**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N17700**

**Ljublana, SI – July 2018**

|  |  |
| --- | --- |
| **Source:** | **Leonardo Chiariglione**  |
| **Title:** | **The MPEG work plan** |

**The MPEG work plan**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Std** | **Part** | **E/A/C** |  | **Motivations** | **Objectives** |
| MPEG-2 | 1 - Systems | AMD | 1 | Carriage of associated CMAF boxes for audio-visual elementary streams in MPEG-2 TS\r\nStructures to carry ISO/IEC 23000-19 CMAF boxes (CMAF Fragments boxes and CMAF initialization header boxes, no mdat box) over MPEG-2 transport stream along with associated audio-visual elementary stream that is designed to be transformed easily to CMAF delivery format. \r\nCarriage of JPEG XS in MPEG-2 TS\r\nMPEG Transport Stream can already carry JPEG 2000 (ISO/IEC 15444-1) for use in broadcast applications. In the meantime, WG1 has specified a new coding scheme, known as JPEG XS, that is more lightweight in terms of complexity, and focused on low-latency applications compared to JPEG 2000. | Carriage of associated CMAF boxes for audio-visual elementary streams in MPEG-2 TS\r\nThe CMAF boxes are carried over MPEG-2 TS metadata stream and it carries only the CMAF metadata boxes and does not carry any audio-visual sample data (mdat box is not carried). The sample data for mdat box will be derived from the associated audio-visual elementary stream.\r\nCarriage of JPEG XS in MPEG-2 TS\r\nThis new standard, numbered ISO/IEC 21122-1, is also intended to be used in broadcast applications, mainly for video transport over IP. Consequently, this Amendment proposal aims at defining in ISO/IEC 13818-1 the necessary syntax to transport this newly specified WG1 standard ISO/IEC 21122-1. |
| MPEG-4 | 10 - Advanced Video Coding | AMD | 1 | AVC is missing several SEI messages that have been defined in HEVC or are in the process of being defined in HEVC, including high-dynamic range colour related information, omnidirectional video related information, and manifest and prefix messages. | Specification of certain additional SEI messages in the AVC context that have been defined in HEVC or are in the process of being defined in HEVC. |
| MPEG-4 | 12 - ISO Base Media File Format | AMD | 1 | Sample groups are increasingly heavily used and there is an opportunity to reduce the size of the mapping box and make files smaller. The file format is being used to carry more diverse kinds of media, and that media needs richer support. There are also various small improvements to be made in various parts of the specification. | To specify following items\r\n• Stereo pairs in video\r\n• Compact sample to group\r\n• Basic multiplexed metadata (but not the advanced stuff)\r\n• Compatible schemes\r\n• Parts of MIME signaling \r\n• Spatial track relationships\r\n• Track groups and IDs |
| MPEG-4 | 12 - ISO Base Media File Format | AMD | 2 |   |   |
| MPEG-4 | 12 - ISO Base Media File Format | COR | 1 |   |   |
| MPEG-4 | 15 - Carriage of NAL unit structured video in the ISO Base Media File Format | AMD | 2 |   | To define two brands, \'hvti\' and \'lhte\' |
| MPEG-4 | 15 - Carriage of NAL unit structured video in the ISO Base Media File Format | COR | 4 |   |   |
| MPEG-4 | 32 - Reference software and conformance for file formats | STD | 1 |   |   |
| MPEG-7 | 15 - Compact descriptors for video analysis | STD | 1 | Industry needs a video description standard to enable scalable instance search in applications such as media production, archiving and security, and other applications that need to match content across many video sequences. | CDVA exploits the temporal redundancy of video by extracting a single compact descriptor to represent a segment rather than individual frames. This enables more compact descriptions for efficient matching of large sets of video, which is robust against changes of view, imaging conditions and transformations (e.g., transcoding, overlays) of video sequences. |
| MPEG-7 | 16 - Conformance and Reference Software for Compact Descriptors for Video Analysis | STD | 1 |   |   |
| MPEG-21 | 22 - User Description | STD | 2 |   |   |
| MPEG-A | 19 - Common Media Application Format | STD | 2 |   |   |
| MPEG-A | 19 - Common Media Application Format | AMD | 2 |   | Support of xHE-AAC and other media profiles |
