

2.1 Reference framework

2.1.1 Machine specification

The tests were performed using two machines:

1) **ASUSTeK COMPUTER INC. TUF GAMING X570-PLUS _WI-FI_** with the following technical specifications:

- Hardware Information:

- **Memory**: 64.0 GiB
- **Processor**: AMD Ryzen™ 9 5950X × 32
- **Graphics**: NVIDIA GeForce RTX™ 3090
- **Disk Capacity**: 5.0 TB

- Software Information:

- **Firmware version**: 4230
- **OS Name**: Ubuntu 24.04.1 LTS
- **Os Type**: 64-bit
- **GNOME Version**: 46
- **Windowing System**: x11
- **Kernel Version**: Linux 6.8.0-51-generic

2) **CPU INTEL I9-14900K** with the following technical specifications:

- **Hardware Information:**

  - **Memory**: 125.0 GiB
  - **Processor**: Intel(R) Core(TM) i9-14900K
  - **Graphics**: Intel Corporation Raptor Lake-s GT1
  - **Disk Capacity**: 7.3 TB

- **Software Information:**

  - **Firmware version**: 4230
  - **OS Name**: Ubuntu 24.04.2 LTS
  - **Os Type**: 64-bit
  - **GNOME Version**: 46
  - **Windowing System**: x11
  - **Kernel Version**: 6.11.0-19-generic

2.1.2 Software

In order to test all the video sequences different softwares have been used:

- **HM** [32]

This software package is the reference software for Rec. ITU-T H.265 | ISO/IEC 23008-2 High efficiency video coding (HEVC). It includes both encoder and decoder functionality. HM implements all the main features of the HEVC codec, including advanced compression methods such as intra-prediction, inter-prediction or in-loop filtering.

- **VVenC - the Fraunhofer Versatile Video Encoder** [33]

The Fraunhofer Versatile Video Encoder provides a publicly available, fast and efficient VVC encoder implementation. VVenC supports real-world encoding features, including frame-level frequency control and perceptually optimized encoding, in order to provide an exible, fast and easy-to-use video encoding solution for the VVC standard. [49] The encoder project includes two encoders: the simpler one which is an encoder app (vvencapp) and a full featured expert encoder (vvencFFapp). The one that has been used for the test is vvencapp.

Instead of using the MPEG reference software for VVC, this optimized version offers a better performance with comparable results.

- **AV1** [34]

The AV1 software package serves as the reference implementation for the AOMedia Video 1 (AV1) codec, developed by the Alliance for Open Media. It includes both encoder and decoder functionality, trying to leverage advanced compression techniques, such as temporal and spatial prediction, adaptive quantization, and loop restoration filtering, to deliver high-quality video at significantly reduced bitrates. All of these peculiarities makes its application ideal for a wide range of applications from streaming to real-time communication.

- **AVM** [35]

AVM, the latest AOMedia software can be seen as a successor of AV1. They both exploit AOMENC/AOMDEC as a bin for encoding and decoding. AVM is a new codec thus not yet supported by all devices and systems. [36]

2.1.3. General information

As in the official experiments within the standardization bodies, to conduct our experiments, RAW video sequences in YUV format 420 have been utilized.

YUV files are uncompressed files that specify the bond among the Y information (Luminance) and U and V (Chrominance). As its analogic predecessor does, YUV uses these three components to represent the color components of Red, Green and Blue.

Luminance is derived from an RGB color by calculating the weighted average of the red, green and blue components.

For standard-definition television, for example, the following formula is used:

Y'[5]=0.299R+0.587G+0.114B or **Y' = 0.2125R + 0.7154G + 0.0721B**

(for most recent and high-definition tvs). [37]

YUV separates brightness information (luminance, Y) from color information (chrominance, U and V).

This separation reflects the sensitivity of the human eye: humans are much more sensitive to brightness details than to color details. Thanks to this feature, it is possible to reduce the amount of data for the color components (U and V) without significantly losing visual quality. This principle is the foundation of techniques such as chroma subsampling (e.g., 4:2:2, 4:2:0), which reduce the overall size of video files.

The YUV format is essential in modern compression algorithms (e.g., H.264, H.265) because it allows luminance and chrominance to be processed separately, with the latter being compressed more aggressively.

The U and V components, also known as chrominance values or color difference values, are derived by subtracting the Y value from the red and blue components of the original RGB color.

U = B - Y'

V = R - Y'

Jointly, the obtained Y, U and V are sufficient to reconstruct the original RGB value.

[5] *(') is usually used to distinguish luma to the luminance, which is a related value and it's usually designated as Y.*

To open a YUV file on every operating system an appropriate software is needed. The software used for the tests on the Ubuntu machine is:

- **Vooya** [38]

Figure 2.1: Vooya's visualization.

- **YUView** [39]

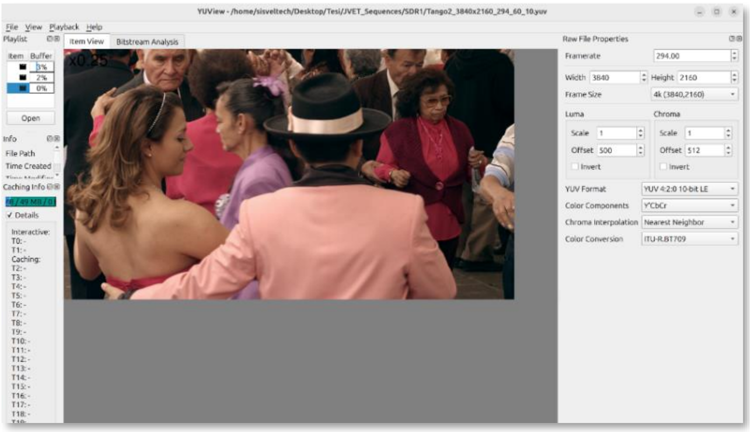


Figure 2.2: YUView's visualization.

2.1.4. Video sequences

The YUV sequences used in this study are 20, and they're all provided by the Joint Video Exploration Team (JVET) and are subject to copyright restrictions. These sequences are made available for research and standardization purposes related to video compression and are not intended for commercial distribution or public sharing. Their usage complies with the terms set by JVET, ensuring adherence to intellectual property regulations.

Snapshots	**Descriptions**
	**Name:** ArenaOfValor **Size:** 1920x1080p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8
	**Name:** BasketballDrill **Size:** 832x480p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** BasketballDrive **Size:** 1920x1080p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** BasketballPass **Size:** 1920x1080p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** BlowingBubbles **Size:** 416x240p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** BQMall **Size:** 832x480p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8

	**Name:** BQSquare **Size:** 416x240p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** BQTerrace **Size:** 1920x1080p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8
	**Name:** Cactus **Size:** 1920x1080p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** FoodMarket4 **Size:** 3840x2160p **Frames:** 300 **Framerate:** 30 Hz **Bitdepth:** 10
	**Name:** FourPeople **Size:** 1280x720p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8
	**Name:** Johnny **Size:** 1280x720p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8
	**Name:** KristenAndSara **Size:** 1280x720p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 8

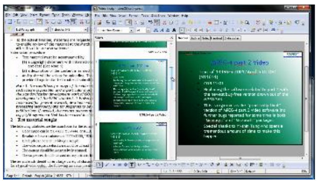
	**Name:** MarketPlace **Size:** 1920x1080p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 10
	**Name:** PartyScene **Size:** 823x480p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8
	**Name:** RaceHorses **Size:** 416x240p **Frames:** 300 **Framerate:** 30 Hz **Bitdepth:** 8
	**Name:** RaceHorses **Size:** 832x480p **Frames:** 300 **Framerate:** 30 Hz **Bitdepth:** 8
	**Name:** RitualDance **Size:** 1920x1080p **Frames:** 600 **Framerate:** 60 Hz **Bitdepth:** 10
	**Name:** SlideEditing **Size:** 1280x720p **Frames:** 300 **Framerate:** 30 Hz **Bitdepth:** 8
	**Name:** SlideShow **Size:** 1280x720p **Frames:** 500 **Framerate:** 50 Hz **Bitdepth:** 8

Table 2.1: Tested video sequences.

The selection of these sequences is based on:

- resolution;
- bit-depth;
- frame rate;
- content of the video.

The most represented scenarios are veryday ones: sport activities, urban landscapes markets, dances or parties. This diversity allows for testing compression, performance and visual quality in contexts with different levels of motion, lighting or detail. High motion scenes, such as BasketballDrill challenge the ability to maintain fidelity and smoothness. Meanwhile others with gradual variations, like FourPeople or Johnny, help assess reproduction quality and the handling of fine details.

2.1.5 Bash command lines

All the parameters used in every single codec guarantee a similar encoding. The main reference is [40] especially for AV1, and starting from that it was possible to repeat the same encoding parameters on the other codecs. To ensure fair comparison, the fixed QP is separately selected for each codec. This removes rate control adaptation between video frames, yielding to an unbiased quality evaluation.

- **HEVC codec:**

```
#Encoding
-i <InputFile.yuv> -c <ConfigurationFile.cfg>  -wdt <Width>
-hgt <Height> -b <BitstreamFile.bin> -f <Frames>
--InputBitdepth <Bitdepth> -q <QP>

#Decoding
-b <BitstreamFile.bin> -o <ReconstructedFile.yuv>
--OutputBitDepth=<Bitdepth> --OutputBitDepthC=<Bitdepth>
```

QP values considered were: 22, 27, 32, 37, 42.

The main tool of the HEVC's codec (HM software) is the configuration file (cfg) which can be considered both as a positive and a negative aspect. It makes the management of the configuration parameters easier without modifying the script or the main code and improves the modularity of the project. Despite it, the cfg must be handled separately. In this study case, the Random Access (RA)[6] configuration was chosen.

Tool	Value
IntraPeriod	32
GOPsize	16
DecodingRefreshType	1
QP	0-51
InternalBitDepth	8
Profile	Main

Table 2.2: Main HM cfg's features.

- **VVC codec:**

```
#Encoding
--preset slow -i <InputFile.yuv> -s <{Width}x{Height}>
-f <FrameRate> -q <QP> -ip <32> -p 1 --profile auto
–format yuv420 -o <OutputFile.bin>

#Decoding
-b <BitstreamFile.bin> -o <ReconstructedFile.yuv>
--OutputBitDepth=<Bitdepth> --OutputBitDepthC=<Bitdepth>
```

QP values considered were: 22, 27, 32, 37, 42.

The main features here are --preset and –profile. The VVenc software does not integrate a cfg file, but the configuration is mainly implemented by the –preset command.
For YUV sequences with a bitdepth of 10, the format and profile parameters must be set on yuv420_10 and main_10 for a correct encoding.

[6] *A Random Access encoding is a video compression technique that allows direct access to specific points in the video stream without the need to decode the entire sequence from the beginning.*

- **AV1 codec:**

```
#Encoding
--width=<Width> --height=<Height> --limit=<Frames>
--fps=<Framerate/1> --cq-level=<QP> --codec=av1 --good
--input-bit-depth=<Bitdepth> --psnr=1 --end-usage=3
--kf-max-dist=32 --kf-min-dist=32 --cpu-used=1
--auto-alt-ref=1 --arnr-maxframes=7 --arnr-strength=5
--passes=1 -o <OutputFile.ivf>

#Decoding
-o <ReconstructedFile.yuv> -o <OutputFile.ivf>
```

QP values considered were: 27, 32, 37, 42, 47.

This command encodes a video using the AV1 codec with specific parameters for resolution, frame rate (--fps), and quality. The –good parameter indicates an encoding modality that is quality optimized, so it balances compression and encoding efficiency, it is the pre-defined option for the libaom-av1 encoder. --kf-max-dist=32, --kf-min-dist=32 sets a fixed keyframe interval corresponding to a 32 Intra Period of HEVC and VVC. --cpu-used=1 configures encoding efficiency because it is able to balance compression quality and encoding speed. Additionally, the command applies temporal filtering parameters (--auto-alt-ref=1, --arnr-maxframes=7, --arnr-strength=5).

- **AVM codec:**

```
#Encoding
--width=<Width> --height=<Height> --limit=<Frames>
--fps=<Framerate/1> --qp=<QP> --good
--input-bit-depth=<Bitdepth> --psnr=1 --end-usage=3
--kf-max-dist=32 --kf-min-dist=32 --cpu-used=1
--auto-alt-ref=1 --arnr-maxframes=7 --arnr-strength=5
--passes=1 -o <OutputFile.ivf>

#Decoding
-o <ReconstructedFile.yuv> -o <OutputFile.ivf>
```

QP values considered were: 110, 135, 160, 185, 210[7].

AVM commands are almost the same as the AV1's ones. In this case, it isn't necessary to specify the codec used and the quality is exposed by the –qp parameter.

The configuration parameters shown in the HM' cfg, originally used for encoding in HEVC, were reproduced for encoding in AV1 and AVM to ensure optimal comparability between the four standards. In particular, the encoding profile and intra period were kept consistent with the HEVC configuration to preserve the compression structure. This process allows impartial comparison of compression performance, video quality and computational complexity between the two formats.

2.1.6 QP's evaluation

The sequences were tested on each codec considering five different quantization parameters (QP). This last one determines the precision of the quantization applied to the coefficients of the video signal transform, which is usually a DCT or one of its variants. It allowed the analysis of the influence of the QP values on the final vision quality and bitrate efficiency.

[7] AVM codec has a different quantization scale. These values are comparable to the other codecs' ones in terms of given results.

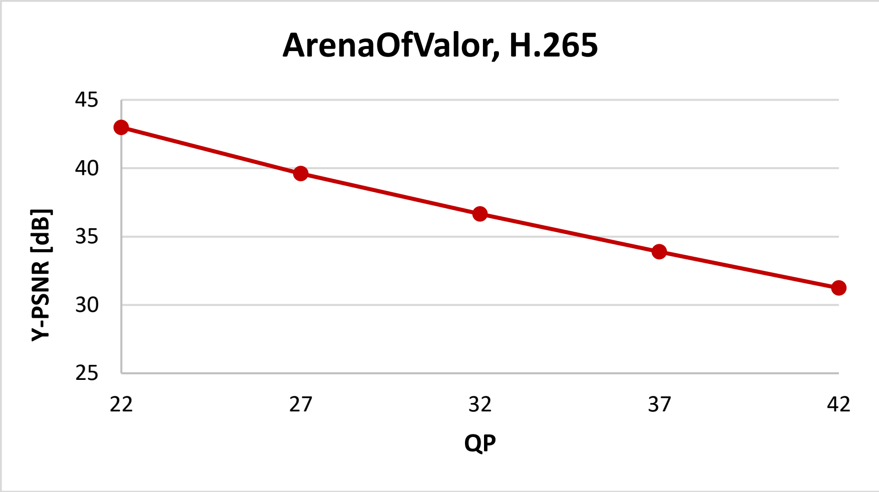


Figure 2.3: ArenaOfValor (HEVC), comparison between QP and PSNR.

A lower QP keeps the majority of the details of the original video resulting in a high-quality sequence but it generates a bigger dimension file. On the other hand, a higher QP increases the compression, it reduces significantly the final bitrate for a lower quality video.

