|  |  |
| --- | --- |
| **ITU – Telecommunications Standardization Sector**  STUDY GROUP 21 Question C/16  **Video Coding Experts Group (VCEG)**  2-8 November 2024, Kemer, TR | Document VCEG-BW10-v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| Question: | C/16 SG21 (VCEG) | | |
| Source: | **Nokia** | Email: | {Justin.Ridge, Jill.Boyce, Miska.Hannuksela, Jani.Lainema, Ville-Veikko.Mattila, Sebastian.Schwarz} at nokia.com |
| Title: | **Requirements for future video coding** | | |
| Purpose: | **Proposal** | | |

# Abstract

This document proposes requirements for a video coding standard to eventually follow H.266/VVC. It includes a non-exhaustive list of motivational scenarios that may be developed into more complete use cases.

# Existing markets and applications

## Broadcast television

Applications such as terrestrial broadcast are spectrum-bound. Increased coding efficiency is key to the introduction of higher quality (resolution, frame rates) or the multiplexing of additional services. Additionally, there are often onerous compatibility requirements which make the use of layered coding schemes attractive. Experience with past video codecs has shown that the incorporation of scalability features early in the design process allows them to be included with minimal effort.

## Video over wireless networks

Many video use cases now rely on wireless networks. Examples include drones or other remotely controlled vehicles and surveillance systems. A significant amount of video streaming traffic is now carried via wireless networks. Although 6G will provide increased bandwidth and improved latency, the amount of video traffic still has the potential to stretch the capacity of a given cell. Furthermore, video traffic in such networks may vary and be subject to high burst rates. This leads to requirements regarding latency, robustness and the ability to switch or adjust coding parameters within the video bit stream.

## Gaming

Gaming requires low latency at high resolution, high frame rate content. Within this area, game streaming and cloud gaming are distinct and have different requirements, e.g. latency. The signal characteristics of gaming video are different from naturally occurring content, and a new video codec should provide improved performance for this type of content.

# Emerging markets and applications

## Video generated via artificial intelligence

By the time a standard is completed, it is likely that video content generated using artificial intelligence will have proliferated. While this video may be used in many applications (and thus be subject to the existing requirements for those applications), it may also have distinct signal characteristics. A new video codec must provide performance gains for AI-generated content as well as other types of content.

## Divergent processing capabilities

Recent video codecs can be implemented on mobile devices and in consumer electronics with relatively little distinction. In the future, there is the potential to make greater use of GPUs, NPUs etc. that may lead to complexity that varies greatly. A new video codec must offer performance improvements at all levels of complexity in both the encoder and decoder, i.e. it is not sufficient for there to only be a gain when the codec becomes tremendously more complex than previous video codecs.

## Trust and authenticity

The ability to generate synthetic video, perform complex analysis in real time, and alter live video signals will increase. In some scenarios it is important to understand the authenticity (‘provenance’) of video content.

## Machine-to-machine communication

The amount of video data sent between “non-human” entities is expected to increase drastically. In such scenarios specific tools may be required.

## Immersive content and communication

Immersive video has particular characteristics and requirements. Near-eye ‘immersive’ displays can involve high resolutions (at or above 4K per eye), and response time to head movement is critical thus bringing a latency requirement. Various projections may be used for 360˚ video, and to achieve feasible data rates viewport extraction is needed. Various rendering approaches may be used, including split rendering and remote rendering, as opposed to performing the entire rendering process near the display device.

# Requirements

## Compression performance

For many applications, compression efficiency is the most important property of a future video coding standard. A substantial improvement in compression efficiency compared to VVC is required for the target application(s); at no point of the entire bit rate range shall it be worse than existing standard(s). 30% bitrate reduction for the same perceptual quality is sufficient for some important use-cases and may justify a future video coding standard. Other use-cases may require higher bit-rate reductions such as 50%.

Subjective visually lossless compression shall be supported. Lossless compression should be supported.

## Complexity

The complexity shall allow for feasible implementation of encoding and decoding within the constraints of the available technology at the expected time of usage.

The design shall support a range of complexity operating points.

The design should support parallel processing implementations with low cost in terms of bitrate overhead. The design should support pipelining of the processing to different types of processing units.

Note 1: Complexity includes: Power consumption, computational power, memory bandwidth etc.

Note 2: Encoding complexity of approximately 10 times or more than that of VVC is acceptable for many applications.

## Complexity vs compression scalability

The codec shall be capable of operating in a range of encoder and decoder complexity configurations and shall provide a reasonable gain over existing standards at each point. That is, the codec shall offer a substantial gain when operating in a higher complexity configuration, and a meaningful gain when operating in a lower complexity configuration.

