INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2017-2020

SG16-TD155-A3/PLEN

STUDY GROUP 16

Original: English

Question(s): 6/16 Macao, China, 16-27 October 2017

TD

Source: Rapporteur Q6/16

Title: Copy of "Requirements for a Future Video Coding Standard v5" (Doc. ISO/IEC JTC1/SC29/WG11 N17074)

Status: NOTED by the SG16 Plenary when approving the JVET ToR (SG16-TD155-A1/PLEN) and requirements ITU-T SG16 Question 6/16 requirements (SG16-TD155-A2/PLEN), on 2017-10-27.

Abstract: This document contains the "Requirements for a Future Video Coding Standard v5" (ISO/IEC JTC1/SC29/WG11 N17074)

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 N17074
July 2017, Torino, IT

Source	**Requirements**
Status	**Approved**
Title	**Requirements for a Future Video Coding Standard v5**

Requirements for a Future Video Coding Standard v5

1 Introduction

The expanding use of more information rich digital video in diverse and evolving context and the still limited transmission and storage capabilities demand `more powerful compression schemes.

2 Estimated industry needs

The following four classes of potential users of a future video coding standard have been identified:

1. The class of application that is likely to accept the most frequent change of algorithm is one where decoding is purely software-based and there is no need for storage of the encoded bitstream (e.g. some videoconferencing applications)
2. The second class is likely to consist of applications with software decoding, but with the need to maintain server farms to deliver pre-encoded content to the end customer (e.g. OTT video streaming)
3. The third class is likely to consist of applications with hardware decoding, but with a consumer expectation of fairly rapid equipment swap-out (e.g. mobile telephony)
4. The class with the greatest barrier to accepting a frequent change of algorithm is likely to consist of those applications that are based on hardware decoders, where the consumer has an expectation of relatively infrequent swap-out (e.g. traditional terrestrial / satellite)

Within each of these four classes, an additional important factor was identified to be the "consuming device ownership".

In the example of broadcasting, there appears to be a lower barrier to change of algorithm in a vertically controlled pay TV market. In this case, the operator is able to balance the long-term commercial benefit of moving to a more efficient compression standard against the cost of accelerating the swap-out of legacy set-top boxes. In a horizontal free-to-air market, where the consuming device is owned by the end customer, there tends to be a political requirement to continue to provide service to residual legacy devices until they represent only a tiny percentage of the population.

In other application areas the opposite market dynamic may apply; if the user owns the consuming device then they may feel that owning the latest device confers enhanced status, thus providing an incentive for manufacturers to provide devices with new features, such as the latest decoder, which in turn provides an incentive to service provider to provide services in the new format.

3 New use cases for existing and emerging markets

3.1 Distinction between existing and emerging markets

Existing markets could be characterized as those where extrapolating from the past can be reasonably expected to provide some useful guidance to predicting the future. Emerging markets could be characterized as those which are so radically different from what has happened before that the past provides no useful guidance to the future.

3.2 Examples of existing markets for video coding

3.2.1 Terrestrial, and Satellite broadcasting

Broadcasting uses two basic business models: free-to-air (either funded by government or by advertising revenue) and pay TV (usually subscription-based, sometimes with additional individual pay-to-view events). Secure encryption using conditional access, to avoid content piracy, is a key technology for pay TV services. Terrestrial broadcasting is generally free-to-air, cable broadcasting is generally pay TV and both business models are used for satellite broadcasting.

For pay TV services including some cable broadcasting, the broadcaster typically provides a “set-top box” for reception and decoding as part of the subscription package, whilst the consumer owns the display. With free-to-air services, the consumer owns the receiving and display devices, either a fully integrated TV or an integrated receiver/decoder plus a separate display. Historically, there is a consumer expectation of a long lifetime of such devices, although the expected lifetimes may be reducing. For example, a report for the BBC Trust in 2009 indicated that the typical replacement cycles for primary digital receiving equipment in the UK was 7 to 8 years for integrated TVs, 5 to 6 years for integrated receiver/decoders. The "public service" nature of free-to-air broadcasting implies that there is a strong political pressure to avoid a situation where consumers get a blank screen with old devices that are only capable of decoding old formats.