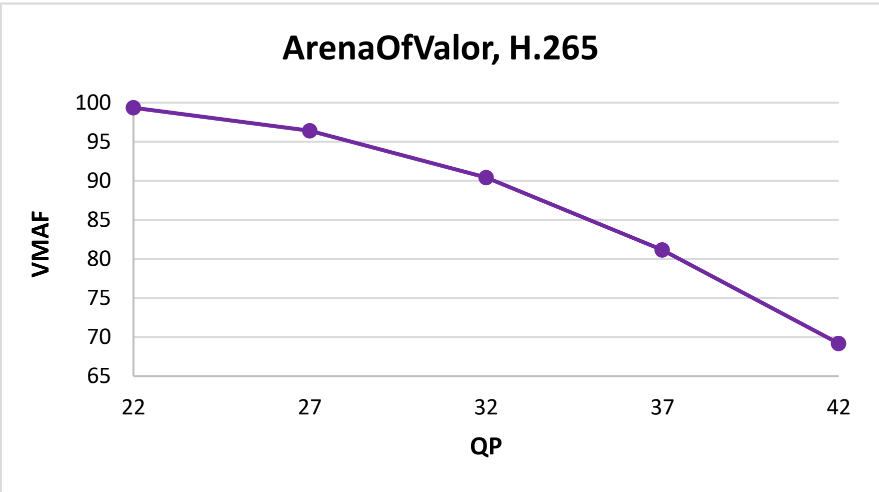


Figure 2.4: ArenaOfValor (HEVC), comparison between QP and VMAF.

HM software, referring to ArenaOfValor sequence was taken into consideration.

As depicted in the figure 2.4, it's clear that for higher QP values the PSNR (see 2.3.1) tends to decrease drastically. So does the video quality.

It’s definitely clear how the QP 22 decoded one is way less degraded than the QP 42. The encoding was quicker, but the compression is visible higher.

Figure 2.5: Quality difference between QP 22 and QP 42 for ArenaOfValor (H.265).

2.2 Quality metrics

The application of any form of processing to an image results in loss of information, and image quality is no exception. Methods for assessing image quality differ primarily in two categories: objective and subjective approaches. Subjective approaches rely on human perception, on the other hand, objective approaches incorporate explicit numeric criteria for evaluation. Both methods employ some sort of comparison with the reference data which includes ground truth images and prior knowledge represented statistically through metrics and tests. [41]

PSNR used to be considered the most reliable one. Despite that, it can be observed the use of other quality metrics rather than PSNR, such as SSIM, VMAF or subjective measures like MOS[8]. Nevertheless, there are several notable differences in how these metrics behave. For instance, PSNR is unbounded, whereas other metrics like SSIM and VMAF do have specific bounds. Moreover, SSIM tends to reach a saturation point at higher bitrates. Another challenge associated with subjective quality assessments, such as MOS, is that their values do not always increase consistently. [42]

2.2.1 PSNR

Peak Signal-to-Noise Ratio, or PSNR, is the most common measure used in the assessment of video quality. This formula determines the quality of an image or video by quantifying the proportion of maximum possible signal power to the noise that degrades the integrity of that image or video. It quantifies the distortion of an image or video resulting from data compression or other factors. With higher PSNR values, the lesser the distortion or the noise and thus the better the quality of the image or video. The measure PSNR indicates the level of fidelity in the processed signal. [43]

PSNR can be also mathematically explained[9]:

$$\mathbf{PSNR} = \mathbf{20log_{10}}\left(\frac{\mathbf{MAX_f}}{\sqrt{\mathbf{MSE}}}\right) \text{ and } \mathbf{MSE} = \frac{\mathbf{1}}{\mathbf{mn}}\sum_{\mathbf{0}}^{\mathbf{m-1}}\sum_{\mathbf{0}}^{\mathbf{n-1}}||\mathbf{f(i,j)} - \mathbf{g(i,j)}||^{\mathbf{2}}$$

where:

f = matrix data of the original image
g = matrix data of our the degraded image
m = pixels' number of rows
i = rows' index
n = pixels' number of colums
j = columns' index
MAX_f = maximum signal value

The PSNR value provides an indication of the fidelity of the processed signal:

[8] *Mean Opinion Score: a numerical measure of the human-judged overall quality of an event or experience.*
[9] *For this implementation, a 2D array of data or matrix was assumed.*

- **Higher PSNR values** indicate better quality and closer similarity to the reference signal.
- **Lower PSNR values** suggest greater distortion or loss of information.

Since many signals have a relatively wide range, the PSNR value is typically expressed on a logarithmic scale in decibels (dB). Generally, values range from 20 to 50.

Despite its advantages, PSNR has limitations and critics. The primary criticism is that PSNR does not always offer a precise alignment with human perception of video quality. PSNR actually calculates the absolute difference between the original and the distorted video. However, the perception of a video may be influenced by several factors, such as the video's content, the viewing conditions, and individual visual acuity. As a result, a video with a high PSNR value might not appear of high quality to a human observer, and vice versa. Overlooking regional errors is one more disadvantage. PSNR tries to consider the errors over the entire video which can ignore certain localized distortions that drastically alter the perceived quality. It is possible that an artifact that is small in nature, but noticeable and occurs within a critical region of the video can drastically skew the viewer's perception yet have little impact on the overall PSNR value.

The reference software used do include PSNR as a standard metric in the output of and encoding process. By calculating PSNR values during encoding, these codecs provide insights into the level of distortion or quality loss introduced by compression. This integration ensures that PSNR is readily available as a performance indicator, facilitating comparisons across different encoding settings and codec implementations.

```
          Bitrate(kbps)  |  PSNR(Y)   |  PSNR(U)   |  PSNR(V)   |  PSNR(Avg)  |  PSNR(Overall)  |  Encoding time (FPS)
--------------------------------------------------------------------------------------------------------------------
Summary:    1611.882400  |  42.478231  |  46.031109  |  45.172613  |  43.282337   |  43.099014       |  67332.2s (0.0 fps)
encoding ends
```

Figure 2.6: Example of an econding output (AVM), to show how PSNR is alrealdy calculated.

2.2.2 Bjøntegaard Delta

The Bjøntegaard Delta (BD) was first proposed by Gisle Bjøntegaard in 2001. This method was thought to be the metric able to compare Rate Distortion curves for fixed Quantization Parameter (QP) based encoded video sequences. This QP-based encoding usually ensures “good” overlap between the two compared Rate Distortion curves, which can be defined as “well-behaved.” The most common practise, which has been done also for this thesis’ work, is to compress each video sequence using different QP values (usually four to evaluate the BD). So, after the encoding, Bitrate and PSNR are automatically calculated for each encoded sequence, and then the BD-Rate and the BD-PSNR values can be computed.

BD-Rate makes it possible to measure the bitrate reduction offered by a codec or codec feature while maintaining the same quality as measured by objective parameters.

The BD-rate can be actually obtained from a specific formula:

$$\boldsymbol{BDRate = 100\left(e^{k} - 1\right)} \text{ where } \boldsymbol{k = \frac{\mathrm{int}_2 - \mathrm{int}_1}{\mathrm{y}_2 - \mathrm{y}_1}} \quad [44]$$

To compare two encoders (or two encoding configurations) the area between the two RD-curves have to be divided by the integration interval.

BD-Rate is measured in percent and expresses the average percentage difference in bitrate of the two data sets at a similar distortion value measured by an objective metric (usually PSNR).

If PSNR distortion value is compared using the bitrate as the basis of integration, the relative metric is known as BD-PSNR (measured in dB). [45]

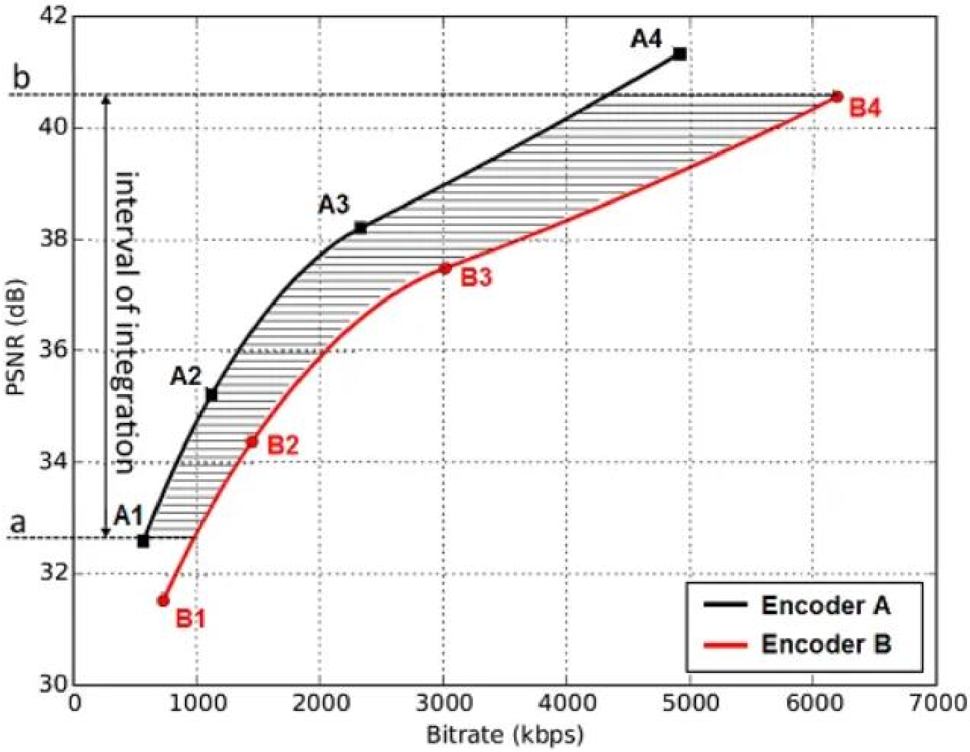


Figure 2.7: Example of curves with different rates. [45]

2.2.3 VMAF

Video Multi-Method Assessment Fusion (VMAF) is an advanced quality metric developed by Netflix to evaluate the perceptual quality of images and videos. Unlike traditional metrics such as PSNR and SSIM[10], which rely on mathematical differences between pixels, VMAF is an objective metric with a strong focus on aligning with subjective perception of quality. It uses machine learning models trained on datasets that include subjective opinions from human viewers, which allows it to approximate human visual judgment. This makes it particularly valuable in assessing how end-users experience visual content, especially in the context of streaming platforms and video compression. It integrates aspects of both traditional pixel-based measurements and perceptual metrics to simulate how humans evaluate visual quality. The VMAF score ranges from 0 to 100, where higher scores indicate better quality and closer similarity to

[10] *Structural similarity index measure is another perceptual quality metrics that quantifies image quality degradation, caused by processing such as compression or by losses in data transmission.*

the reference content. Moreover, by incorporating multiple metrics, VMAF provides a nuanced assessment of visual quality.

VMAF's main negative aspect might be its computational resources, which are way more than simpler metrics. [46]

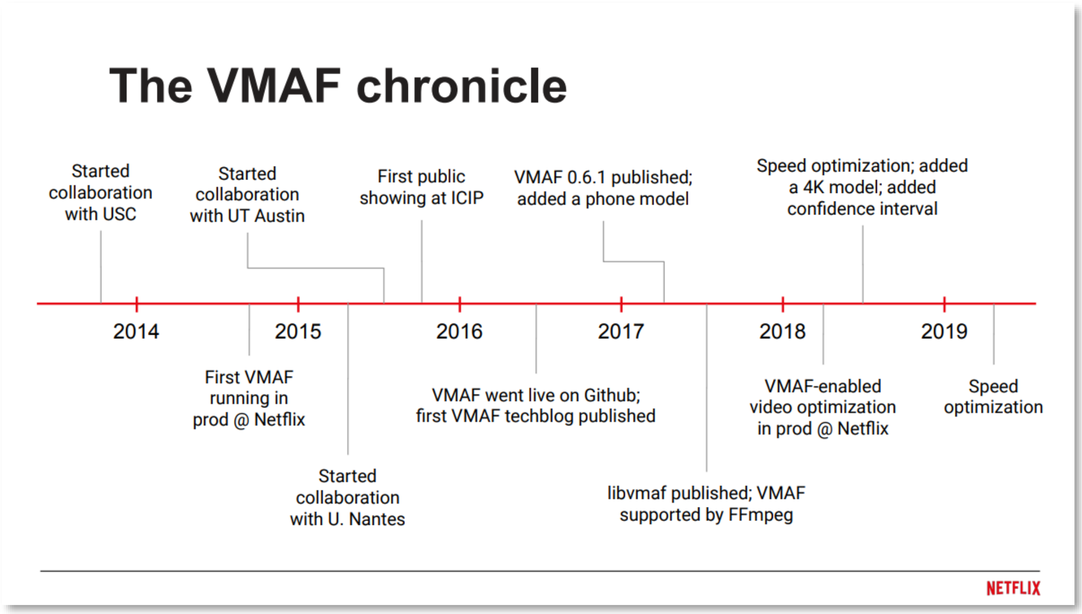


Figure 2.8: The VMAF chronicle.

Chapter 3: Results and evaluation

The most relevant results obtained from the analysis of the parameters and QPs of interest mentioned in section 2.1.5 will be presented below.

A single pass encoding has been done and all encodings were implemented in the YCbCr 4:2:0 color space.

Coding performance is measured using the metrics described in 2.2 and all the sequences are referred to table 2.1.

3.1 BD-Rate evaluation

The tables below present the results of the BD-Rate and BD-PSNR calculation. The input data consists of Bitrate and Y-PSNR values for the different video sequences. Negative values indicate a Bitrates reduction, meaning a gain in compression efficiency compared to the reference codec in the comparison. The more negative the value, the more efficient the codec is in reducing bitrate while maintaining the same quality. The opposite happens for the BD-PSNR. To compute the BD-Rate and the BD-PSNR values only the first four QP values were considered.

These results provide insights into the performance improvements achieved by newer codecs over older standards.

Sequence	H.265 vs H.266	AV1 vs AVM	H.266 vs AVM
ArenaOfValor	-22%	-15%	1,40%
BasketballDrill	-28%	-17,20%	-1,70 %
BasketballDrive	-29,60%	-15,30%	6,60%
BasketballPass	-17,8%	-12,10%	-2%
BlowingBubbles	-21%	-9,30%	13,30%
BQMall	-22,20%	-13,50%	8%

BQSquare	-28,70%	-15,90%	6,50%
BQTerrace	-28,70%	-10,60%	14,70%
Cactus	-31,40%	-16,20%	15,70%
FoodMarket4	-31,70%	-12,20%	14,80%
FourPeople	-25,30%	-16,80%	15,30%
Johnny	-29,10%	-17,94%	14,50%
KristenAndSara	-24,60%	-18,50%	17%
MarketPlace	-37,60%	-14%	14%
PartyScene	-24,60%	-9,90%	-6,90%
RaceHorses	-16,60%	-16,10%	-2,20%
RaceHorses	-15,90%	-15,90%	-5,10%
RitualDance	-23,20%	-15,10%	-0,70%
SlideEditing	-52,20%	-38,60%	15,10%
SlideShow	-44,10%	-18,50%	8,20%

Table 3.1: BD-Rate's results.

Sequence	**H.265 vs H.266**	**AV1 vs AVM**	**H.266 vs AVM**
ArenaOfValor	1,1 dB	0,65 dB	-0,05 dB
BasketballDrill	1,4 dB	0,74 dB	0,07 dB
BasketballDrive	0,85 dB	0,28 dB	0,15 dB
BasketballPass	0,96 dB	0,66 dB	0,10 dB
BlowingBubbles	0,97 dB	0,41 dB	-0,5 dB
BQMall	0,98 dB	0,47 dB	-0,29 dB
BQSquare	1,25 dB	0,61 dB	0,22 dB
BQTerrace	0,55 dB	0,1 dB	0,19 dB
Cactus	0,94 dB	0,30 dB	-0,32 dB
FoodMarket4	1,11 dB	0,841 dB	-0,1 dB
FourPeople	0,96 dB	0,38 dB	-0,41 dB

Johnny	0,72 dB	0,23 dB	-0,40 dB
KristenAndSara	0,74 dB	0,35 dB	-0,36 dB
MarketPlace	1,28 dB	0,36 dB	0,44 dB
PartyScene	1,2 dB	0,45 dB	0,29 dB
RaceHorses	0,85 dB	0,927 dB	0,11 dB
RaceHorses	0,64 dB	0,22 dB	0,02 dB
RitualDance	1,24 dB	0,71 dB	0,01 dB
SlideEditing	12,21 dB	5,98 dB	-1,97 dB
SlideShow	4,405 dB	1,51 dB	-0,56 dB

Table 3.2: BD-PSNR's results.

