|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

|  |  |
| --- | --- |
| N41 | 2020/10/21 |
| Source | MPAI Video Group |
| Editors | Ajay Luthra, Leonardo Chiariglione |
| Title | Draft MPAI-EVC Use Cases and Requirements |
| Target | MPAI Video |

**ABSTRACT**

This document describes various applications that use digital video coding technologies & standards, and their operating environments and attributes. It also describes the requirements that the next generation of video coding standard should meet to satisfy the needs of those applications.

# INTRODUCTION

Innovations in the areas of Artificial Intelligence (AI) and Machine Learning (ML) technologies and their implementations have been increasing exponentially during last few years. Their usage can now be found in wide range of applications and areas, including Image Processing, Image Recognition, eCommerce, Workplace Communication, Healthcare, Agriculture, Cybersecurity, Finance, Autonomous Vehicles, Supply Chain Management, Manufacturing etc.

MPAI AI-Enhanced Video Coding (MPAI-EVC) is focused on understanding how AI/ML can help in improving the performance of existing digital video coding technologies and standards. It is expected that if AI/ML based technologies are found to help in significantly improving the performance over state-of-the-art existing video codecs, an MPAI Video coding standard will be developed.

Section 2 describes various Use Cases that use digital video coding technologies & standards, and their operating environments and attributes. Section 3 describes the requirements that the next generation of video coding standard should meet to satisfy the needs of those Use Cases.

It is a work in progress. This topic was first discussed during the MPAI-EVC’s teleconference on Oct 13, 2020. It is expected to be further refined and developed during future meetings as well as via email discussions.

# Use Cases / Applications

## Entertainment TV Content Distribution

This application includes sending entertainment content to home. It is sometimes also referred as “Watch TV’ application. Content can be viewed on a large screen TV as well as small screens of mobile devices. Complexity of the encoders can be significantly higher than that of decoders. Content gets compressed in two modes:

* Real time encoding
* Off line encoding

Real time encoding is used for live contents like sports, news etc. Off-line encoding is used for stored content for application like On-Demand TV, Streaming Video etc. Compressed content can also be stored either in the network for nPVR (networked Personal Video Recoding) application or in-home DVR (Digital Video Recorder), including whole home DVR. In-home DVR can also distribute the stored compressed content to other devices in home, including other TVs and mobile devices.

Transcoding of the content may also be done either to change the bit rates or to use another encoding standard that can be decoded by a receiving device. Transcoding functionality can be implemented in IRDs (Integrated Receiver Decoder), commercial inserting devices or advanced nPVRs/DVRs.

Commercials may also be inserted in some content. Commercial inserters may also perform transcoding and/or transrating to match the coding standard and/or the bit rate to that of the main stream.

In summary, following are the attributes of the environment associated with this application:

* Network: Guaranteed QoS, as well as, Best Effort
* Protocols: MPEG-2 TS as well as IP based ABR
* Content Types: Natural video (Camera captured content), Computer generated graphics, Hybrid
* Content Attributes
	+ HD, 4k, 8k
	+ Aspect ratio: 16:9 is the most popular
	+ SDR and HDR
	+ Standard and Wide Color Gamut
	+ 4:2:0 Color format
	+ Frame Rates: up to 120 Hz
* Compression Mode: Real Time and Off-line
* Compression type: Lossy
* End-to-end delay: High
* Implementation architecture
	+ Decoders: Software as well as Hardware/ASIC based
	+ Encoders
		- Software as well as Hardware/ASIC based
		- Stand alone as well as Cloud based
		- Significantly more complex than decoders
	+ Transcoders
		- Decode and re-encode the content to convert the compression standard and/or bit rates
		- Software as well as Hardware/ASIC based
		- Stand alone as well as Cloud based
* Viewing Environment: 2D
* Accessibility: To be able to do random access with 1 to 6 sec delay. Picture level accessibility is not necessary and long Group of Pictures (GOP) structures can be used.

## Video Games, including eSports and Cloud Gaming

…[ Comment: Insert brief introduction]

Following are the attributes of the environment associated with this application:

* Network: Best Effort
* Protocols: IP based ABR
* Content Types: Computer generated graphics, Hybrid
* Content Attributes
	+ HD, 4k, 8k
	+ Aspect ratio: 16:9 is the most popular
	+ SDR and HDR
	+ Standard and Wide Color Gamut
	+ 4:2:0 Color format
	+ Frame Rates: up to 120 Hz
* Compression Mode: Real Time
* Compression type: Lossy
* End-to-end delay: Low
* Implementation architecture
	+ Decoders:
		- Software as well as Hardware/ASIC based
		- Stand alone as well as Cloud based
	+ Encoders
		- Software as well as Hardware/ASIC based
		- Stand alone as well as Cloud based
* Viewing Environment: 2D, including Head Mounted Devices (HMD)

## Videoconferencing

……[ Comment: Insert brief introduction]

Following are the attributes of the environment associated with this application:

* Network: Best Effort
* Protocols: IP based ABR
* Content Types: Natural video (Camera captured content), Computer generated graphics, Hybrid
* Content Attributes
	+ Mainly HD, sub-HD
	+ Aspect ratio: 16:9 is the most popular
	+ Mainly SDR
	+ Mainly Standard Color Gamut
	+ 4:2:0 Color format
	+ Frame Rates: up to 60 Hz
* Compression Mode: Real Time
* Compression type: Lossy
* End-to-end delay: Low
* Implementation architecture
	+ Decoders: Mainly software based
	+ Encoders:
		- Mainly software based
		- Stand alone
* Viewing Environment: 2D