At the same time, there is an expectation of ever-increasing video quality. Video resolution has increased from standard definition TV (SDTV) to high definition TV (HDTV) and now 4K ultra high definition TV (UHDTV), with 8K UHDTV broadcasting already launched in Japan. The traditional frame rates of 25 and 30 fps have increased to 50/60 fps in the first phase of UHDTV, with 100/120 fps expected in a second phase. The colour gamut and dynamic range have also been extended from BT.601 (SDTV) and BT.709 (HDTV) up to BT.2020 (UHDTV) and BT.2100 (HDR).

Conversely, there is constant pressure to reduce the spectrum available for broadcasting, particularly in the case of terrestrial broadcasting. Up until now, this has been partially offset by the introduction of more efficient channel coding and modulation, typically at the same time as the introduction of more efficient video coding. For example, the original terrestrial digital TV services launched in the UK in 1998 used MPEG-2 video coding with DVB-T channel coding and modulation, which gave a capacity of about 27.1 Mbit/s in an 8MHz channel. The DVB-T2 multiplex launched in 2011 gave a capacity of about 40.2 Mbit/s in an 8MHz channel, for a similar level of robustness, and used AVC video coding. However, since the modulation performance is now approaching the Shannon limit, there is little potential for further improvement in this aspect of the system and hence greater reliance on more efficient video compression.

For further information on broadcasting applications, see MPEG input document M36097, "Analysis of an Existing Market for Video Coding: TV Broadcasting in the UK".

3.2.2 Cable Services

Cable services are provided over a broad set of platforms including QAM based, IP based, and wireless based networks. It is important to recognize that the receiver devices may be IP television sets, QAM and IP set top boxes, game consoles, and mobile devices, and that it is necessary for cable operators to deliver a consistent experience, and quality of service across these platforms. Note that the above devices are both customer-owned and service operator owned devices, and the frequency with which they are replaced varies.

3.2.3 Managed IPTV via fixed telecommunication services

Managed IPTV services are typically delivered to the home on the telecommunications network that was originally designed and installed to carry voice frequencies over distances of several km from the local telephone exchange to the home. The xDSL modem technologies use higher frequencies than voice services, so the signals attenuate more rapidly with distance from the exchange. Providing increased bit-rates to the consumer therefore requires extending the fibre network closer to the home to reduce the length of the twisted pair connection, typically deploying a Fibre to the Cabinet topology with VDSL from the street cabinet to the home, capable of offering speeds of up to about 40Mbit/s. Much higher bitrates can be achieved by replacing the external twisted pair network entirely by fibre: a topology known as Fibre to the Premises or Fibre to the Home.

Almost all IPTV services today use AVC coding with hardware-based decoding. The decoding is typically performed by dedicated set-top boxes (STBs) supplied by the service provider as part of a subscription package, but other decoding devices such as games consoles are also sometimes used. It is desirable to improve compression efficiency to reduce distribution costs by optimizing bandwidth, but a change of compression standard requires the replacement of STBs, for which a replacement cycle of 5 years or more is generally sought. It is also desirable to control operational costs by limiting the number of formats in use at the same time, thus maximizing service platform interoperability.

The historical pattern has been to introduce a new compression standard, together with a higher resolution video format, approximately once every 10 years. However, there appears to be a law of diminishing returns in further increases of video resolution. About 50% improvement in compression efficiency remains a desirable goal when introducing a new compression standard, although the barrier to change is likely to be lowered when software decoding implementations become practical.

3.2.4 Professional content production and primary distribution

Professional video content is typically captured at a higher bit depth, chrominance resolution and bit-rate than will be used for the final version of the content after post-production. This practice facilitates the use of special effects, such as chroma key compositing, as well as ensuring that there is sufficient information captured to enable general post-production enhancement, such as colour correction. Movie content is generally captured as 4:4:4, whereas broadcasting content has traditionally been captured using a 4:2:2 format, reflecting the traditional use of interlace in the final broadcast transmission.