These values are an important reference for the discussion of the results, because they offer a precise numerical vision of the actual differences between the various codecs for each single sequence analyzed.

	H.265 vs H.266	**AV1 vs AVM**	**H.266 vs AVM**
BD-Rate	-29,15%	-15,93%	8,03%
BD-PSNR	1,71 dB	0,77 dB	-0,18 dB

Table 3.3: avarange BD-Rate and BD-PSNR values.

3.1.1. Computer generated content

The results' discussion will start from the computer graphics scenes, considering, for example, ArenaOfValor and SlideShow. The graphs below show the bond between the bitrate and video quality in PSNR. These sequences represent videos with computer graphics scenes, meaning that compression may be influenced by graphical details and textures.

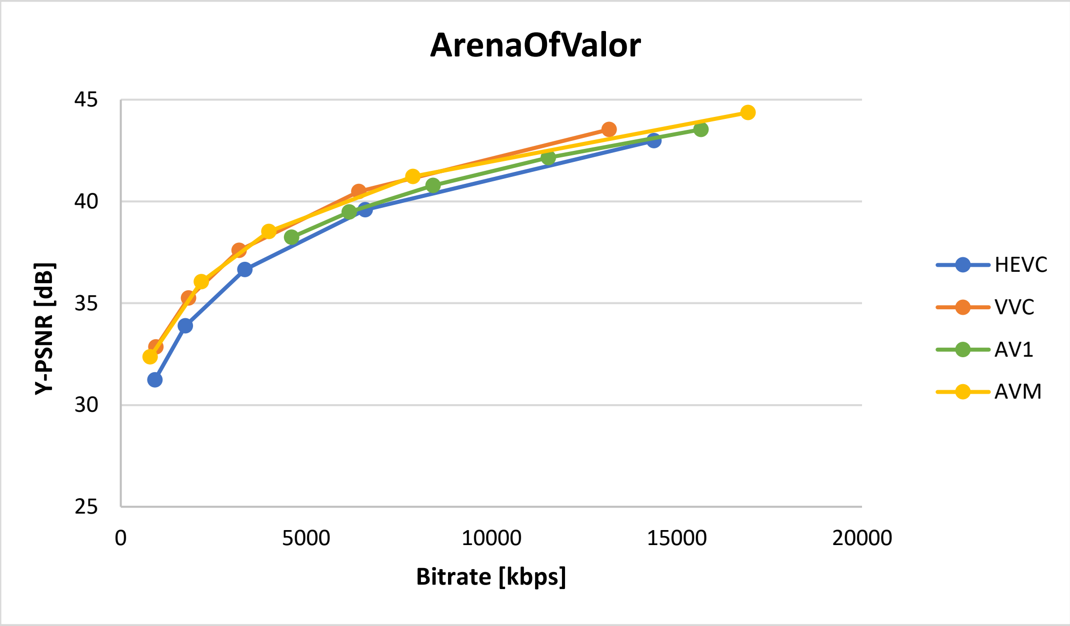


Figure 3.1: ArenaOfValor's Bitrate-PSNR graphic.

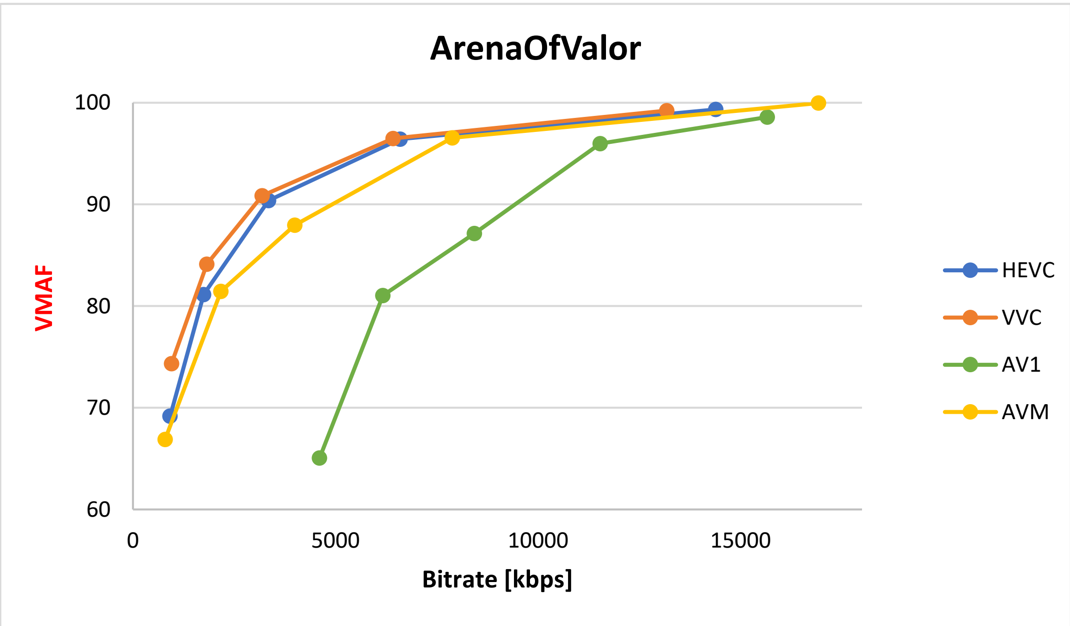


Figure 3.2: ArenaOfValor's Bitrate-VMAF graphic.

As it is clearly visible in figure 3.1, VVC, AV1, and AVM offer higher quality than HEVC for most bitrates, with AVM seemingly providing the best overall results. So, for ArenaOfValor at the higher bitrates the best results are given by AV1 and AVM.

Considering only the AOMedia's codecs: it's definitely noticeable that AVM gives better results rather than AV1, it can be seen especially taking the VMAF's results into consideration and the BD-Rate value in table 3.1 where AVM gains the 15%.

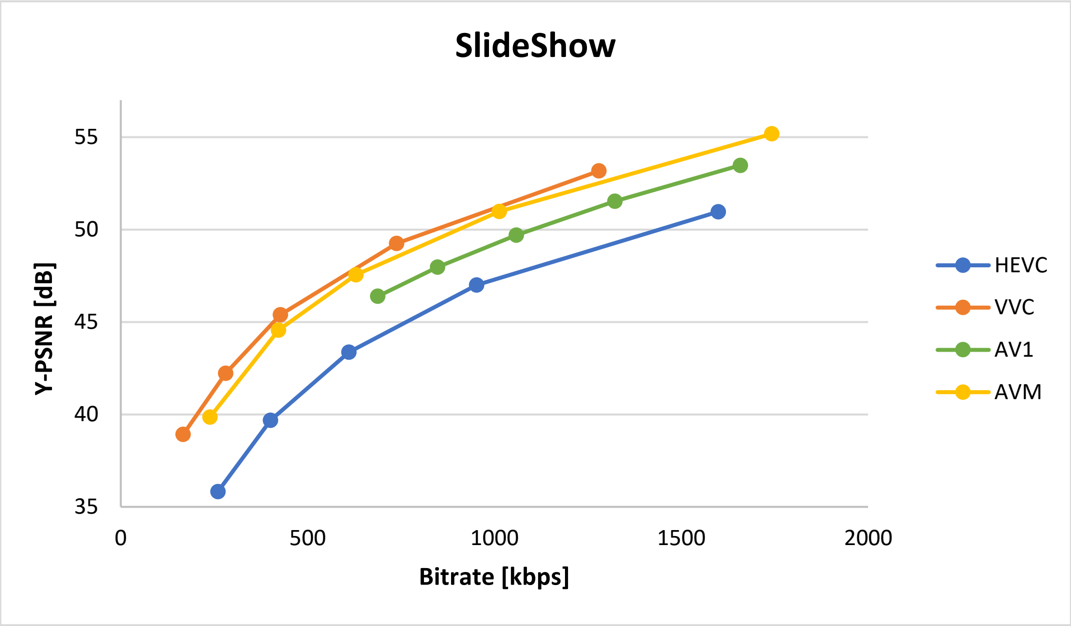


Figure 3.3: SlideShow's Bitrate-PSNR graphic.

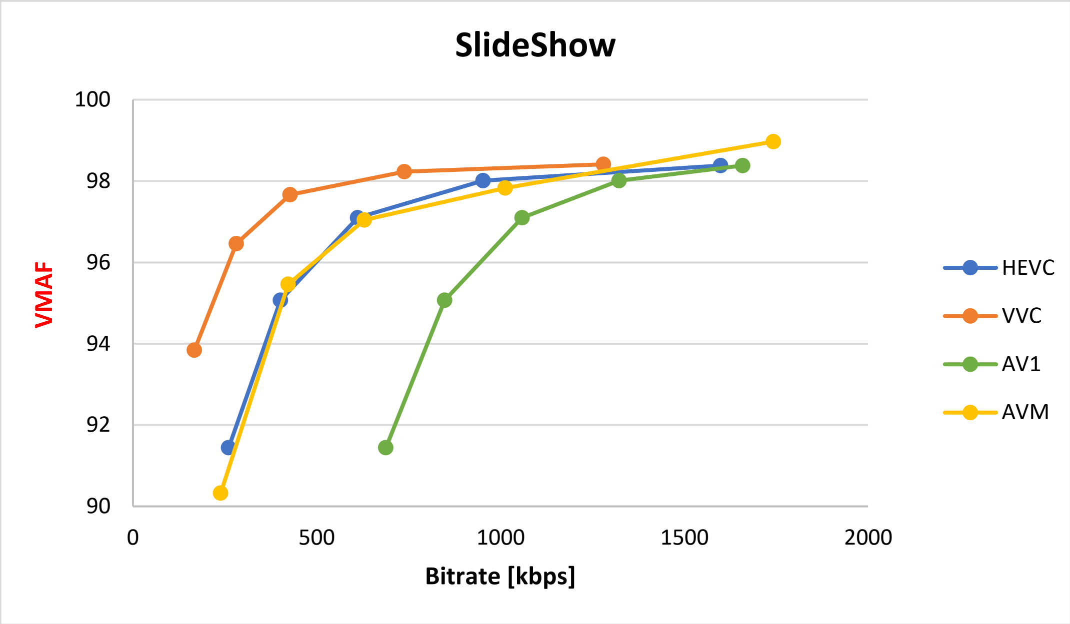


Figure 3.4: SlideShow's Bitrate-PSNR graphic.

The second scenario, SlideShow, shows a greater variation between the codecs, with VVC and AVM outperforming HEVC at lower bitrates.

The PSNR values for SlideShow appear significantly higher compared to ArenaOfValor, and the curves show a huge difference between the codecs. This carriage can be related to several factors linked to the nature of computer-generated scenes and the features of the SlideShow's content:

- the sequence consists of slides, text and very simple graphics with low motion animation and few complex textures;
- compression is more efficient on such content because codecs can encode static or predictable regions with fewer bits while maintaining high quality;
- most advanced codecs (such as VVC or AVM) use precise techniques like intra-prediction[11] or block partitioning[12] to better encode areas that present little variation.

[11] *Intra-prediction: it uses only information from the same frame to predict the blocks.*
[12] *Block partitioning: it is a technique used in video codecs to divide an image (frame) into smaller blocks to optimize compression.*

The higher PSNR values in SlideShow are a result of its low complexity, minimal motion, and well-defined edges, making it highly compressible. This contrasts with ArenaOfValor, where motion, textures, and dynamic shading introduce more compression challenges, resulting in lower PSNR even at higher bitrates. As already depicted in ArenaOfValor, also the VMAF's results are higher for the AVM codec compared to AV1.

In general, the difference between codecs is definitely noticeable both at lower but in particular at higher bitrates, which is expected as compression efficiency diminishes with increased bitrate.

3.1.2 High-motion content

Video sequences that show more dynamic scenario require more motion estimation, to predict a movement between consecutive frames and to reduce the amount of data to store. So, more movement requests more motion estimation, with the consequence of a bigger usage of inter-frame compression.

For this type of analysis, BasketballPass and PartyScene are taken into consideration. These both shows high-dynamic scenes, with young boys playing Basketball and young girls dancing at a birthday party.

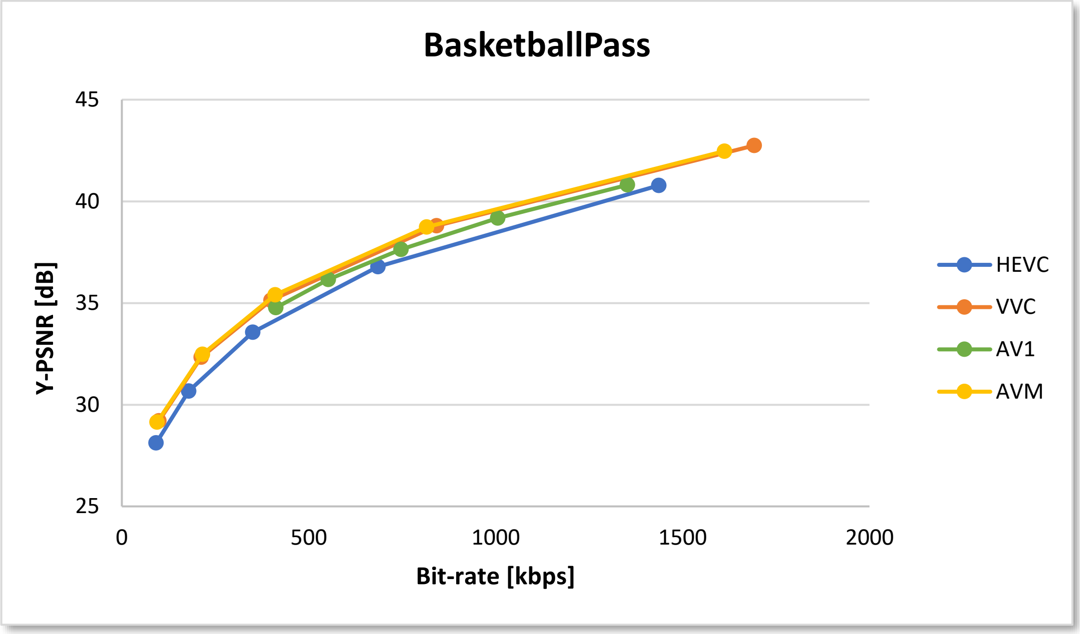


Figure 1.5: BasketballPass' Bitrate-PSNR graphic.

