|  |  |
| --- | --- |
| **ITU – Telecommunications Standardization Sector**  STUDY GROUP 21 Question 6  **Video Coding Experts Group (VCEG)**  76th Meeting: 27 March – 4 April 2025, by teleconference | Document VCEG-BX25-v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| Question: | 6/21 (VCEG) | | |
| Source: | **ITU-T Q6/21** | Email: | [gary.sullivan@dolby.com](mailto:gary.sullivan@dolby.com) |
| Title: | **Draft of use cases and requirements for potential next-generation video coding standard beyond VVC capability** | | |
| Purpose: | Approved output of Q6/21 | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Abstract**

This document contains a set of draft requirements for a next-generation video coding standard with capabilities beyond VVC (Rec. ITU-T H.266 | ISO/IEC 23090-3), for potential joint standardization by ITU-T SG21 and ISO/IEC JTC 1/SC 29. These draft requirements were agreed in a joint meeting session of Q6/21, ISO/IEC JTC 1/SC 29/WG 2 (MPEG Requirements), JVET, and ISO/IEC JTC 1/‌SC 29/AG 5 (MPEG Visual Quality Assessment) on Monday 31 March 2025.

**1 Use cases and applications**

*Use cases:*

* Anticipated use cases include:
  + Mobile communication and entertainment.
  + On-demand streaming, download-and-play, and storage-media-based applications.
  + Live streaming.
  + Mobile video content uploading for cloud storage and social networking services (SNS) applications.
  + Low-delay interactive communication, e.g. for remote collaboration.
  + Ultra-low delay communication and entertainment, e.g. cloud gaming or autonomous driving.
  + Broadcast television.
  + Digital cinema and large-screen digital imagery.
  + Immersive applications, such as virtual, augmented and mixed reality for communication, interaction and entertainment.
  + Stereoscopic immersive content for playback in large venues, domes or HMDs
  + Surveillance and smart homes.
  + Video editing.
  + Professional content production and contribution
  + Machine analysis of video content.
  + Internet of Things, industrial applications, automotive, robotics, and digital medicine.
  + Low-delay 3D representation for remote communication
  + Representation of content to allow the rendering of a viewpoint from a limited specified perspective
  + Single- and multiview representations for medical applications
  + Adaptation to device capabilities with regards to coding complexity
  + Video coding for machine-to-machine communication
  + Trust and authenticity of video content, including identification of AI generated and AI modified video content
  + Low-cost edge transcoding should be considered for carrying multiple video representations given that clear benefits can be demonstrated.
* Specific content types to be addressed
  + User-generated content.
  + High-dynamic range and extended colour gamut content.
  + High frame-rate and sports content.
  + Content with film grain and content with digital noise.
  + Professionally authored content.
  + Screen content.
  + Gaming content.
  + AI-generated content.
  + Stereoscopic and multi-view content.
  + Mixed-source content (e.g., mixtures of camera-captured and rendering-generated content).
  + Representations of assets created from multiple sensors with different characteristics
  + Content with associated sample-based maps (e.g. for depth or alpha).
  + Monochrome and non-visual content.
  + Coding of alpha and/or depth components together with tristimulus colour components.
  + Other video modalities such as video atlases consisting of multiple distinct image regions from the same or different sensors, occupancy maps, and object instance ID maps.

**2 Compression capability**

*Requirement***:**

* The standard shall be capable of providing a substantial bit rate reduction at similar subjective quality compared to the VVC Main 10 profile over a wide range of bit rates and a wide range of content characteristics, including challenging content. For the entire anticipated operation range, it shall provide some bit rate reduction relative to existing standard(s) for equivalent subjective quality.
* Visually lossless compression shall be supported. Mathematically lossless compression should be supported.

**3 Complexity**

Complexity refers to computational resource consumption (in terms of battery drainage, power consumption, computing cycles, memory capacity, memory bandwidth, etc.) based on typical computing architectures and parallelization mechanisms.