It is expected that the practice of using higher bit depth, chrominance resolution and bit-rate for production and primary distribution will continue in the future. For example, it is expected that professional content intended to be broadcast as 10 bit 4:2:0 UHDTV will tend to be captured as 12 bit 4:2:2 or 4:4:4. Although the use of the 4:2:2 format is currently well-established in the broadcasting industry it is likely to decline in the long term, since in the absence of interlace it is probably preferable to balance the horizontal and the vertical chroma resolution.

3.2.5 Digital cinema

A key consideration in the digital cinema market is the accurate maintenance of the artistic intent, through visually lossless compression, careful control of the ambient light levels and accurate replication of the director’s chosen values of luminance and chromaticity. Stereoscopic 3D content is more important for digital cinema than for other market segments, since the market expectation is to pay higher ticket prices for such content.

Digital cinemas generally follow the specifications produced by the Digital Cinema Initiatives (DCI), a joint venture of major motion picture studios. The DCI specification uses JPEG 2000 intra-frame coding to achieve visually lossless compression at an average total bitrate of about 80 to 125Mbit/s, constrained to a peak of 250Mbit/s per eye. The video format is 12 bit 4:4:4 using P3 colour space, with a peak luminance level of at least 48cd/m^2. The video resolutions are based on either 2K (2048×1080 pixels) or 4K (4096×2160 pixels, which is different from the “4K” used in broadcasting). The vast majority of movies still use the traditional 24 fps frame rate, although a small number of movies have recently been shot at 48 fps and there is consideration of introducing higher frame rates in the future: 60, 72, 96, and 120 fps.

The replacement of analogue by digital cinema resulted in a substantial reduction in workflow costs for the studios, who therefore provided financial incentives to encourage the transition. However, there appears to be minimal benefit in introducing a more efficient coding standard in these cinemas in the future, unless it somehow resulted in improved revenue, whilst any transition process would entail tangible costs due to the need for parallel workflows. On the other hand, there may be greater potential for enhancing the “premium” forms of cinema, such as IMAX. There has been some criticism that the current digital version of IMAX is noticeably inferior to the original film version.

3.2.6 Home cinema and packaged media

Similarly to the previous section, home cinema systems aim to create a cinema-like experience in the home, with large display screens and surround sound audio, within the quality constraints imposed by consumer-priced equipment. The content is typically played from packaged media such as an optical disc (e.g. Blu-ray), although streamed or server-based playout may become more important in the future.

3.2.7 Surveillance

There are four basic trends for the video surveillance industry:

1. Shift from analogue to IP-based video surveillance
2. Upgrade from SD to HD resolution
3. Intelligent video surveillance
4. Move from wired to wireless connectivity

IP-based HD video surveillance combines all three and will also be important for cloud based video analysis for intelligent surveillance. In the longer term, UHD with 4K resolution may be a potential market.

A typical IP-based HD video surveillance system includes the following parts:

- Network cameras with video processing, encoding and IP transmission functions
- Network infrastructure. Private networks are usually built for enterprise or city security, while public internet is generally used for consumer security
- Storage cloud for uploaded video content, often saved for weeks or even months, depending on the application requirements
- Video analysis cloud, used when requested by the system manager or by pre-defined security patterns. Alternatively, such intelligent analysis could be implemented in the camera side, to analyse the uncompressed video directly

HEVC has been adopted already in this market, to improve the video quality and reduce bandwidth and storage costs. A future new codec with improved video compression performance and acceptable complexity could be adopted relatively quickly, since the industry chain for video surveillance is quite short.

3.3 Examples of Internet based markets for video coding

3.3.1 Introduction

Internet based delivery of content creates new opportunities for the delivery of content with personalized advertising or even the modification of content depending on consumer preferences.

3.3.2 Over-the-top (OTT) services: IPTV via unmanaged networks

Broadband IP connectivity to the home can be provided using a range of wired and wireless technologies. Wired networks currently provide the highest bit-rate and reliability, typically using either xDSL modems based on the twisted pair telephony network or else DOCSIS cable modems based on the hybrid fibre-coax cable TV network or fiber to the home. OTT services including video on demand services, are also provided directly by cable operators. The quality of service achievable with both of these approaches has improved significantly over time, to the extent that it has become practical for “over-the-top” (OTT) TV services to be offered by a different organisation from that managing the IP network.