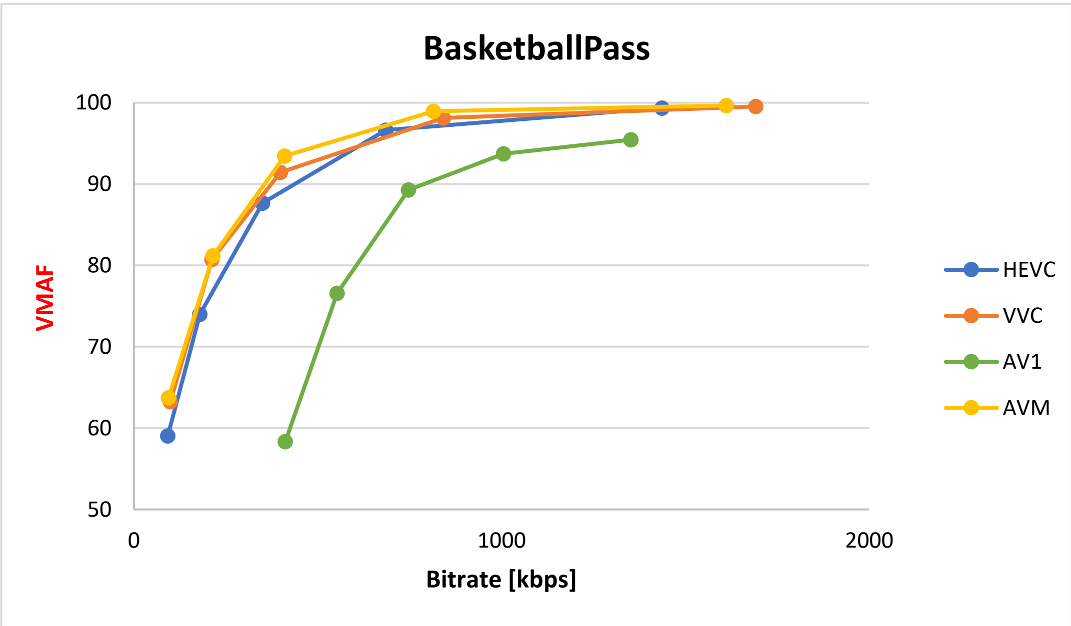


Figure 3.6: BasketballPass' Bitrate-VMAF graphic.

As foreseeable, VVC and AVM do give the best performance, demonstrating that it is optimal in handling scenes with high motion. AVM offers a little improvement comparing to the others at the lower bitrates. Conversely, at higher bitrates, VVC tends to have a better performance with a delta of (approximately) 200 compared to AVM. In fact, referring to table 3.1, AVM loses 1,40% of BD-Rate value towards VVC.

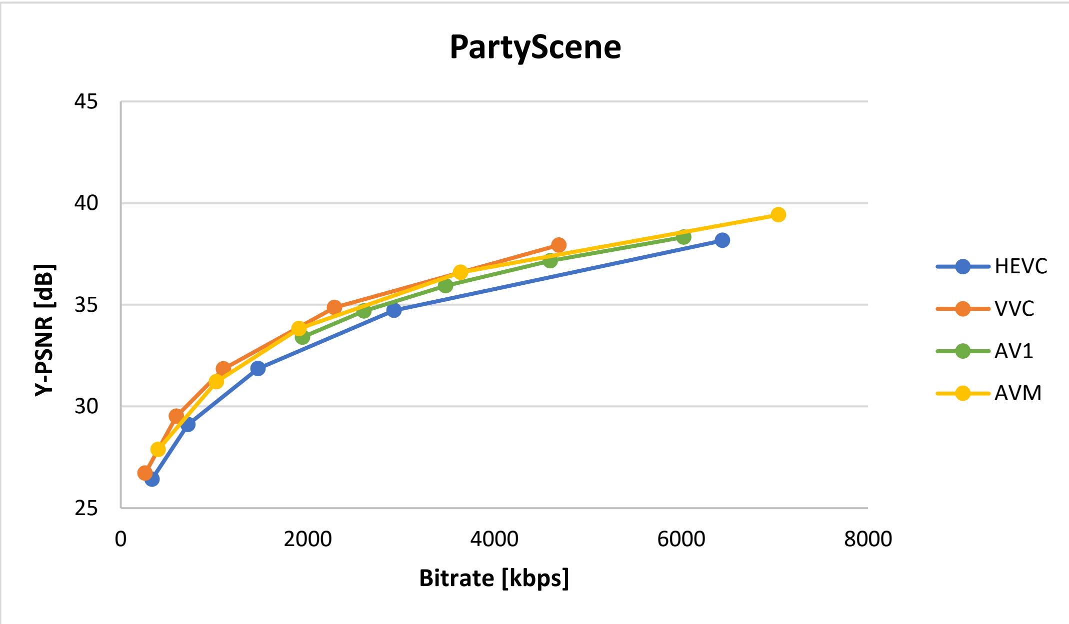


Figure 3.7 PartyScene's Bitrate-PSNR graphic.

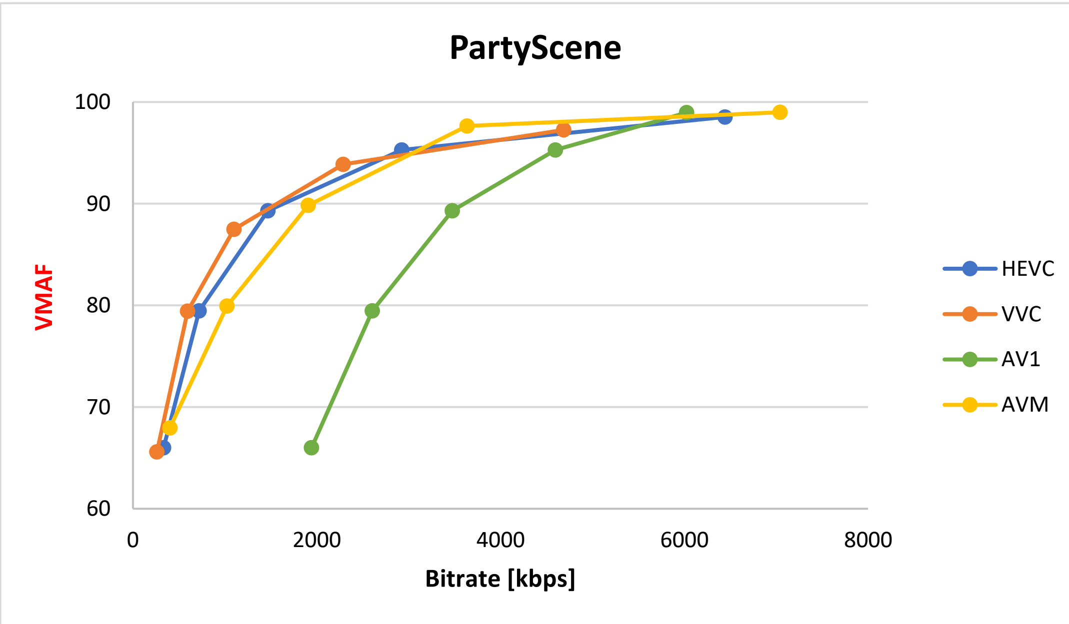


Figure 3.8: PartyScene's Bitrate-VMAF graphic.

Even in this case, as shown in figure 3.7, the differences between codecs are reduced at high bitrates (>5000 kbps). AVM and VVC show the best compression efficiency, with a clear advantage over HEVC and AV1. HEVC requires much more bitrate to achieve acceptable quality, confirming that it is less optimized for dynamic scenes.

In this case, AVM seems to work slightly better than its MPEG rival, both at higher and lower bitrates. But, notwithstanding, referring to Table 3.1, AVM shows a 6,90% BD-Rate increase compared to VVC.

BasketballDrill and BasketballDrive have a similar trend to BasketballPass' one. Indeed, the sequences all show fast movements of basketball players, with different resolutions.

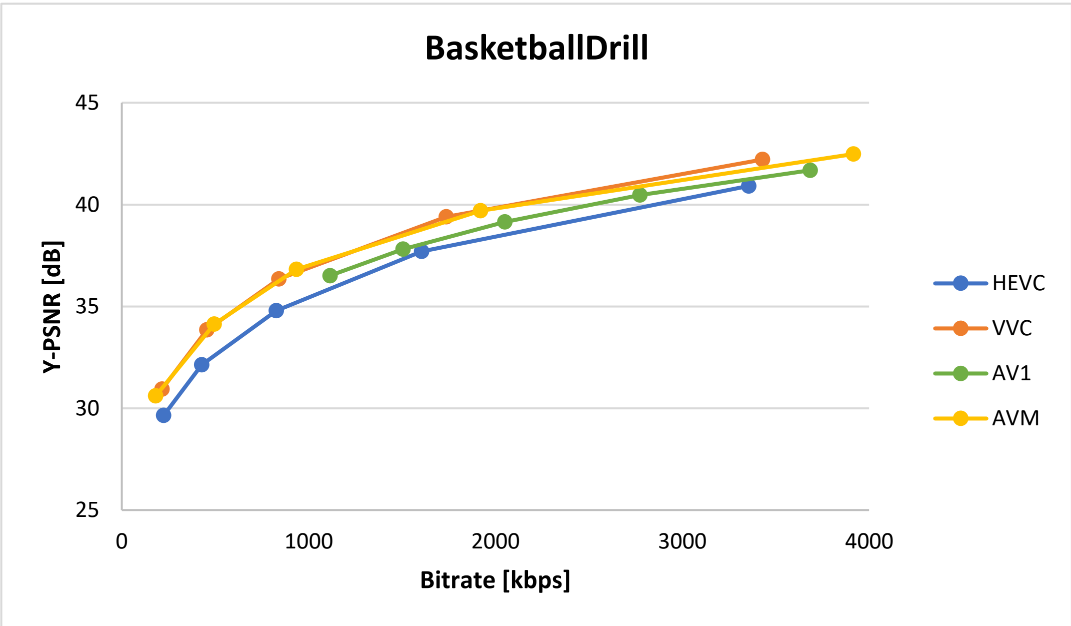


Figure 3.9: BasketballDrill's Bitrate-PSNR graphic.

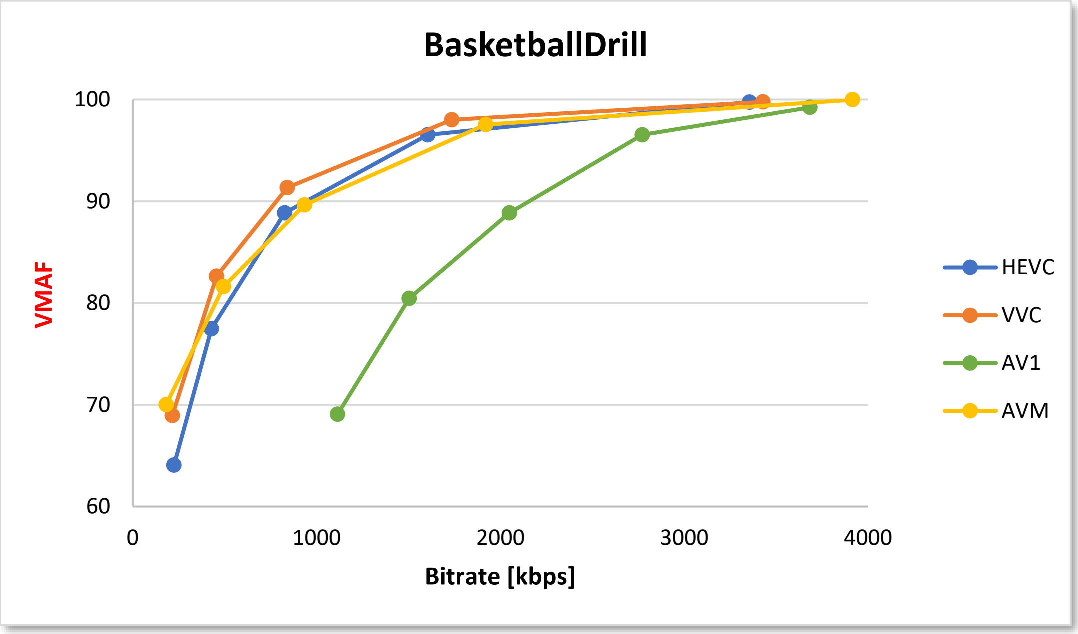


Figure 3.10: BasketballDrill's Bitrate-VMAF graphic.

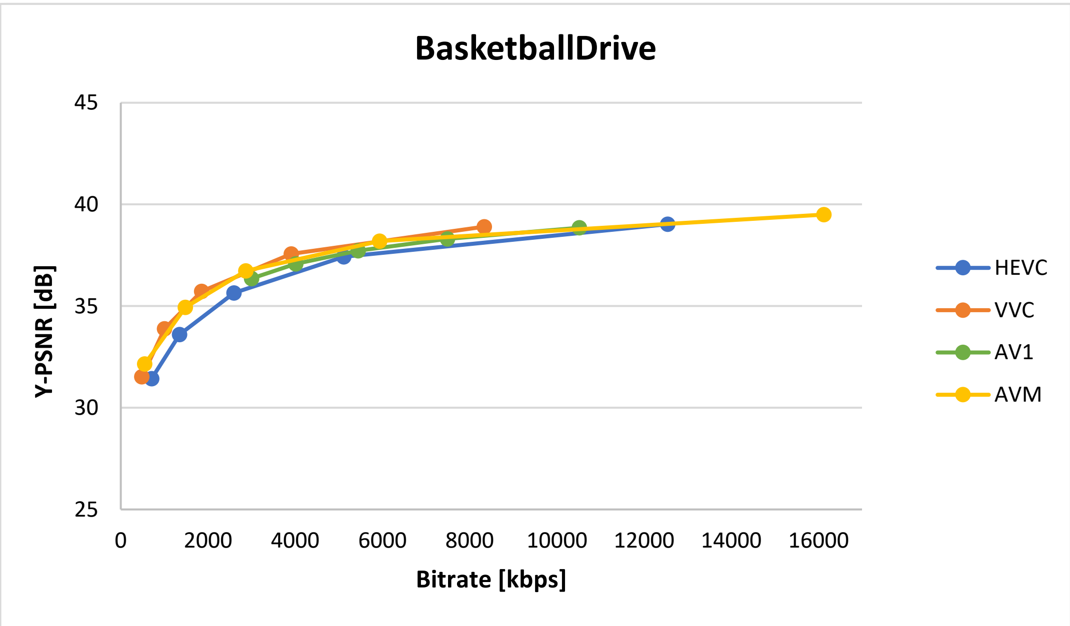


Figure 3.11: BasketballDrive's Bitrate-PSNR graphic.

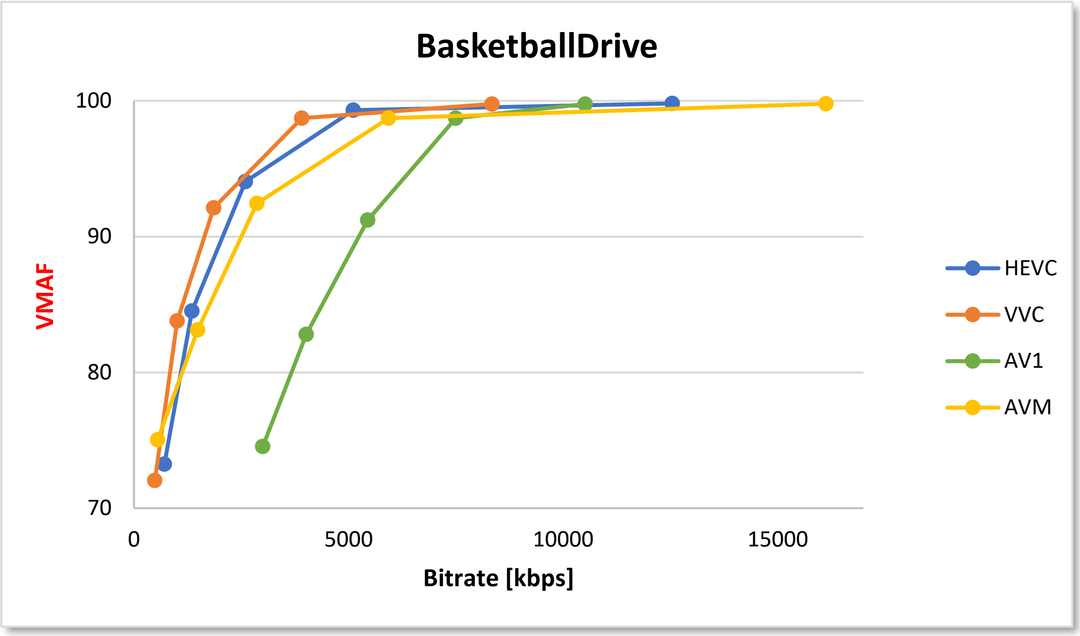


Figure 3.12: BasketballDrive's Bitrate-VMAF graphic.