| MPEG-A | 21 - Visual Identity Management Application Format | STD | 1 | There are widespread concerns on user privacy arising from sharing pictures in social media services. People can be on a picture taken by someone else, either intentionally or by mistake, and such picture can be posted on a social media service without any permission of the person captured on the picture and possibly without the person even being aware to be on the picture. Social media service operators try to provide some ways to manage such cases but it seems quite limited. Same happens for various video capturing devices such as CCTVs. | To specify a framework for managing privacy of users on the pictures or videos when pictures or videos are shared among users. |
| MPEG-A | 22 - Multi-Image Application Format | STD | 1 |   |   |
| MPEG-B | 7 - Common Encryption for ISO Base Media File Format Files | AMD | 1 | There are cases where it may be desirable or needed to have multiple keys, with their associated Ivs, for a single sample. For example, when a scalable or tiled media bitstream was represented by multiple tracks in a file, each of the tracks protected with its own keys, multiple keys per sample description is required to re-package the bitstream as a single track in the file. | To support multiple keys per sample using the following tools:\r\n1. extension of the seig sample group\r\n2. extension of the sample auxiliary info data for CENC |
| MPEG-B | 13 - Media orchestration | AMD | 1 |   |   |
| MPEG-B | 14 - Partial File Format | STD | 2 |   |   |
| MPEG-B | 15 - Carriage of Web Resources in ISOBMFF | STD | 1 |   |   |
| MPEG-D | 3 - Unified Speech and Audio Coding | STD | 2 |   |   |
| MPEG-D | 4 - Dynamic Range Control | AMD | 3 |   |   |
| MPEG-D | 4 - Dynamic Range Control | AMD | 4 |   |   |
| MPEG-D | 5 - Uncompressed Audio in MP4 FF | STD | 1 |   |   |
| MPEG-V | 7 - Reference Software and Conformance | STD | 4 |   |   |
| MPEG-H | 1 - MPEG Media Transport | STD | 3 |   |   |
| MPEG-H | 1 - MPEG Media Transport | AMD | 3 | Need to support Virtualized Network Function environment including virtualized MANE |   |
| MPEG-H | 1 - MPEG Media Transport | COR | 4 |   |   |
| MPEG-H | 2 - High Efficiency Video Coding | STD | 4 | This activity will create a new profile of HEVC that that will have an encoding of a single (i.e. monochrome) colour plane and will be restricted to a maximum of 10 bits per sample. This profile will complement other 10 bit profiles, such as the Main 10, Main 10 Still Picture, and Main 4:4:4 10 profiles. It is expected to be used in a variety of applications that may require signalling of 10 bit monochrome auxiliary information, such as depth information and alpha planes. The specification of additional supplemental enhancement information will also be included, including fisheye video, SEI manifest, and SEI prefix messages. | Specification of a profile of HEVC that that will have an encoding of a single (i.e. monochrome) colour plane and will be restricted to a maximum of 10 bits per sample, in a manner otherwise consistent with the prior specified “range extensions” profiles of HEVC. The specification of additional supplemental enhancement information will also be included, including fisheye video, SEI manifest, and SEI prefix messages. |
| MPEG-H | 2 - High Efficiency Video Coding | AMD | 1 |   |   |
| MPEG-H | 3 - 3D Audio | AMD | 5 | Marketplace requires that MPEG-H 3D Audio support all metadata used by 3D audio production tools. | To specify technology to deliver and process metadata from 3D audio production tools. |
| MPEG-H | 4 - MMT Reference Software | AMD | 1 |   |   |
| MPEG-H | 6 - 3D Audio Reference Software | STD | 2 |   |   |
| MPEG-H | 7 - MMT Conformance testing | STD | 1 |   |   |
| MPEG-H | 8 - HEVC Conformance testing | AMD | 1 |   |   |
| MPEG-H | 10 - MPEG Media Transport FEC Codes | AMD | 1 |   | To describet the two-stage FEC scheme implementation as one stage FEC or two stage FEC by one entity, and may be cascaded and added by two or more (if more than two stages) entities on the delivery path if needed. Similarly, the layer aware FEC may be implemented as one layer FEC or two layer FEC by one entity, and may be cascaded and added by two or more (if more than two layers) entities on the delivery path. |
| MPEG-H | 13 - MMT Implementation Guidelines | STD | 4 |   |   |
| MPEG-H | 13 - MMT Implementation Guidelines | TR | 3 | The MMT Implementation Guidelines describe the usage of MMT for multipath delivery, layer aware FEC and so on. | To extend the MMT Implementation Guidelines because MMT has added more technologies |
| MPEG-DASH | 1 - Media presentation description and segment formats | AMD | 5 | There are a number of technologies in DASH that have been under consideration for some time and the DASH subgroup plans to publish a new amendment to advance some of these technologies. | Device information and other extensions are planned to be included in this amendment. |
| MPEG-DASH | 1 - Media presentation description and segment formats | COR | 3 |   |   |