OTT content may be provided to the user without additional charge (e.g. BBC iPlayer) or else as a pay TV service (e.g. Netflix). OTT technology has enabled new players to enter the pay TV market, some of whom adopt a more international business model than is typical for traditional pay TV providers. Video on demand (VOD) subscription services are well-suited to OTT delivery, since they can be provided by a scalable content caching architecture and are less susceptible to potential problems of peaking in network demand that could detract from the quality of a popular live event delivered over unmanaged IP networks.

The first widely available HEVC-encoded UHDTV services were offered by the new OTT service providers. These providers are expecting a similar package of future video quality improvements as other TV service providers: higher resolution, higher frame rate, wider colour gamut and higher dynamic range. A key factor for OTT service providers is the network caching cost, so it is helpful to limit the number of coding formats in use at the same time.

Multiple video representations are frequently seen in OTT services. Scalability may be beneficial, both to save storage capacity and to improve the user experience through better adaptive streaming algorithms. It is important that the cost of carrying multiple content representations is low, especially for the highest layer. Edge transcoding may be a beneficial alternative for supporting multiple representations and should be considered for a future video codec

3.3.3 User generated content (UGC)

User generated content (UGC) may originate from various sources, such as smartphones, action cameras, drone cameras, wearable devices, dashcam, digital SLR, captured screen content, etc.

Future user generated content could include live streaming of an event attended by thousands, via smartphones. Such content, captured by multiple cameras, could potentially be reconstructed in 3D.

3.3.4 Video on demand via mobile telecommunication networks

Video on demand (VOD) is already an established system for video delivery to the home through wired internet connections. During the past few years, mobile phones have significantly increased in capabilities in terms of processing power and display quality. At the same time, tablets and laptop computers have become increasingly mobile through connections to 3G and LTE networks.

Consumers who subscribe to VOD services at home expect to be able to use the same VOD services with mobile devices. It is likely that it will become increasingly important to provide reliable services regardless of what device and connection type is used for accessing the content. Providers of VOD services will have to face challenges not only in terms of infrastructure, caching and delivery but also in terms of content creation and ensuring a certain level of Quality of Experience. One factor to take into account is the wide variety of display sizes on which the services may be consumed.

3.3.5 Video conferencing

Key requirements for video conferencing include low delay, the ability to tolerate diversity of networks and devices, rapid adaptation to varying network environments, relatively low bit-rates with easy to encode content, etc. Both point to point and multipoint scenarios are important and the number of encoders and decoders is fairly symmetric. The expected video quality may vary from very low to very high. Scalability is likely to be a useful feature for many video conferencing applications. There is a trend towards software implementations where the typical complexity is an important factor (as well as worst-case complexity).

For mass market native deployments by mobile network operators, 3GPP and GSMA are defining specifications for how to set up point-to-point and multiparty video calls and the specifications have rapidly been adopting new video codecs, AVC and HEVC. The video resolution has been increasing together with increased capacity in the networks and improved capabilities in the user equipment (mobile phones) and is currently defined up to 720p.

Deployment and usage of video telephony and video conferencing in mobile networks is constantly increasing but is not yet frequently practiced by a broader audience. For example, in LTE networks a voice-only telephone call can seamlessly be extended to a video-call as soon as both endpoints support it. But limited bitrate and coverage remains major obstacles and there is currently work ongoing on improving video telephony Quality of Service through definitions of efficient video error recovery mechanisms in 3GPP.

There are also mobile applications for professional Video Conferencing as a Service (VCaaS) products. For example, an insurance company could use VCaaS to enable remote customer service at the scene of an accident to inspect the damage. For this type of application, a high quality of service is typically required, e.g. 1080p video to allow the insurance company to use the video as evidence. The terminal device for this kind of application would typically be a smartphone or some kind of mobile IP camera with wireless uploading functions.