Despite the different results obtained, all the VMAF's score overcome the 55 score, which means that the compressed sequence can be considered reliable compared to the original ones.

3.1.3 Low-motion content

Contrary to what was analyzed in 3.1.2, some other sequences require way less motion estimation, because the depicted scenario is less dynamic. This kind of sequences (e.g., surveillance, interviews) have been always used as a reference for video coding tests, because it ensures a better compression due to the bigger lack of movements showed. From table 2.1 FourPeople, Johnny and KristenAndSara can be taken.

FourPeople shows an interview with three men and a woman talking to each other. Johnny is a single person interview to a man. KristenAndSara instead, shows a short conversation between two young women, with a dark blue background. The scenes are therefore static and without particular movements, except facial ones.

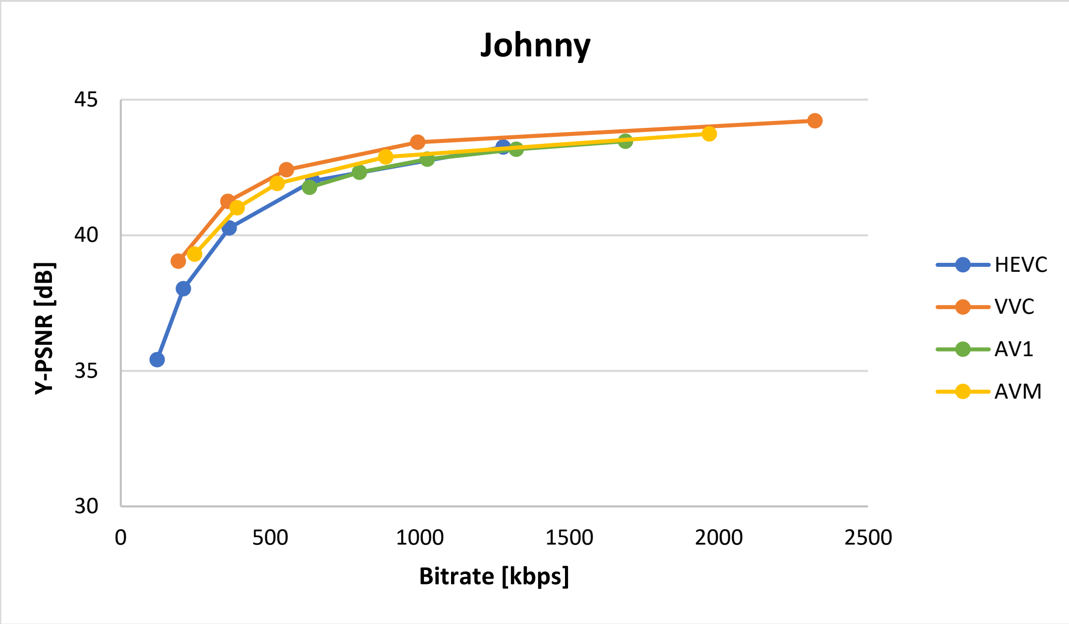


Figure 3.13: Johnny's Bitrate-PSNR graphic.

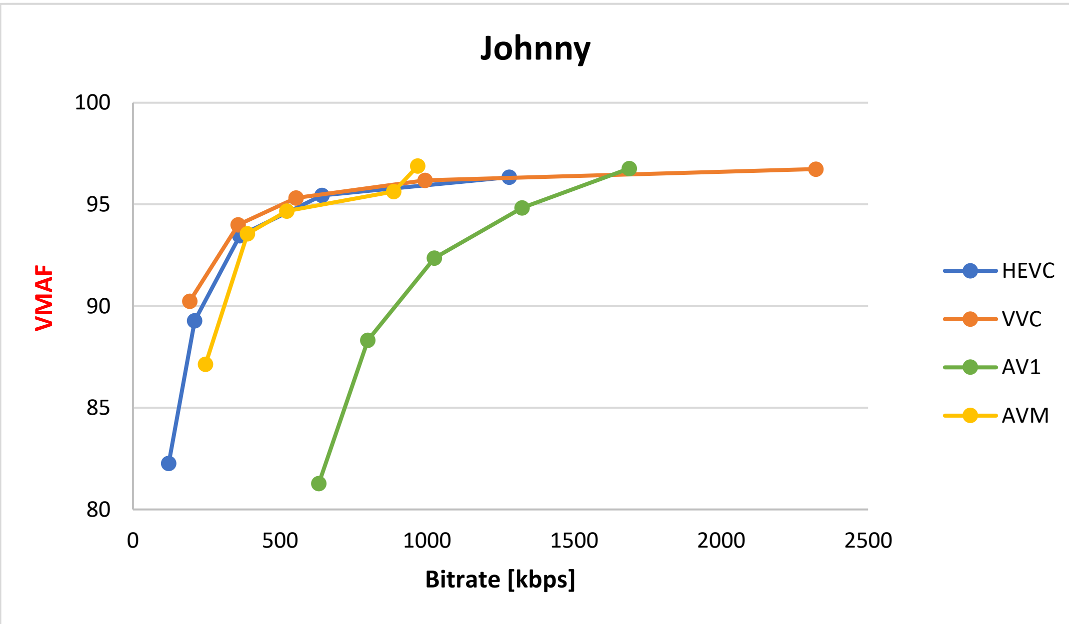


Figure 3.14: Johnny's Bitrate-VMAF graphic.

The trend in all the sequences is mighty similar. What immediately catches the eye is how HEVC handles this type of sequences, returning such lower PSNR values in comparison to other codecs but as the Bitrate increases, it converges with the other codecs.

AV1 is close to VVC and AVM, but slightly less efficient.

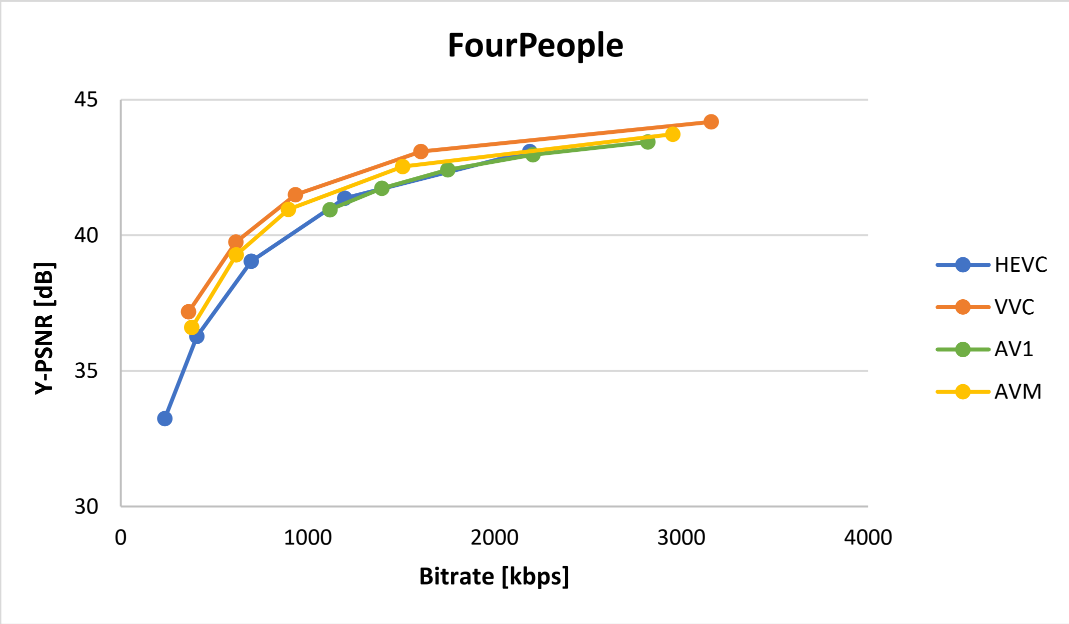


Figure 3.15: FourPeople's Bitrate-PSNR graphic.

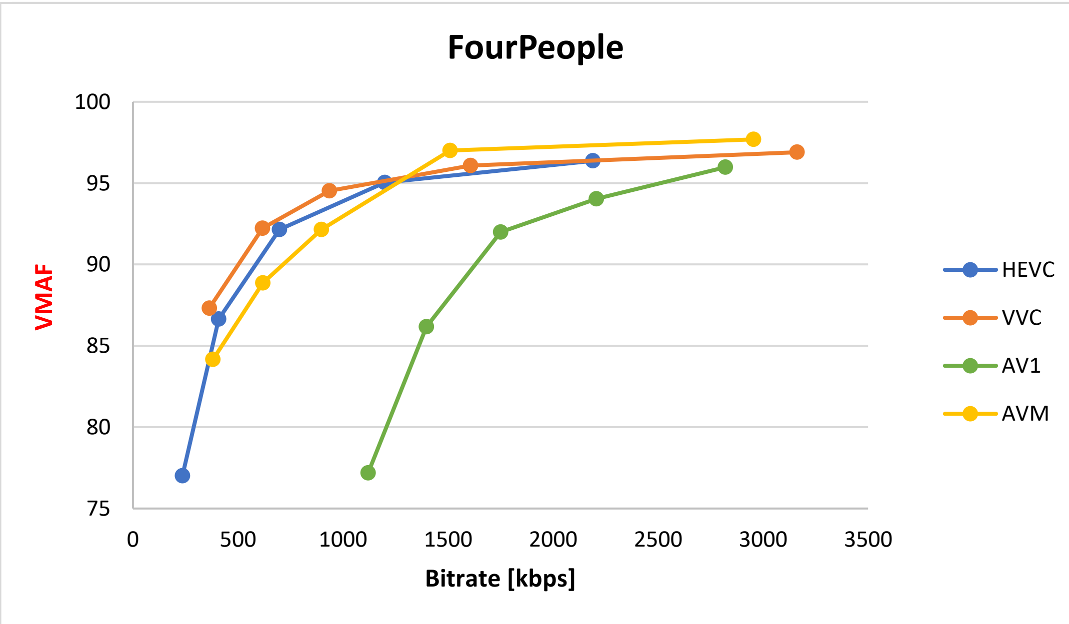


Figure 3.16 FourPeople's Bitrate-VMAF graphic.

Even referring to the VMAF's scores, VVC and AVM are the best choices for higher video quality with less bandwidth consumption by reaching the highest score at lower bitrates.

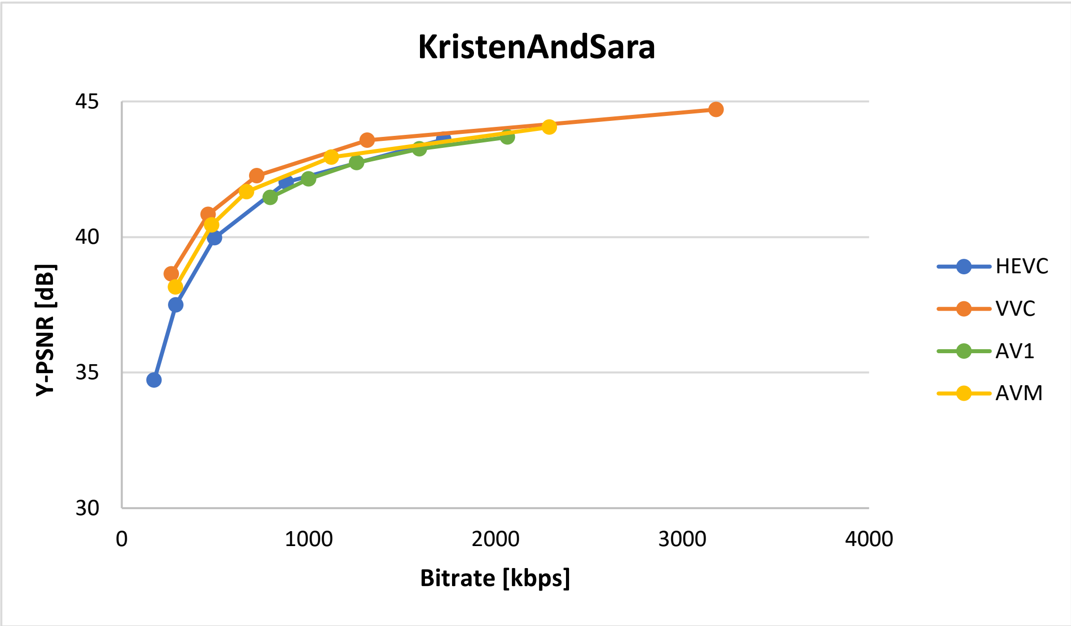


Figure 3.17: KristenAndSara's Bitrate-PSNR graphic.

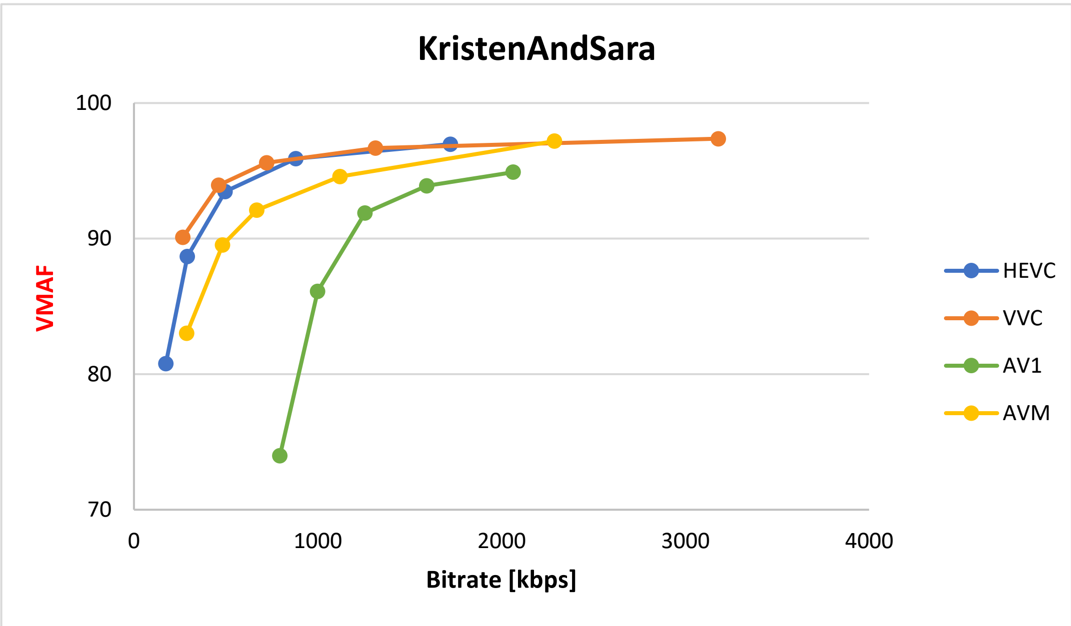


Figure 3.18: KristenAndSara's Bitrate-VMAF graphic.

For static scenes with speaking individuals, VVC and AVM are the most efficient codecs, achieving high quality with lower bitrate requirements. HEVC is the least efficient, requiring significantly higher bitrates for comparable quality, while AV1 falls in between but remains close to VVC and AVM.