| MPEG-DASH | 2 - Reference software and conformance | AMD | 1 |   |   |
| MPEG-DASH | 3 - Implementation guidelines | TR | 2 |   | Guidelines for design and deployment of streaming media delivery systems using ISO/IEC 23009 (MPEG-DASH) including content generation, client implementation, and examples of deployment scenarios. |
| MPEG-DASH | 7 - Delivery of CMAF content with DASH | TR | 1 |   |   |
| MPEG-I | 1 - Technical Report on Immersive Media | TR | 1 |   | - Define a body of terminology (vocabulary) to be used across the Project.\r\n- Define the qualitative elements of an immersive experience at production &amp; consumption.\r\n- Provide one or more integrated and architectural views on how these elements contribute to an overall immersive experience and how they are combined. \r\nâ�¢ Define an architectural view on the compression and coded representation of elements of immersive experiences as well as the coded representation and delivery of a full media experience, taking into account the individuality of the experience, while enabling scalable and efficient individual delivery as well as mass distribution. \r\n- Document standardization requirements to create interoperability in end-to-end systems. Such aspects are expected to include Audio, Video, Graphics and Systems with capture and rendering, as well as appropriate interfaces with sensors that record navigation in the immersive audiovisual space, as well as suitable formats for cost-conscious delivery to mass markets. |
| MPEG-I | 2 - Omnidirectional MediA Format | STD | 2 |   |   |
| MPEG-I | 3 - Versatile Video Coding | STD | 1 | Industry needs a standard providing more video compression and new features | 1. Develop 2D video coding technology which could improve the compression performance or give new functionality, as compared to HEVC including the development of test cases and evaluation methodologies for assessment of such benefits are investigated.\r\n2. Develop video compression that can be applied to 360ᵒ Video (3DoF) |
| MPEG-I | 4 - Immersive Audio | STD | 2 | MPEG-H 3D Audio may need to be extended to support initial forms of immersive experiences | To explore how MPEG Audio technology, including MPEG-H 3D Audio, can be extended to support AR and VR use cases. When concrete use cases and requirements are established, new work will begin. |
| MPEG-I | 5 - Point Cloud Compression | STD | 1 | Technologies allow the capure of 3D point clouds typically with multiple cameras and depth sensors in various setups producing thousands up to billions of points when realistically reconstructed scenes are represented. Point clouds can have attributes such as colors, material properties and/or other attributes and are useful for real-time communications, for GIS, CAD and cultural heritage applications. | To specify lossy compression of 3D point clouds employing efficient geometry and attributes compression, scalable/progressive coding, and coding of point clouds sequence captured over time with support of random access to subsets of the point cloud. |
| MPEG-I | 6 - Immersive Media Metrics | STD | 1 | A consistent method to capture, measure and analyse such impact is essential to quantify and assess the VR product and application performance and effectiveness, maximize feelings of presence and enjoyment, and further optimize the product and experience design. While it is challenging to quantify the super accurate immersive level or emotional impact from the aggregate data, it is critically important to identify the basic objective metrics needed for a quality VR experience for MPEG-I use cases. | To specify the metrics and measurement framework to enhance the immersive media quality and experiences. It also includes a client reference model with observation and measurement points to define the interfaces for the collection of the metrics. |
| MPEG-I | 7 - Immersive Media Metadata | STD | 1 | In MPEG-I several standards will require similar information such as description about a projection. So, instead of having duplicated information in many standards, this standard will provide a single consolidated reference of information | To define common immersive media metadata focusing on immersive video (including 360° video), images, audio, and timed text. This metadata can be referenced by various other standards. |