## Picture formats

The new standard shall support rectangular picture formats that will include all commonly used picture formats, ranging at least from VGA to 8Kx4K. Picture formats of arbitrary size shall also be supported, within limits specified by Levels.

## Colour Spaces and Colour Sampling

1. YCbCr colour spaces with 4:2:0 sampling, 10 bits per component shall be supported
2. YCbCr/RGB 4:4:4, YCbCr 4:2:2 and monochrome should be supported.
3. Bit depths up to 16 bits per component should be supported
4. High dynamic range shall be supported (peak brightness 1000, 4000, and 10000 nits dynamic range greater than 16 f-stops)
5. Wide gamut colour BT2100 shall be supported
6. Additional modalities (transparency, depth, etc.) should be supported

## Frame Rates

The new standard shall support fixed and variable rational frame rates starting from 0Hz. Maximum supported frame rate shall be at least 240Hz, with upper limits specified by levels.

## Source Video Content Characteristics

The new standard shall support the encoding of the full variety of characteristics of video content encountered in the envisioned applications (to the maximum extent feasible). This includes (electronic and film) camera-captured scenes (e.g., HDR/WCG, 3D), text and graphics mixed into a camera-captured video source, rendered animation content, rendered computer graphics, computer desktop or mobile device display content, scrolling text over camera-captured video, imagery from medical and other application-specific sensors, video game content, content generated via artificial intelligence, content not designed for direct human viewing (e.g. depth, alpha channels, image features, feature masks), content intended for machine processing or analysis, etc.

## Scanning Methods

Support for progressive scanning shall be required for all Profiles and Levels.

## Low Delay

Encode plus decode latency of less than one frame duration shall be supported.

Resolution switching shall be possible when operating in a low delay configuration.

## Random Access and “Trick Modes”

The standard shall support random access to certain positions in time of a stored video stream and allow fast channel switching in the case of multi-channel services.

Pause, fast forward, normal speed reverse, and fast reverse access to a stored video bitstream shall be supported.

Random access to positions in space shall be supported.

Intra-only coding shall be supported.

## Error Resilience

Video bitstream segmentation and packetization methods for the target networks shall be supported.

The video layer and its interfaces to the network layer should be designed in a way such that relevant error resilience measures can effectively and flexibly be applied for networks needing error recovery, e.g. networks subject to burst errors. Proper balance of increase in complexity, loss in coding efficiency and benefits achieved by the error resilience measures at the coding layer should be achieved.

Note: Frame-level error resilience should be considered in the interactive/low-delay networks (currently, in the mobile networks, packet losses of up to 20% is observed).

In addition, the error resilience should be considered with regard to video game applications.

## Buffer Models

Buffer models, including hypothetical reference decoders (HRDs), shall be specified for target applications.

## Interface to System Layers

The standard shall be designed to permit efficient adaptation and integration with the target system and delivery layers. The standard shall also be designed to facilitate support of video-related services for accessibility (e.g. closed captions, descriptive video services), multiple languages, emergency alerts, and alternate content (e.g. advertisements, blackouts). The standard shall facilitate trust and authentication of source content.

## Support for Adaptive Streaming

The standard shall support fast representation switching in the case of adaptive streaming services that offer multiple representations of the same content, each having different properties (e.g. spatial resolution or sample bit depth). The standard shall enable the use of efficient prediction structures (e.g. so-called open groups of pictures) without compromising from the fast and seamless representation switching capability between representations of different properties, such as different spatial resolutions.

## Scalability

Scalability modalities (such as temporal, spatial, and SNR scalability) shall be supported.

## Multiview video

The standard shall support the coding of stereo and multiview content with optional depth.

## Machine-to-machine video

The standard shall support the coding of content envisioned for machine consumption.

## Transcoding

Low-cost edge transcoding should be considered for carrying multiple video representations given that clear benefits can be demonstrated.

## Omnidirectional video and projections

The codec shall efficiently code monoscopic and stereoscopic content captured using one or more lenses and potentially represented by common omnidirectional projection formats.

The codec should enable efficient extraction of a desired viewport from the compressed bit stream.

The codec should enable inclusion of parameters in the bitstream related to any correction due to optics or pre-processing, which may be useful to the rendering device in order to accurately reproduce the scene.

## Volumetric video

The codec should efficiently code 2D representations of volumetric video.

## Trust and authentication

The video codec should enable authenticity and trust mechanisms.