*Requirements:*

* The standard’s complexity shall allow for feasible implementation at the expected time of usage. Technologies such as neural networks would not be precluded from consideration as a core element of the design, provided the expressed requirements are satisfied.
* Real-time decoding shall be feasible at the expected time of usage.
* The standard shall enable encoders to trade off complexity and coding efficiency in an economically viable way. Real-time encoding on a wide range of devices, including mobile devices, with adequate coding efficiency advantage over existing standards shall be feasible at the expected time of usage. Non-real-time encoding with further improvement of coding efficiency should be feasible at the expected time of usage.
* The standard may provide multiple decoder complexity configurations and shall provide a reasonable gain over existing standards for each such configuration. That is, the standard shall offer a substantial gain when operating in a higher complexity configuration, and a meaningful gain when operating in a lower complexity configuration.

**4 Bit depth and colour sampling**

*Requirements:*

* Commonly applied functionalities and formats that can be readily supported without major architectural modifications shall be supported in the core coding design, including the following:
  + 4:2:0 colour sampling and monochrome
  + Bit depths of 8 to 10
  + Arbitrary picture size and frame rates (incl. variable frame rate)
  + Variable picture size within a coded video sequence
  + Omnidirectional video
  + Stereoscopic 3D, multi-view, depth, alpha, and auxiliary components
* The standard shall be capable of representing video signals with bit depths of 8 and 10, and should be capable of representing video signals with bit depths ranging from 8 to 16 and non-video signals with bit depths ranging from 1 to 16 bits.
* The standard shall be capable of representing video signals with 4:2:0 and monochrome colour sampling, and should be capable of representing video signals with colour samplings ranging up to 4:4:4.

**5 Resolutions, frame rates, etc.**

*Requirements:*

* The standard shall be capable of representing pictures and video signals with spatial resolution up to 8K×4K, and should be capable of representing pictures and video signals with spatial resolution larger than 8K×4K.
* The standard shall be capable of representing video signals with spatial resolutions with a square or portrait aspect ratio.
* The standard shall be capable of representing video signals with temporal resolutions ranging up to 240 fps, and should be capable of representing video signals with temporal resolutions higher than 240 fps.
* The standard shall be capable of representing still pictures.
* The standard shall be capable of representing pictures with wide colour gamut and high dynamic range (e.g., ITU-R BT.2100) as well as traditional dynamic range colour formats (e.g., ITU-R BT.709).
* The standard shall be capable of coding progressively scanned video signals, and may support the coding of other scanning formats.
* The standard shall support omnidirectional video and projection formats, including enabling extraction of a desired viewport from the compressed bitstream. The codec should support inclusion of parameters in the bitstream for correction of optics.

**6 End-to-end delay**

*Requirements:*

The standard shall enable low end-to-end delay operation, efficiently enabling interactive and conversational applications.

* The codec shall support applications requiring operation without picture reordering.
* The codec shall be capable of providing an effective tradeoff of delay and coding efficiency.
* The codec shall support encode-decode latency of less than one frame duration (a.k.a. ultra-low delay).

Spatial resolution switching shall be possible when operating in a low-delay configuration.

**7 Scalability functionalities**

*Requirements:*

* Scalability modalities (such as temporal, spatial, and picture-quality scalability) shall be efficiently supported.
* The standard shall support the efficient coding of stereoscopic and multiview content with optional depth in a view-scalable manner.

**8 Random access and "trick mode" support**

*Requirements***:**

* The standard shall have support for random access points in the video bitstream for functionality such as channel switching and program chapter access.
* The standard shall have support for pause, fast forward, normal speed reverse, and fast reverse access to a stored video bitstream.
* The standard shall have support for spatial random access in the video bitstream for functionality such as extraction, transmission, and decoding of only one or more rectangular regions of the pictures, for efficiently enabling applications such as virtual reality, 360o video streaming, and region-of-interest stream adaptation.

**9 Packet loss resilience**

* The video layer and its interfaces to the network layer shall be designed in a way such that relevant data loss resilience measures can effectively and flexibly be applied for networks needing loss recovery, e.g. networks subject to burst errors. Proper balance of increase in complexity, loss in coding efficiency and benefits achieved by loss resilience measures at the coding layer should be achieved.
* Frame-level loss resilience should be applicable for interactive/low-delay networks, including channels with up to 20% packet loss.
* The loss resilience should be effective in ultra-low delay applications.
* Error concealment techniques to handle loss of data corresponding to parts of a coded picture should be considered

**10 Buffer model**

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

**11 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 authentication of source content.

**12 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.
* The standard shall enable the use of efficient prediction structures (e.g. so-called open groups of pictures) such that fast and seamless representation switching is supported.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_