| MPEG-I | 8 - Network-based Media Processing | STD | 1 | Recent developments in multimedia have brought significant innovation and disruption to the way multimedia content is consumed. With the emergence of VR and AR/MR applications, users can interact and navigate the consumed content along multiple degrees of freedom. Advanced media processing technologies (e.g., network stitching for VR service, super resolution for enhanced visual quality, transcoding, viewport extraction for 360Â° video) require too much compute power to be executed on modern mobile devices. | Network-based Media Processing (NBMP) will be a framework that allows service providers and end users to describe media processing operations that are to be performed by the network. NBMP describes the composition of network-based media processing services out of a set of network-based media processing functions and makes these network-based media processing services accessible through Application Programming Interfaces (APIs). NBMP framework allows content and service providers to describe, deploy, and control media processing for their content in the network. The NBMP Framework will be interoperable with existing Cloud platforms and is designed to integrate with multiple network environments such as 5G. |
| MPEG-CICP | 3 - Audio | COR | 1 |   |   |
| MPEG-CICP | 4 - Usage of video signal type code points | TR | 1 | Industry should be helped to avoid common content processing mistakes due to lack of understanding of approporiate combinations of video properties used, such as colour indication code points. With the increased usage of high-dynamic range and the increased use of look-up tables in television systems, these content processing mistakes could increasingly become magnified. | To provide guidance on combinations of video properties that are widely used in industry production practices. It will document the usage of colour-related code points and description data for video content production. |
| MPEG-G | 1 - Transport and Storage of Genomic Information | STD | 1 | genomic information as everyday practice in several fields, but the growing volume of data generated becomes a serious obstacle for a wide diffusion. The lack of an appropriate representation and efficient compression of genomic data is widely recognized as a critical element limiting its application potential. ISO/TC 276 and MPEG have combined their respective expertise and missions to develop a compression standard capable of providing new effective solutions for genomic information processing applications. | Transport and storage of genomic sequencing data and associated metadata with the capability of accessing these data sets efficiently, e.g. selective fast browsing, searching and access capabilities directly in compressed form. |
| MPEG-G | 2 - Genomic Information Representation | STD | 1 | The development of Next Generation Sequencing (NGS) technologies enable the usage of genomic information as everyday practice in several fields, but the growing volume of data generated becomes a serious obstacle for a wide diffusion. The lack of an appropriate representation and efficient compression of genomic data is widely recognized as a critical element limiting its application potential. ISO/TC 276 and MPEG have combined their respective expertise and missions to develop a compression standard capable of providing new effective solutions for genomic information processing applications | The objectives of the standard are to provide:\r\n- A transport format specification that supports a file format for storage scenarios and a packet format for streaming scenarios that are mutually convertible.\r\n- A compressed representation for sequence reads, quality values and alignment information that enable efficient selective access to genomic regions, data classes and associated information.\r\n- Standard APIs for selective access to the compressed genomic information and the conversion to and from MPEG-G files of commonly used genomic data formats.\r\n- Reference SW for the normative decoding process and informative encoding, conformance methodology. |
| MPEG-G | 3 - API for Genomic Information Representation | STD | 1 | The development of Next Generation Sequencing (NGS) technologies enable the usage of genomic information as everyday practice in several fields, but the growing volume of data generated becomes a serious obstacle for a wide diffusion. The lack of an appropriate representation and efficient compression of genomic data is widely recognized as a critical element limiting its application potential. ISO/TC 276 and MPEG have combined their respective expertise and missions to develop a compression standard capable of providing new effective solutions for genomic information processing applications. | To specify the API to access genomic informatiom to \r\n1. Simplify the usage and manipulation of sequencing data sets for genomic analysis applications\r\n2. Ensure interoperability of transport and storage formats at all levels of the various processing pipelines. |