## Social Media

## ……[ Comment: Insert brief introduction]

Following are the attributes of the environment associated with this application:

* Network: Best Effort
* Protocols: IP based ABR
* Content Types: Natural video (Camera captured content), Computer generated graphics, Hybrid
* Content Attributes
	+ Mainly sub-HD, HD
	+ Aspect ratio:
		- Main Video –Various
		- Video banners – 2:1, 6:1, 8:1,
		- Vertical video in social media - 4:5, 2:3, and 9:16
	+ Mainly SDR
	+ Mainly Standard Color Gamut
	+ 4:2:0 Color format
	+ Frame Rates: up to 60 Hz
* Compression Mode: Real Time and Off-line
* Compression type: Lossy
* End-to-end delay: High
* Implementation architecture
	+ Decoders: Mainly software based
	+ Encoders:
		- Mainly software based
		- Stand alone
* Viewing Environment: 2D

## Drones and Remote Sensing

…

## Medical video

...

## Telemedicine

…

## Security

…

## Digital Cinema

…

## Content Creation

…

# Requirements

## Coding Efficiency

* Must be able to provide around 25 to 50% bit rate reduction over existing state-of-the-art video coding standard(s) for similar visual quality

## Content Types

* Must be able to provide desired improvement in the coding efficiency for wide range of content types: Natural video (Camera captured content), Computer generated graphics, Hybrid, Video gaming content.

## Content Attributes

* Must be able to provide desired improvement in the coding efficiency for the content with wide range of attributes:
	+ Up to 8k resolution
	+ Rectangular video with wide range of aspect ratios, including video banners and vertical video
	+ SDR and HDR
	+ Standard and Wide Color Gamut
	+ 4:2:0, 4:2:2 and 4:4:4 Color formats (initial focus is on YUV based coding)
	+ Frame Rates: up to 120 Hz
	+ Bit depth: Initial focus is on 10 bits video

## Compression Types

* Initial focus is on Lossy compression and visually Lossless compression
* (Is mathematically Lossless mode required?)

## Compression Modes

Standard shall allow the encoders designed for operating in the following modes:

* Real Time
* Off-Line Coding

## Viewing Environment

* 2D, including HMD
* 3D (Stereoscopic) (Note: not under consideration in Phase 1)
* Immersive (Note: not under consideration in Phase 1)

## Distribution Networks

Digital video content gets distributed via multiple distribution networks, like Cable, Satellite, Telco, Cellular networks (e.g. 5G), Over-the-Air and storage media. From the QoS point of view, the distribution channels provided by these networks fall in two categories:

* Guaranteed channel bandwidth and capacity
	+ The available bandwidth in this case is generally known and fixed. It also has low error rate after FEC.
	+ Typically, MPEG-2 TS protocol is used.
	+ Due to FEC, the error rate at the video layer is small
* Best effort channel capacity
	+ Mode of operation here is to provide user the bandwidth that is available at the time of the distribution. The available bandwidth in this case is not known a-priori and is also time-varying.
	+ Typically, TCP/IP based Adaptive Bit Rate (ABR) streaming is used.
	+ Due to TCP/IP, the packet loss at the video layer is absent.

Standard must be able to support distribution of video over both the network-types above.

## Rate Control

## CBR, VBR, Capped VBR

## End-to-end delay

Standard shall be able to support various end-to-end delay configurations:

* High Delay (> 100 ms to off-line encoding)
* Low Delay (30 msec < Delay <100 ms)
* Very Low Delay (< 30 msec; less than one picture period)

## Accessibility

Standard shall allow entering/accessing into the compressed video bit stream with varying accuracy:

* Picture level/resolution (Editing, Splicing etc.)
* Greater than one picture lag/resolution (channel change, commercial insertion etc.)

## Implementation

Friendly to multiple implementation architectures:

* Hardware
* CPU
	+ Should be possible to design encoder with architecture that allows tradeoffs among coding speed vs coding efficiency vs number of available CPUs/Cores
* CPU+GPU/FPGA
	+ Should be possible to design encoder with architecture that allows tradeoffs among coding speed vs coding efficiency vs number of available CPUs/Cores and presence of GPU and/or FPGA
* CPU+GPU+NPU
	+ Should be possible to design encoder with architecture that allows tradeoffs among coding speed vs coding efficiency vs number of available CPUs/Cores and presence of GPU and/or FPGA and/or NPU
* ASICs
* System
	+ Stand alone
	+ Cloud based
		- Support Virtual Machine architectures
		- Should be possible to design encoder with architecture that allows tradeoffs among coding speed vs coding efficiency vs number of CPUs/Cores available
* Complexity
	+ In large number of applications, encoder complexity can be significantly higher than the decoder complexity.
		- Off-line encoding applications can tolerate very high asymmetry between encoder-decoder complexity. Those applications may also do multi-pass encoding.
		- Real time encoders have relatively less complexity than that of off-line encoders.
		- In some applications with two-way communication, especially those using mobile devices, it may be more desirable to have less asymmetry between encoder and decoder complexity.

Standard should provide the capability to efficiently trade-off encoder and decoder complexities to match the needs of various applications.

## Error resilience

## Not required

## Backward Compatibility/Scalability

* Not the focus/requirement of Phase 1

…

# References

[1] Place holder