In figure 3.15, AVM and VVC behave almost the same, meanwhile in figure 3.17, VVC, reaches a higher PSNR for way higher bitrates.

Even the BD-Rate value shows a gain towards AVM. In fact, in FourPeople the BD-Rate is 15,30%, in Johnny is 14,50% and in KristenAndSara is 17%.

3.2 Complexity profiling

To yield a more complete study, a complexity profiling[13] has been done on AV1 and the newest AVM codec. The purpose of this profiling was to:

- **Identify critical issues**: understand which parts of the code or functions need more time to be executed, in order to optimize the code;
- **Evaluate complexity**: give a more detailed view on the computational complexity;
- **Understand the general trend**: get to know how functions interact with each other, by obtaining an execution flux.

To get the profiling results back with an encoding process, it is necessary to enable the `gprof` and `-pg` command from the command line by:

- `cmake -DCONFIG_GPROF=1 -DCMAKE_C_FLAGS="-pg" -DCMAKE_CXX_FLAGS="-pg” -DCMAKE_EXE_LINKER_FLAGS="-pg" / make` for AV1
- `cmake -DCONFIG_GPROF=1 -DCMAKE_C_FLAGS="-pg" -DCMAKE_CXX_FLAGS="-pg" / make` for AVM

[13] *It is a performance analysis of the code that measures the execution time of each called function.*

These commands allow the profiling activation while compiling the compilation files. When the encoding process comes to an end, a **gmon.out** file is created[14]. This last mentioned contains all the information described so far and it must be converted into a .txt file, in order to be read.

Flat Profile is a type of output generated by code profiling tools (like gprof). Here, every row shows the called function and in the different columns there are information about:

- <u>% time</u>: time percentage occupied by the function.
- <u>Cumulative seconds</u>: cumulated time until that function.
- <u>Self seconds</u>: time taken by the function excluding calls to other functions.
- <u>Calls</u>: number of calls to the function
- <u>s/call</u>: avarange time per single call.
- <u>Name</u>: name of the function.

For the profiling analysis, the video sequence MarketPlace[15] is taken as reference for both encoders.

First, AV1 was profiled. Here are the 10 AV1 functions that take the longest to execute:

1) **av1_optimize_txb**: optimizes transform blocks (TXB), looking for the best frequency representations of the coefficients. It is an expensive and central operation in compression.
2) **obmc_diamond_search_sad**: search for the best motion vector with "diamond" strategy and OBMC (Overlapped Block Motion Compensation), using SAD[16] metrics.
3) **aom_sad16x16x4d_avx2**: calculation of SAD between 16x16 blocks on 4 candidates, optimized with AVX2[17].
4) **upsampled_pref_error**: calculation of predictive error on upsampled images. Used in comparison between prediction and original block.

[14] *The encoding process does not change from the others described, because the activation of the profiling just allows the creation of the gmon.out file.*
[15] *The analysis was done considering the encoding of only the first 100 frames out of 600 and with 42 (for AV1) and 210 (for AVM) as QP values.*
[16] *SAD: Sum of Absolute Reference.*
[17] *AVX2: Advanced Vector Extension (AVX) is a more advanced CPU instruction set, which basically allows the CPU to execute the same instruction on multiple pieces of data.*

5) **aom_convolve8_horiz_avx2**: horizontal convolution (8-tap filter) on images, implemented with AVX2. Part of filter and interpolation operations.
6) **aom_convolve8_vert_avx2**: as above, but vertical convolution.
7) **aom_upsampled_pred_sse2**[18]: generation of upsampled predictions from reference frames.
8) **aom_masked_sad32x32_avx2**: calculation of SAD on 32x32 blocks with masks, used for compound predictions.
9) **aom_comp_mask_pred_avx2**: composite prediction with weighted masks. Key function for inter-composite modes.
10) **av1_wedge_sse_from_residuals_avx2**: calculation of similarity for wedge masks (used in composite prediction) from residuals.

Function	% Time	Self seconds	Calls
av1_optimize_txb	10,52%	8,10 s	42669774
obmc_diamond_search_sad	5,61%	4,32 s	25047431
aom_sad16x16x4d_avx2	4,20%	3,23 s	117307565
upsampled_pref_error	3,66%	2,82 s	174773756
aom_convolve8_horiz_avx2	3,34%	2,57 s	114341103
aom_convolve8_vert_avx2	2,62%	2,02 s	116074115
aom_upsampled_pred_sse2	1,97%	1,52 s	175158457
aom_masked_sad32x32_avx2	1,81%	1,39 s	35060217
aom_comp_mask_pred_avx2	1,59%	1,22 s	72221688
av1_wedge_sse_from_residuals_avx2	1,59%	1,22 s	59070656

Table 3.4: summary of the AV1 Flat Profile's results.

In first place is av1_optimize_txb, which alone takes up more than 10% of the total time. This is a critical function within the video encoding process, because it deals with

[18] *SSE: Streaming SIMD Extensions is a single instruction, multiple data (SIMD) instruction set extension to the x86 architecture.*

optimizing the transformed blocks (TXB), i.e., those blocks that contain the coefficients of the discrete transform (typically DCT). The goal is to find the most efficient representation while minimizing data loss and size. Although it is not called as many times as other functions, each execution of it is particularly “heavy,” hence the high overall time.

Next comes obmc_diamond_search_sad, a function dedicated to motion estimation, which uses a “diamond” search to find the best motion vectors within the overlapping motion compensation (OBMC). This is also a key step in inter-frame video coding, as it allows reducing temporal redundancy between frames.

It is interesting to note that the very two functions most frequently called throughout the profiling are **upsampled_pref_error** and **aom_upsampled_pred_sse2**. Although they are not the most “expensive” in terms of absolute time, the fact that they are executed hundreds of millions of times indicates how central they are to the encoder's operating cycle. They represent two fundamental components of the inter-frame prediction phase, demonstrating how upsampling and prediction quality assessment are massively repeated operations, but optimized to be extremely light and fast. This explains why, although they have a very high number of calls, their computational weight remains relatively low.

Then, by examining the Flat Profile obtained, it can be said the 10 AVM functions with the greatest impact on performance and the highest time percentage are:

1) **av1_highbd_warp_affine_sse4_1** : implements the affine transform for motion compensation in high-bit-depth images, exploiting SSE4.1 instructions. It is one of the most expensive operations in video encoding.
2) **fwd_stxfm_avx2**: performs the forward transform on image blocks, a key step in AV1 compression. Optimized with AVX2 for faster speed.
3) **search_tx_type.constprop.0**: determines the best transform type for each block, a critical operation for output quality.
4) **aom_highbd_sad16x16x4d_avx2**: calculates the absolute difference between image blocks, used in motion estimation and motion compensation mechanisms.
5) **av1_convolve_symmetric_highbd_avx2**: performs convolution on image blocks to apply preprocessing filters.

6) **av1_wedge_sse_from_residuals_avx2**: uses residuals to optimize the selection of wedge masks, improving the quality of inter-frame prediction.
7) **refinemv_highbd_pad_mc_border**: refines the calculation of motion vectors by applying padding to edges to improve the accuracy of motion estimation.
8) **make_masked_inter_predictor**: creates a masked inter-frame predictor by combining multiple frame references with specific weights.
9) **compute_distortion_block_avx2**: calculates the distortion of image blocks, which is useful for selecting the optimal compression mode.
10) **av1_highbd_dist_wtd_convolve_2d_avx2**: is an advanced function that performs two-dimensional convolution with weights based on the distance between reference frames, improving inter-frame prediction and visual quality in AVM encoding.

Function	% Time	Self seconds	Calls
av1_highbd_warp_affine_sse4_1	9,83%	15,55 s	35477615
fwd_stxfm_avx2	7,26%	11,48 s	83796072
search_tx_type.constprop.0	2,57%	4,07 s	7768310
aom_highbd_sad16x16x4d_avx2	2,56%	4,05 s	134458243
av1_convolve_symmetric_highbd_avx2	2,36%	3,73 s	187965600
av1_wedge_sse_from_residuals_avx2	2,29%	3,62 s	19635286
refinemv_highbd_pad_mc_border	2,23%	3,53 s	77817124
make_masked_inter_predictor	2,14%	3,38 s	7503777
compute_distortion_block_avx2	2,09%	3,31 s	4311510
av1_highbd_dist_wtd_convolve_2d_avx2	1,95%	3,08 s	33030070

Table 3.5: summary of the AVM Flat Profile's results.

The data shows that the most called function is fwd_stxfm_avx2 with 83796072 calls. However, the function that takes the most execution time is av1_highbd_warp_affine_sse4_1 which accounts for 9,83% of the total time, with 15,55 seconds of exclusive execution time and 35477615 calls. This function is related to motion compensation, a computationally expensive operation in video codecs because it involves affine transformations at high bit depth. This would explain why, while it is not the most called function, it is still the one with the greatest impact in terms of computation time.

Analysis of the data shows that although the functions with the highest percentage of execution time (% Time) are complex operations such as av1_highbd_warp_affine_sse4_1 or fwd_stxfm_avx2, the most frequently called functions are others, such as subtract_8x8 (348144445 calls) and aom_count_primitive_quniform (290276269 calls).

The **subtract_8x8** function is involved in calculating the difference between pixel blocks, an essential step in video compression to determine the residual to be encoded. This explains why it is executed so many times: each image block must be processed individually to improve encoding efficiency.

Similarly, **aom_count_primitive_quniform**, with 290 million calls, appears to be an operation related to uniform quantization of transform coefficients, another crucial step in compression to reduce the amount of data needed to represent the image without compromising visual quality too much. **aom_highbd_sad16x16_avx2** (259 million calls) is also a key function, as it calculates the SAD between image blocks. This metric is critical for motion estimation, as it helps identify the best reference block between frames, reducing the overall bitrate of the video.

In general, these functions, while occupying a relatively small percentage of the total execution time, are called many times because they operate on small blocks of data repeatedly during the encoding process. This highlights the trade-off between the number of calls and the impact on overall processing time.

3.2.1 Comparison of profiling of AV1 and AVM

Analyzing the profiling results of AV1 and AVM, some significant differences emerge in the distribution of computational time and frequency of calls to the heavier functions.

Both codecs have a similar computational structure, with highly optimized functions exploiting AVX2 and SSE4.1 instructions, but the most expensive operations differ in some key aspects.

In AVM profiling, the most expensive function is av1_highbd_warp_affine_sse4_1, which accounts for 9,83% of the total time and is called 35M times. This function is related to motion compensation with affine transformations, indicating that AVM devotes a significant part of its computation to fitting blocks between frames. In AV1, on the other hand, the most expensive function was av1_optimize_txb, which deals with the optimization of the quantization and transform of the coefficients, accounting for more than 10% of the total time. This suggests that AV1 places more emphasis on compression efficiency at the transform stage, while AVM invests more in motion prediction.

Another notable difference is in the number of calls. In AVM, the function fwd_stxfm_avx2 has a very high number of executions (83M calls) with a significant cost of 7,26%, indicating that the Forward Transform phase is one of the most critical operations. In contrast, in the profiling of AV1, the highest number of calls belonged to functions such as upsampled_pref_error and aom_upsampled_pred_sse2, which while extremely frequent, had a smaller impact on total time. This shows that AV1 performs a lot of light and repeated operations in prediction, while AVM seems to have a more concentrated load distribution on specific functions.

An interesting comparison concerns filtering and convolution operations. In AVM, we find functions such as av1_convolve_symmetric_highbd_avx2, which takes up 2,36% of the time, while in AV1 similar functions (e.g., aom_convolve8_horiz_avx2) were higher in the ranking. This suggests that AV1 depends more on horizontal and vertical filtering operations, probably to improve prediction quality and reduce visual artifacts.

Finally, functions such as make_masked_inter_predictor and refinemv_highbd_pad_mc_border appear among the top 10 in AVM, but did not have such a large impact in AV1. This may indicate that AVM devotes more resources to improving inter-frame prediction through compensation techniques and edge padding, while AV1 optimizes compression more through transform processing.

Conclusions

The main objective of this thesis work was to investigate the state of the art in the field of video compression through a comparative analysis of four of the most recent and relevant codecs: HEVC (H.265), VVC (H.266), AV1 and AVM (AOMedia Video Model). The focus was particularly on the AVM codec, the most recent of those considered, in order to understand whether it actually succeeded in achieving the promised results in terms of compression efficiency, visual quality, and computational complexity.

From the experimental results that emerged, it can be said that AVM represents a concrete evolution from AV1, offering better compression efficiency, as demonstrated by the negative BD-Rate values in most of the tested sequences. In particular, AVM showed a significant average gain over its predecessor, demonstrating a good ability to adapt to both high and low visual complexity content.

However, the comparison with VVC revealed a more multifaceted picture. While AVM comes very close to VVC's performance, in some cases even surpassing it (such as in low-motion sequences), limitations emerge in the compression of high-motion content, where VVC continues to hold an advantage, both in terms of perceived quality and required bitrate. Profiling analyses have shown highly run-time consuming functions, particularly in affine transforms and inter-frame prediction on high bit-depth content. This indicates that although AVM is highly qualitative in performance, it requires non-negligible computational power, which may limit its deployment on low-power devices or in real-time scenarios.

Alternatively, the positive side of AVM is in the fact that it is royalty-free technology, which makes it an extremely wanted contender for bulk adoption, mainly by web and mobile segment applications and streaming organizations.

Hence, it can be said that AVM has adequately fulfilled the objectives, registering concrete improvements compared to AV1 and approaching VVC performance. However, there are outstanding issues, particularly its high computational complexity, which will need to be addressed to obtain its complete and universal adoption. In mind its young age and continuous growth of the project, it is reasonable to expect that subsequent iterations of the codec will continue to enhance on performance and efficiency, solidifying its place among the leading standards in the field of video compression. In the coming years, the video codecs that are discussed in this thesis not only represent the video compression state of the art, but also occupy a strategic role in shaping the distribution and consumption of multimedia content in the next few years. Market needs are evolving very rapidly with the viewing of high-resolution content (4K, 8K) increasing, and the proliferation of immersive experience such as virtual reality and augmented reality, and low-latency over-the-top (OTT) delivery on mobile networks. In such a scenario, the AVMs and VVCs will play a key role in meeting these new demands since they can preserve high visual quality at low bitrates. In particular, AVM, with its royalty-free approach, may prove to be an enabling technology for web-based applications and mobile devices, where scalability, lightness, and accessibility play a crucial role.

In addition, the advent of artificial intelligence in streaming, video analytics, and surveillance systems implies massive use of highly efficient visual streams, which need versatile codecs optimized for integration with machine learning models. In this sense, standardization and deployment of adaptive and open-source codecs such as AVM could accelerate innovation, facilitating adoption in emerging contexts such as the metaverse, cloud gaming, and video transmission in 5G and future networks (6G).

At the same time, it is foreseeable that the evolution will be increasingly oriented toward personalization of video quality through perceptual metrics (such as VMAF) and artificial intelligence-based optimization algorithms. Codecs of the future will not only be more efficient, but also more “intelligent,” capable of adapting video quality in real time according to the context of use, the available network, and user preferences. The thesis highlighted how the choice of codec is never neutral, but depends on a complex balance between quality, compression, computational complexity and economic constraints. Thoroughly understanding these dynamics is critical to meeting the future challenges of audiovisual distribution and to designing solutions that are truly sustainable and efficient in the long term.

In conclusion, it can be expected that the competition between supposed royalty-free solutions (such as AVM) and formalized standards (such as VVC) will shape a hybrid codec ecosystem, in which multiple solutions will coexist depending on specific needs. The key will be interoperability, power efficiency, and the ability to evolve rapidly to keep up with the unstoppable digital transformation.