3.3.6 Screen content

Screen content applications require coding solutions that can efficiently compress computer-generated content with high visual quality. Examples would be wireless display, display wall, digital operating room (DiOR), transmission and storage of medical (2D+t) image data, virtual desktop infrastructure (VDI), screen/desktop sharing and collaboration, factory automation display, supervisory control and data acquisition (SCADA) display, automotive/navigation display and cloud gaming.

For further information on use cases for screen content applications, see document N14174, Requirements for future extensions of HEVC in coding screen content.

3.3.7 Gaming

A new category of application relates to 3D gaming, in which games are not necessarily played on local devices. Currently, games can be streamed from cloud to a heterogeneous set of devices, including notebooks, gaming tablets / portable devices, etc. In this category of applications, a server renders 3D games in a cloud server, and streams the game to any device with a wired or wireless broadband connection. This allows anyone to play full featured games from anywhere in the Internet. An additional category of applications is the broadcast of video games, which are played by people over the Internet, in real-time or for later viewing. The game streaming market is considered to be a sizeable and fast-growing market.

The game video content compression is usually considered to be much more challenging compared to the natural camera-captured video due to relatively sharp edges and larger motion.

3.3.8 Multispectral content

Multispectral content generally represents continuous tone representations of data that may be outside the range of the visible spectrum. Examples would include multispectral and hyperspectral imaging, remote sensing, radar, weather imaging, time of flight camera imaging and non-medical ultrasonic imaging. For some applications, having the encoder be less complex than the decoder may be considered, as the encoder may be remote and subject to power constraints. The component depths may be up to 16 bits and quality up to visually lossless and even mathematically lossless may be required.

4 Emerging Markets

4.1 Augmented and Virtual reality

Augmented and virtual reality platforms that include video coding are now emerging. The following virtual reality use cases have been identified.

4.1.1 Multicast streaming/broadcast

Virtual reality capture involves capturing the scene from many perspectives. Depending on the optics, this could involve a relatively small number of cameras, e.g. using fisheye lenses or similar, or it could involve a larger array of cameras covering a full 360 degrees, such as the eight lenses used in Nokia OZO.

The content from each camera could be "pre-stitched" so that any offset or overlap between cameras is taken into consideration and a single "scene" is composed prior to encoding. Alternatively, each "view" could be coded separately (or as a separate layer), potentially with some type of compensation between them.

Because there is no "feedback channel" in this use case, the capture device does not know which perspective the end user will take; thus capture must be complete (i.e. from every perspective). The quantity of video data being captured in this scenario means that some type of compression needs to be done near to the capture point, and that compression efficiency is extremely important.

From the client perspective, rendering the VR content may involve a spherical projection system, in which case the entire compressed bit stream is decoded and rendered. In this case, the decoder or renderer should take into account any warping or optical distortion from the capture process, so that the rendered video is displayed correctly.

Alternatively, rendering may involve displaying a subset of the captured video, e.g. using a VR headset[1]. In this case, rather than rendering all the captured video, a particular view is extracted according to sensory feedback from the viewing apparatus. Such feedback occurs continually, so that as the viewer moves his or her eyes or turns his or her head, a different view is projected. Because the capture and transmission is complete, delay is not a significant codec issue in this use case.

4.1.2 Unicast streaming

The perspective requested by the viewer(s) is identified dynamically according to sensory feedback, and requested from the capture end. Thus a subset of the captured video, corresponding to the view that is desired by the viewer(s), is streamed to the decoding device(s). However, there is a tight delay constraint –whether or not the video is "live", the rendered video must update promptly to avoid a "lag" between physical movement and the display.

4.1.3 Single-user point-to-point

Live single-user applications mean that the view may be "negotiated", i.e. the capture device potentially knows which view the user is interested in and can selectively request the far-end to capture and compress it. This use case is somewhat analogous to video conferencing due to the ability to negotiate, constraints on delay, etc. However, the system may still have to account for optical distortions and for the greater uncertainty due to the viewer's ability to arbitrarily change viewing perspective.

[1] M36488, "[FTV AHG] The head mounted display application for Free-viewpoint video service", G. Bang et al., 112th meeting (Warsaw, PL), June 2015.