Appendix

1) HEVC:

Sequence	QP	Bitrate	Y-PSNR	VMAF Score
ArenaOfValor	22	14390,83	42,98	99,34
ArenaOfValor	27	6594,78	39,59	96,42
ArenaOfValor	32	3352,03	36,65	90,40
ArenaOfValor	37	1745,10	33,89	81,14
ArenaOfValor	42	919,54	31,23	69,19
BasketballDrill	22	3355,90	40,91	99,75
BasketballDrill	27	1605,25	37,70	96,52
BasketballDrill	32	826,52	34,79	88,85
BasketballDrill	37	428,51	32,13	77,46
BasketballDrill	42	224,60	29,65	64,07
BasketballDrive	22	12542,81	39,03	99,81
BasketballDrive	27	5115,25	37,42	99,32
BasketballDrive	32	2596,31	35,65	94,04
BasketballDrive	37	1348,73	33,60	84,53
BasketballDrive	42	708,76	31,44	73,25
BasketballPass	22	1436,60	40,79	99,36
BasketballPass	27	684,86	36,79	96,64
BasketballPass	32	350,06	33,56	87,66
BasketballPass	37	179,24	30,67	73,99
BasketballPass	42	91,96	28,12	59,03
BlowingBubbles	22	1486,31	38,08	96,94
BlowingBubbles	27	698,62	34,87	93,67
BlowingBubbles	32	349,61	31,98	87,53

BlowingBubbles	37	168,94	29,17	77,77
BlowingBubbles	42	80,22	26,59	64,35
BQMall	22	3319,13	40,29	99,80
BQMall	27	1616,50	37,79	99,40
BQMall	32	854,11	35,16	96,89
BQMall	37	457,80	32,43	88,96
BQMall	42	245,10	29,72	77,06
BQSquare	22	1608,53	37,84	99,14
BQSquare	27	619,38	34,45	97,77
BQSquare	32	312,44	31,99	95,04
BQSquare	37	175,78	29,63	89,97
BQSquare	42	99,40	27,01	81,38
BQTerrace	22	22600,41	36,88	98,38
BQTerrace	27	5082,49	35,31	97,31
BQTerrace	32	2394,59	34,07	95,70
BQTerrace	37	1245,80	32,44	90,84
BQTerrace	42	653,76	30,37	81,96
Cactus	22	11473,99	38,40	98,96
Cactus	27	5091,94	36,96	96,20
Cactus	32	2638,02	35,14	91,07
Cactus	37	1372,84	32,94	82,55
Cactus	42	716,16	30,65	71,08
FoodMarket4	22	19401,85	43,96	93,15
FoodMarket4	27	9901,15	42,44	92,99
FoodMarket4	32	5441,95	40,45	91,32
FoodMarket4	37	2992,09	38,04	89,56
FoodMarket4	42	1662,01	35,45	86,61
FourPeople	22	2189,15	43,10	96,37

FourPeople	27	1198,69	41,37	95,04
FourPeople	32	698,40	39,04	92,16
FourPeople	37	408,21	36,27	86,65
FourPeople	42	235,44	33,24	77,01
Johnny	22	1279,91	43,25	96,34
Johnny	27	642,83	42,00	95,44
Johnny	32	363,04	40,27	93,47
Johnny	37	209,44	38,03	89,28
Johnny	42	121,66	35,41	82,27
KristenAndSara	22	1722,75	43,60	96,97
KristenAndSara	27	880,37	42,03	95,90
KristenAndSara	32	495,94	39,97	93,48
KristenAndSara	37	289,93	37,49	88,68
KristenAndSara	42	173,48	34,72	80,78
MarketPlace	22	14041,34	40,57	99,28
MarketPlace	27	6340,13	38,54	97,39
MarketPlace	32	3073,87	36,38	91,40
MarketPlace	37	1455,69	34,09	81,09
MarketPlace	42	680,01	31,81	67,85
PartyScene	22	6442,34	38,16	97,26
PartyScene	27	2925,20	34,72	93,86
PartyScene	32	1469,31	31,86	87,48
PartyScene	37	719,52	29,12	79,43
PartyScene	42	333,83	26,43	65,59
RaceHorses	22	1101,75	39,40	99,95
RaceHorses	27	523,67	35,55	98,61
RaceHorses	32	262,56	32,31	90,40
RaceHorses	37	128,88	29,42	76,67

RaceHorses	42	63,99	27,05	61,93
RaceHorses	22	4173,16	38,87	99,98
RaceHorses	27	1692,66	35,61	99,17
RaceHorses	32	828,26	32,90	92,19
RaceHorses	37	406,28	30,22	79,36
RaceHorses	42	190,62	27,83	65,38
RitualDance	22	11156,89	44,11	99,84
RitualDance	27	5665,75	40,96	98,50
RitualDance	32	3041,78	37,92	92,06
RitualDance	37	1604,14	34,98	80,23
RitualDance	42	830,38	32,27	66,30
SlideEditing	22	1302,16	49,85	97,53
SlideEditing	27	1001,37	45,70	96,92
SlideEditing	32	760,08	41,04	95,55
SlideEditing	37	586,42	36,18	92,25
SlideEditing	42	436,08	31,32	84,69
SlideShow	22	1598,93	50,97	98,38
SlideShow	27	952,84	47,01	98,01
SlideShow	32	611,09	43,37	97,10
SlideShow	37	401,02	39,69	95,07
SlideShow	42	260,21	35,84	91,44

2) VVC:

Sequence	QP	Bitrate [kbps]	Y-PSNR [dB]	VMAF Score
ArenaOfValor	22	13178,12	43,54	99,23
ArenaOfValor	27	6423,10	40,48	96,48
ArenaOfValor	32	3193,40	37,60	90,86
ArenaOfValor	37	1827,49	35,26	84,12
ArenaOfValor	42	952,13	32,85	74,35
BasketballDrill	22	3430,06	42,21	99,77
BasketballDrill	27	1736,75	39,40	98,00
BasketballDrill	32	841,02	36,35	91,33
BasketballDrill	37	455,59	33,84	82,64
BasketballDrill	42	215,19	30,94	68,95
BasketballDrive	22	8342,10	38,91	99,77
BasketballDrive	27	3911,10	37,56	98,73
BasketballDrive	32	1858,81	35,72	92,11
BasketballDrive	37	1005,48	33,88	83,79
BasketballDrive	42	486,49	31,52	72,04
BasketballPass	22	1691,82	42,75	99,54
BasketballPass	27	841,88	38,81	98,14
BasketballPass	32	398,75	35,13	91,46
BasketballPass	37	212,34	32,34	80,74
BasketballPass	42	99,05	29,20	63,27
BlowingBubbles	22	1503,88	39,05	96,57
BlowingBubbles	27	741,51	36,10	93,81
BlowingBubbles	32	368,80	33,19	88,56
BlowingBubbles	37	208,81	30,88	81,74
BlowingBubbles	42	95,21	28,05	69,80
BQMall	22	2867,16	40,54	99,79

BQMall	27	1440,27	38,27	99,27
BQMall	32	736,42	35,63	96,60
BQMall	37	431,16	33,36	90,61
BQMall	42	222,58	30,57	79,46
BQSquare	22	1251,92	38,09	98,36
BQSquare	27	542,69	35,27	96,78
BQSquare	32	280,66	32,82	94,21
BQSquare	37	176,18	30,86	91,00
BQSquare	42	96,81	28,18	84,16
BQTerrace	22	9343,99	36,27	97,63
BQTerrace	27	3299,77	35,19	96,49
BQTerrace	32	1595,81	33,96	94,15
BQTerrace	37	921,75	32,66	89,99
BQTerrace	42	471,10	30,75	81,85
Cactus	22	8417,09	38,42	98,38
Cactus	27	3897,68	37,15	95,60
Cactus	32	1920,31	35,37	90,34
Cactus	37	1082,42	33,60	83,83
Cactus	42	526,84	31,30	72,55
FoodMarket4	22	17160,82	44,21	43,80
FoodMarket4	27	8959,04	43,08	43,45
FoodMarket4	32	4654,24	41,29	42,29
FoodMarket4	37	2661,33	39,37	39,61
FoodMarket4	42	1312,36	36,61	34,98
FourPeople	22	3161,48	44,18	96,90
FourPeople	27	1606,92	43,09	96,08
FourPeople	32	935,91	41,50	94,53
FourPeople	37	616,61	39,75	92,22

FourPeople	42	363,52	37,18	87,32
Johnny	22	2322,88	44,22	96,74
Johnny	27	993,79	43,43	96,19
Johnny	32	554,91	42,42	95,32
Johnny	37	358,28	41,24	94,00
Johnny	42	193,12	39,04	90,24
KristenAndSara	22	3180,72	44,70	97,37
KristenAndSara	27	1314,55	43,57	96,68
KristenAndSara	32	723,02	42,26	95,60
KristenAndSara	37	462,20	40,83	93,95
KristenAndSara	42	265,65	38,64	90,09
MarketPlace	22	12059,34	40,96	99,15
MarketPlace	27	5180,10	39,16	96,83
MarketPlace	32	2396,89	37,17	91,11
MarketPlace	37	1281,94	35,35	84,01
MarketPlace	42	602,68	33,04	73,10
PartyScene	22	4689,66	37,94	98,53
PartyScene	27	2289,22	34,87	95,28
PartyScene	32	1099,61	31,85	89,32
PartyScene	37	595,39	29,52	79,48
PartyScene	42	260,97	26,71	66,02
RaceHorses	22	1199,79	40,56	99,98
RaceHorses	27	621,96	37,22	99,57
RaceHorses	32	285,61	33,57	94,23
RaceHorses	37	145,35	30,80	82,75
RaceHorses	42	59,87	27,66	64,42
RaceHorses	22	3530,45	38,71	99,97
RaceHorses	27	1605,45	36,02	99,35

RaceHorses	32	712,55	33,03	92,83
RaceHorses	37	352,53	30,54	81,73
RaceHorses	42	139,22	27,81	64,89
RitualDance	22	10590,22	45,00	99,94
RitualDance	27	5615,58	42,16	99,15
RitualDance	32	2901,56	38,98	93,11
RitualDance	37	1637,95	36,26	83,91
RitualDance	42	780,13	33,1699	70,89
SlideEditing	22	1302,16	49,85	97,29
SlideEditing	27	1001,37	45,70	96,59
SlideEditing	32	760,08	41,04	94,50
SlideEditing	37	586,42	36,18	90,81
SlideEditing	42	436,08	31,32	81,88
SlideShow	22	1598,93	50,97	98,41
SlideShow	27	952,84	47,01	98,23
SlideShow	32	611,09	43,37	97,66
SlideShow	37	401,02	39,69	96,46
SlideShow	42	260,21	35,84	93,84

3) AV1:

Sequence	QP	Bitrate [kbps]	Y-PSNR [dB]	VMAF Score
ArenaOfValor	27	15661,96	43,54	98,59
ArenaOfValor	32	11537,24	42,15	95,99
ArenaOfValor	37	8429,45	40,78	87,14
ArenaOfValor	42	6167,99	39,48	81,05
ArenaOfValor	47	4608,69	38,2400	65,07
BasketballDrill	27	3685,46	41,68	99,23
BasketballDrill	32	2773,44	40,48	96,52
BasketballDrill	37	2050,65	39,15	88,85
BasketballDrill	42	1505,12	37,80	80,46
BasketballDrill	47	1114,53	36,5100	69,07
BasketballDrive	27	10517,37	38,85	99,77
BasketballDrive	32	7497,89	38,31	98,73
BasketballDrive	37	5444,21	37,72	91,21
BasketballDrive	42	4016,63	37,07	82,79
BasketballDrive	47	3001,29	36,3600	74,54
BasketballPass	27	1352,35	40,81	95,46
BasketballPass	32	1005,66	39,18	93,72
BasketballPass	37	747,20	37,64	89,27
BasketballPass	42	552,65	36,15	76,57
BasketballPass	47	411,95	34,7700	58,36
BlowingBubbles	27	1498,68	38,23	99,27
BlowingBubbles	32	1167,28	37,16	91,8
BlowingBubbles	37	893,78	35,99	89,79
BlowingBubbles	42	676,63	34,76	86,27
BlowingBubbles	47	511,32	33,5200	70,1

BQMall	27	3295,92	40,32	97,80
BQMall	32	2559,06	39,51	96,11
BQMall	37	1958,16	38,55	95,90
BQMall	42	1509,03	37,55	86,60
BQMall	47	1170,02	36,5	75,60
BQSquare	27	1378,80	37,73	99,37
BQSquare	32	1015,80	36,68	97,00
BQSquare	37	761,21	35,65	95,66
BQSquare	42	584,88	34,70	88,92
BQSquare	47	451,57	33,7	80,80
BQTerrace	27	14293,24	36,49	99,02
BQTerrace	32	8881,58	36,02	98,10
BQTerrace	37	6533,73	35,65	95,22
BQTerrace	42	4920,77	35,27	91,20
BQTerrace	47	3757,49	34,84	81,99
Cactus	27	10596,91	38,31	98,75
Cactus	32	8027,98	37,87	95,89
Cactus	37	6046,76	37,32	91,96
Cactus	42	4500,28	36,68	82,51
Cactus	47	3439,56	35,97	70,83
FoodMarket4	27	15110,30	43,44	99,93
FoodMarket4	32	11558,97	42,81	98,60
FoodMarket4	37	8774,72	42,03	93,38
FoodMarket4	42	4559,96	40,01	81,99
FoodMarket4	47	6547,11	36,4	73,88
FourPeople	27	2821,14	43,44	95,98
FourPeople	32	2206,73	42,97	94,04
FourPeople	37	1750,70	42,42	92,00

FourPeople	42	1397,05	41,74	86,17
FourPeople	47	1120	40,95	77,20
Johnny	27	1688,38	43,47	96,77
Johnny	32	1323,52	43,17	94,84
Johnny	37	1025,26	42,80	92,37
Johnny	42	798,53	42,32	88,32
Johnny	47	631,9	41,77	81,27
KristenAndSara	27	2064,32	43,68	94,90
KristenAndSara	32	1593,41	43,25	93,90
KristenAndSara	37	1257,91	42,74	91,88
KristenAndSara	42	999,77	42,15	86,10
KristenAndSara	47	794,3	41,46	73,99
MarketPlace	27	12842,82	40,62	99,88
MarketPlace	32	9628,85	39,92	97,49
MarketPlace	37	7040,29	39,10	91,55
MarketPlace	42	5103,75	38,21	80,16
MarketPlace	47	3704,55	37,27	70,55
PartyScene	27	6026,55	38,33	98,96
PartyScene	32	4597,99	37,17	95,28
PartyScene	37	3476,11	35,94	89,32
PartyScene	42	2603,93	34,69	79,48
PartyScene	47	1944,89	33,41	66,02
RaceHorses	27	1019,16	39,16	99,15
RaceHorses	32	770,50	37,69	97,88
RaceHorses	37	575,65	36,20	90,99
RaceHorses	42	427,97	34,77	77,10
RaceHorses	47	318,05	33,38	62,56
RaceHorses	27	3519,57	38,50	99,79