4.2 Automotive applications

There is increasing use of video content, both internally for safety applications and externally for applications such as enhanced car navigation. Applications and scenarios related to fully Connected Cars will emerge more and more in the near, mid, and long-term future.

Currently, we already have vehicle-to-infrastructure (V2I) communication to achieve infotainment, on-line navigation, remote diagnostics, safety & security, communications. In the next generation, we will have vehicle-to-vehicle (V2V) and vehicle-to-infrastructure communications to warn drivers potential risks and improve traffic efficiency as basic sets of application. In the long term and as an enhanced set of application, we can have accident-free, driver-less vehicles and also smart vehicles achieving road/environment harmony.

Video will be a major element in a number of these scenarios. As an example, the driver-less car will require compression algorithms allowing computer vision based on compressed camera input. Key images features allowing vision recognition should be carefully preserved. In the use case of vehicle-to-vehicle video communication, the low delay will be also an important requirement.

4.3 Media-centric Internet of Things

Internet of Things (IoT) applications are expected to be deployed in greater numbers in the future. The use of video sensors over a range of resolutions will play an increasingly important role in these applications. Some IoT applications require low power consumption.

5 Requirements

5.1 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 HEVC Main Profile 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.

5.2 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.

5.3 Colour Spaces and Colour Sampling

a) YCbCr colour spaces with 4:2:0 sampling, 10 bits per component shall be supported

b) YCbCr/RGB 4:4:4 and YCbCr 4:2:2 should be supported.

c) Bit depths up to 16 bits per component should be supported

d) High dynamic range shall be supported (peak brightness 1000, 4000, and 10000 nits dynamic range greater than 16 f-stops)

e) Wide gamut colour BT2100 shall be supported

f) Auxiliary channels (transparency, depth, etc.) should be supported

5.4 Frame Rates

The new standard shall support fixed and variable rational frame rates starting from 0Hz, with upper limits specified by levels.

5.5 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, etc.

5.6 Scanning Methods

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

5.7 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 should support parallel processing implementations with low cost in terms of bitrate overhead.

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

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

5.8 Low Delay

Encode plus decode latency as low as one frame duration shall be supported. For some applications, an even lower latency should be supported.

5.9 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.

5.10 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 the video game applications.

5.11 Buffer Models

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

5.12 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).

5.13 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.

5.14 Scalability

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

5.15 Multiview video

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

5.16 Transcoding

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

5.17 Omnidirectional [j1]Video[2] and projections

- The next-generation video codec shall efficiently code monoscopic and stereoscopic content captured using one or more lenses and the combined result mapped using non-uniform projections to one or two rectangles, respectively.
  Example: Good compression efficiency shall be achieved using wide field-of-view lenses, and equirectangular panoramas.,.
- The next-generation video codec should enable efficient extraction of a desired viewport[3] from the compressed bit stream.
  Example: An extracted viewport should have the substantially same coding efficiency as this perspective coded on its own.
- The next-generation video 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.

[2] Omnidirectional video: image or video that enable *rendering* according to the user's viewing orientation, if consumed with a head-mounted device, or according to user's desired viewport, otherwise, as if the user was in the spot where and when the video was captured

[3] Viewport: part of the spherical video that is currently displayed and viewed by the user

NOTE: The size of the viewport can be smaller than or equal to the field of view.

6 Profiles and Levels

The standard shall define profiles and levels targeted at different application scenarios that are of interest to industry.

7 Timeline

The tentative timeline targets FDIS of the first version of the future video coding specification by the end of 2020.

8 References

[1] ISO/IEC JTC 1/SC 29/WG 11 document N15279, “AHG on Future Video Coding Standardization Challenges” , February 2015, Geneva

[2] ISO/IEC JTC 1/SC 29/WG 11 document M35458, “AHG on industry needs for Future Video Coding”, February 2015, Geneva.

[3] ISO/IEC JTC 1/SC 29/WG 11 document N15050, “Presentations of the Brainstorming Session of the Future of Video Coding Standardization”, October 2014, Strasbourg

[4] ISO/IEC JTC 1/SC 29/WG 11 document N14174, “Requirements for an extension of HEVC for coding of screen content”, Jan. 2014, San Jose