RaceHorses	32	2453,37	37,23	99,20
RaceHorses	37	1757,76	36,05	91,93
RaceHorses	42	1301,12	34,93	79,13
RaceHorses	47	966,84	33,78	67,44
RitualDance	27	10883,53	44,43	99,87
RitualDance	32	8269,58	43,20	98,50
RitualDance	37	6192,20	41,84	92,60
RitualDance	42	4595,71	40,42	81,01
RitualDance	47	3.408,84	38,97	69,11
SlideEditing	27	1781,16	54,33	96,52
SlideEditing	32	1562,91	52,49	94,23
SlideEditing	37	1366,44	50,56	93,18
SlideEditing	42	1174,58	48,41	92,23
SlideEditing	47	1004,82	46,23	84,22
SlideShow	27	1658,50	53,48	98,38
SlideShow	32	1322,46	51,54	98,01
SlideShow	37	1058,84	49,71	97,10
SlideShow	42	848,22	47,97	95,07
SlideShow	47	688,07	46,39	91,44

4) AVM:

Sequence	QP	Bitrate [kbps]	Y-PSNR [dB]	VMAF Score
ArenaOfValor	110	16930,88	44,37	99,97
ArenaOfValor	135	7883,96	41,23	96,55
ArenaOfValor	160	3996,08	38,53	87,96
ArenaOfValor	185	2172,14	36,06	81,46
ArenaOfValor	210	794,61	32,36	66,90
BasketballDrill	110	3916,88	42,48	99,98
BasketballDrill	135	1920,73	39,69	97,54
BasketballDrill	160	935,22	36,82	89,66
BasketballDrill	185	494,76	34,13	81,64
BasketballDrill	210	182,00	30,61	70,02
BasketballDrive	110	16131,27	39,50	99,79
BasketballDrive	135	5935,34	38,20	98,73
BasketballDrive	160	2867,07	36,73	92,45
BasketballDrive	185	1481,89	34,94	83,14
BasketballDrive	210	551,34	32,16	75,04
BasketballPass	110	1611,88	42,47	99,66
BasketballPass	135	815,16	38,74	98,96
BasketballPass	160	410,06	35,40	93,47
BasketballPass	185	215,41	32,47	81,19
BasketballPass	210	94,34	29,14	63,73
BlowingBubbles	110	1841,50	39,44	99,71
BlowingBubbles	135	963,45	36,74	98,11
BlowingBubbles	160	523,65	34,11	96,61
BlowingBubbles	185	306,96	31,74	82,74
BlowingBubbles	210	191,28	29,34	73,82

BQMall	110	3938,50	41,08	99,83
BQMall	135	1974,10	39,11	99,27
BQMall	160	1070,03	36,87	95,26
BQMall	185	619,03	34,55	90,13
BQMall	210	278,02	31,27	73,45
BQSquare	110	1864,26	39,36	99,81
BQSquare	135	846,04	36,65	98,75
BQSquare	160	468,93	34,48	94,15
BQSquare	185	297,96	32,61	86,59
BQSquare	210	246,28	30,96	82,65
BQTerrace	110	30868,45	37,38	99,99
BQTerrace	135	7434,29	35,91	96,19
BQTerrace	160	3355,22	35,03	93,65
BQTerrace	185	1825,03	33,98	87,97
BQTerrace	210	756,69	31,82	80,62
Cactus	110	15879,38	38,78	98,86
Cactus	135	6077,78	37,68	94,73
Cactus	160	3108,60	36,31	90,22
Cactus	185	1699,60	34,62	83,96
Cactus	210	609,11	31,62	73,15
FoodMarket4	110	16443,22	44,38	97,15
FoodMarket4	135	15443,25	43,29	96,09
FoodMarket4	160	8020,53	41,76	94,23
FoodMarket4	185	1707,78	39,79	89,37
FoodMarket4	210	552,23	35,39	85,44
FourPeople	110	2954,63	35,39	97,69
FourPeople	135	1509,42	43,73	97,00
FourPeople	160	898,06	42,53	92,16

FourPeople	185	618,04	40,96	88,87
FourPeople	210	380,75	39,28	84,16
Johnny	110	968,70	36,60	96,89
Johnny	135	886,65	43,74	95,63
Johnny	160	523,71	42,89	94,67
Johnny	185	389,56	41,90	93,56
Johnny	210	246,79	41,01	87,14
KristenAndSara	110	2288,23	39,30	97,19
KristenAndSara	135	1121,35	44,05	94,57
KristenAndSara	160	667,91	42,95	92,11
KristenAndSara	185	481,74	41,67	89,52
KristenAndSara	210	287,68	40,44	83,01
MarketPlace	110	16894,37	38,15	99,55
MarketPlace	135	7750,88	41,31	96,36
MarketPlace	160	3938,80	39,74	90,84
MarketPlace	185	2083,18	37,99	86,17
MarketPlace	210	584,06	36,17	70,16
PartyScene	110	7041,52	39,42	98,99
PartyScene	135	3637,68	36,60	97,64
PartyScene	160	1908,52	33,82	89,84
PartyScene	185	1023,60	31,21	79,93
PartyScene	210	402,00	27,89	67,96
RaceHorses	110	1229,04	40,79	99,92
RaceHorses	135	633,51	37,66	91,75
RaceHorses	160	329,76	34,16	93,31
RaceHorses	185	172,17	31,30	82,52
RaceHorses	210	65,65	27,97	65,00
RaceHorses	110	2216,16	38,71	99,93

RaceHorses	135	1962,38	36,79	97,33
RaceHorses	160	989,97	34,29	92,46
RaceHorses	185	522,22	31,81	83,96
RaceHorses	210	169,06	28,41	66,01
RitualDance	110	6313,73	44,89	99,94
RitualDance	135	3394,98	42,69	98,03
RitualDance	160	1837,24	39,79	94,36
RitualDance	185	586,84	36,86	87,10
RitualDance	210	182,32	32,14	73,86
SlideEditing	110	1085,61	53,51	99,99
SlideEditing	135	754,12	49,39	96,72
SlideEditing	160	574,63	45,67	94,05
SlideEditing	185	433,71	42,09	90,75
SlideEditing	210	267,08	34,90	81,18
SlideShow	110	1742,64	55,18	98,97
SlideShow	135	1013,03	50,98	97,83
SlideShow	160	629,37	47,55	97,04
SlideShow	185	422,59	44,57	95,46
SlideShow	210	238,90	39,86	90,33

Bibliography and sitography

[1] Exploring the Video Coding for Machines Standard: Current Status
and Future Directions - Dongmin Leea, Sangkyun Jeona, Yeonghun Jeonga, Joonsoo Kimb, and Jeongil Seoa

[2] Recent Standard Development Activities on Video Coding for Machines - Wen Gao, Shan Liu, Xiaozhong Xu, Manouchehr Rafie, Yuan Zhang, Igor Curcio, https://arxiv.org/abs/2105.12653

[3] Cisco Annual Internet Report (2018–2023) White Paper, https://www.cisco.com/c/en/us/solutions/collater al/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[4] The Ultimate Guide to Video Marketing - Alicia Collins, Megan Conley, https://blog.hubspot.com/marketing/video-marketing

[5] Alexa, The top 500 sites on web, https://gist.github.com/elimisteve/69077a93d21b8bf8a02a362c830fbcb1

[6] YouTube Official Blog, YouTube for Press, https://blog.youtu be/press

[7] The Mobile Internet Phenomena Report, May 2021, https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2021/Phenomena/MIPR%20Q1%202021%2020210510.pdf

[8] An Overview of Recent Video Coding Developments in MPEG and AOMEDIA - Andre Seixas Dias, Saverio Blasi, Fiona Rivera, Ebroul Izquierdo, Marta Mrak, https://zenodo.org/records/1494930

[9] https://www.iso.org/standards.html

[10] https://www.iso.org/about/members

[11] https://www.iso.org/stages-and-resources-for-standards-development.html

[12] https://www.itu.int/hub/membership/our-members/directory/?myitu-members-states=true&request=countries

[13] https://www.itu.int/hub/membership/our-members/directory/?myitu-industry=true&request=sector-members]

[14] https://www.itu.int/hub/membership/our-members/directory/?myitu-smes=true&request=associates

[15] https://www.itu.int/hub/membership/our-members/directory/?myitu-academia=true&request=academia

[16] https://www.itu.int/hub/membership/our-members/directory/?myitu-organisations=true&request=list-reg-int

[17] Draft text of ISO/IEC DIS 14496-1 5th edition Systems, https://www.mpeg.org/standards/MPEG-4/1/

[18] https://leonardo.chiariglione.org/5-3-how-mpeg-develops-standards/

[19] MPEG issues Call for Proposals for AI-based Point Cloud Coding, wg2n00365, https://www.mpeg.org/146th-meeting-of-mpeg/

[20] https://blog.chiariglione.org/how-does-mpeg-actually-work/

[21] Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics - Lingyu Duan, Jiaying Liu, Wenhan Yang, Tiejun Huang, Wen Gao, https://ieeexplore.ieee.org/document/9180095

[22] Codifica video: gli standard di compressione ISO/IEC MPEG - ITU-T - Paola Sunna, http://www.crit.rai.it/eletel/2014-1/141-4.pdf

[23] MPEG-2: where did it come from and what is it? - O.J. Morris

[24] Overview of the H.264/AVC video coding standard, Fraunhofer Institute for Telecommunications (Heinrich Hertz Institut), Berlin, Germany - T. Wiegand , G.J. Sullivan, G. Bjontegaard, A. Luthra

[25] Overview of the High Efficiency Video Coding (HEVC) Standard - Gary J. Sullivan, Jens-Rainer, Ohm, Woo-Jin Han, Thomas Wiegand

[26] https://www.dacast.com/it/il-blog-degli-esperti-di-video-dacast/hevc-video-codec/

[27] Performance Analysis of H.265/HEVC (High-Efficiency Video Coding) with Reference to Other Codecs - N. Minallah, S. Gul, M.M. Bokhari

[28] Overview of the Versatile Video Coding (VVC) Standard and Its Applications - Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle ChenGary j. Sullivan, Jens-Ranier Ohm

[29] https://www.dacast.com/blog/h266-vvc-versatile-video-coding/

[30] A Technical Overview of AV1 - Jingning Han, Bohan Li, Debargha Mukherjee, Han Chiang, Adrian Grange, Cheng Chen, Hui Su, Sarah Parker, Sai Deng, Urvang Joshi, Yue Chen, Yunqing Wang, Paul Wilkins, Yaowu Xu, James Bankoski

[31] https://www.coconut.co/articles/unveil-av2-codec-nextgen-video-streaming

[32] https://github.com/listenlink/HM

[33] https://github.com/fraunhoferhhi/vvenc

[34] https://aomedia.googlesource.com/aom

[35] https://gitlab.com/AOMediaCodec/avm

[36] https://www.coconut.co/articles/unveil-av2-codec-nextgen-video-streaming

[37] https://learn.microsoft.com/it-it/windows/win32/medfound/about-yuv-video

[38] https://www.offminor.de/

[39] https://ient.github.io/YUView/

[40] Comparison of Compression Efficiency between HEVC/H.265, VP9 and AV1 based on Subjective Quality Assessments

[41] Image quality metrics: PSNR vs. SSIM - Alain Horé - Djemel Ziou

[42] Bjøntegaard Delta (BD): A Tutorial Overview of the Metric, Evolution, Challenges, and Recommendations, Nabajeet Barman, Maria G. Martini, Yuriy Reznik

[43] https://www.coconut.co/articles/psnr-peak-signal-quality-measure

[44] https://vicuesoft.com/blog/titles/bd_rate_one_name_two_metrics/

[45] https://medium.com/innovation-labs-blog/bjontegaard-delta-rate-metric-c8c82c1bc42c

[46] https://netflixtechblog.com/vmaf-the-journey-continues-44b51ee9ed12

[47] https://italiarappginevra.esteri.it/it/litalia-e-ooii/diplomazia-scientifica/lunione-internazionale-delle-telecomunicazioni-itu/

[48] https://www.itu.int/en/ITU-T/about/Pages/default.aspx

[49] VVenC Fraunhofer Versatile Video Encoder v1.2.0 Jens Brandenburg, Adam Wieckowski, Tobias Hinz, Ivan Zupancic, Benjamin Bross

Acknowledgments

Prima di iniziare col ringraziare tutte le persone a me più vicine, ci tengo a fare un primo ringraziamento che talvolta risulta scontato.

Grazie alla vita. Grazie per avermi fatto nascere in questa famiglia (allargata e non), per avermi dato queste possibilità e per avermi indirizzato a compiere determinate scelte che mi hanno portato, a loro volta, a conoscere queste persone e a sviluppare determinate passioni e capacità, diventando così la persona che sono oggi.

Grazie alla mia famiglia: mia mamma, mio padre e mia sorella. Grazie mamma e papà per avermi dato le ali ed insegnato a volare, lasciandomi libera di intraprendere il mio cammino senza giudizio, standomi accanto come figure presenti, ma mai ingombranti. Grazie mamma per essere il mio lato più emotivo, e papà per essere quello più pratico. E grazie per avermi dato una compagna di vita, Sara, che sono sicura spiccherà il volo. A voi, devo tutto.

Grazie ai miei nonni, pietre miliari della mia vita. Grazie per esserci sempre stati e dell'amore intrinseco dimostrato con ogni minimo gesto. So che la distanza a volte può essere difficile, ma a volte basta un poco per poterla colmare. E per nonno Costantino, so che mi stai guardando da lì su, so quanto saresti stato fiero e non posso che ringraziarti infinitamente per essere stato la stella che mi ha guidato più di tutte le altre.

Parlando di famiglia, sento di dover aggiungere un ringraziamento speciale.

Grazie a Domenico. Ti cito qui, perché come mi dici tu sempre, "sei tu la mia famiglia". Grazie per essere un uomo con la U maiuscola. Grazie per avermi insegnato che l'amore non significa limiti, ma soprattutto nuovi orizzonti. Grazie per avermi visto bella sempre, per avermi tenuto la mano in ogni situazione e per non essermi stato mai né un passo avanti, né un passo indietro, ma sempre e comunque accanto. Grazie perché con te le calorie diventano solo numeri ed un volo con due o più ore di ritardo sembra poi non pesare così tanto. Per la prima volta ho davvero imparato che tutto è affrontabile, perché da quando ci sei tu nella mia vita, ogni più piccolo problema ma anche ogni più grande gioia, è sempre divisa per due.

Grazie alle mie coinquiline (presenti e passate), so che vivere con me a volte risulti una sfida, ma voi vi siete rivelate tutte delle gran giocatrici. In particolare ringrazio la coinquilina per eccellenza, Alessandra, la più longeva e forse la più temeraria. Grazie perché un inciucio mio diventa anche tuo, e come hai detto anche tu, essere coinquiline può non esser facile, ma noi sappiamo essere molto di più.

Grazie a tutti i miei amici, torinesi e non. Mi sento immensamente fortunata ad avervi, perché so quanto non sia scontato. Grazie, in particolare, a Luca e Lucy, i miei miglior amici per antonomasia nonché fratelli mancati. Grazie, perché lo scorrere degli anni sembra meno pesante, se ci siete voi come mio punto fisso. Un grazie inoltre a tutti i miei compagni di università, per aver reso questo percorso meno tortuoso, tra una risata ed uno spritz alle Panche, resterete sempre impressi nella mia mente come gli anni più belli, nonostante lo stress e le paure che li hanno accompagnati.

Infine, ma non per importanza, devo ringraziare colei con cui ho più condiviso negli ultimi ormai 5 anni, la mia amica, sorella, complice e confidente, Giada. Anche se all'apparenza possiamo sembrare poco simili: tu nord io sud, tu luna io sole, tu argento io oro, il legame che ci unisce va oltre ogni apparente differenza. Grazie per essere stata la mia principale figura di riferimento in questi anni. Senza te, sono sicura che questo percorso sarebbe stato decisamente diverso.

Ringraziare me stessa a questo punto sarebbe quindi vano, perché io senza tutti voi non sarei nulla.