| MPEG-G | 4 - Reference Software | STD | 1 |   |   |
| MPEG-G | 5 - Conformance | STD | 1 |   |   |
| MPEG-IOMT | 1 - IoMT Architecture | STD | 1 | Industry considers the Internet of Things (IoT) and SDOs make plans for related standards. \r\nMPEG has defined a specific instance of Thing called Media Thing (MThing), defined as a Thing able to sense and/or act on physical or virtual objects \r\nMThings may be connected to form complex distributed systems ``called Internet of Media Things (IoMT)`` where MThings interact between them and humans. | To describe the architecture of systems for Internet of Media Things. Internet of Media Things (IoMT) is a particular case of IoT (that by definition has the communication capability and it may sense or act on a physical or virtual object), with the specificity that an IoMT has media related multi-sensorial capabilities such as audio, visual, haptics. |
| MPEG-IOMT | 2 - IoMT Discovery and Communication API | STD | 1 | Industry considers the Internet of Things (IoT) and SDOs make plans for related standards. \r\nMPEG has defined a specific instance of Thing called Media Thing (MThing), defined as a Thing able to sense and/or act on physical or virtual objects \r\nMThings may be connected to form complex distributed systems &amp;acirc;ï¿½ï¿½ called Internet of Media Things (IoMT) &amp;acirc;ï¿½ï¿½ where MThings interact between them and humans.\r\nThese APIs for the media things facilitate for discovering other media things in the network |   |
| MPEG-IOMT | 3 - IoMT Media Data Formats and API | STD | 1 | Industry considers the Internet of Things (IoT) and SDOs make plans for related standards. \r\nMPEG has defined a specific instance of Thing called Media Thing (MThing), defined as a Thing able to sense and/or act on physical or virtual objects \r\nMThings may be connected to form complex distributed systems â�� called Internet of Media Things (IoMT) â�� where MThings interact between them and humans.\r\nThese APIs for the media things facilitate connecting and exchanging data between media things. The APIs also provide means for supporting media tokens and its wallet addresses to access functionalities, resources, and data from media things.\r\nThe data for media things consist of user commands (e.g., setup information) from a system designer, (raw or processed) sensed data, actuation information, and information for characteristics and discovery. | To specify data formats of input and output for media sensors, media actuators, media storages, media analyzers, etc.\r\nSensed data or analysed data can be processed further by media analyzers to extract semantic information. The standard does not specify how the process is carried out but only the interfaces. |
| Explorations | 7 - Immersive Video | EXP | 1 | Technology is making available different ways of offering a user an\r\nimmersive experience surrounding him/her with a large field of view\r\nvideo (up to 360 degrees) through Virtual Reality goggles or large 3D\r\nvideo walls. | To study immersive video where different viewpoints are presented to\r\nthe user`s surroundings, corresponding to rotational head movements\r\nonly (so-called Three Degrees of Freedom, 3DoF), possibly augmented\r\nwith a virtual or physical translational body/head movement in a\r\nlimited volume around a central position (referred to as 3DoF+). |
| Explorations | 7 - Immersive Video | EXP | 2 | Computational imaging technology offers users ways for immersive experiences with six degrees-of-freedom in limited volume free navigation, providing more freedom of user movement than in 3DoF+. Eventually, full-6DoF will be achieved (any translation and rotation in space), synthesizing virtual viewpoints from multiple, fixed cameras set up in various arrangements (planar arrangement, cameras in an arc, 360 divergent, etc). | To provide normative improvements on compression of 6DoF content on top of the state of the art anchor. The improvements are evaluated simultaneously on decoded views and synthesized views. |
| Explorations | 7 - Immersive Video | EXP | 3 |   |   |
| Explorations | 11 - Future Video Coding | EXP | 1 |   |   |
| Explorations | 28 - In advance signalling of MPEG containers content | EXP | 1 | MPEG defines several container formats, in particular ISOBMFF and MPEG-2 TS. Files conformant to these formats may contain multiple media streams, each of which may conform to different media formats, with different profiles and levels. There are several file consumption scenarios under which the full content of the file is not available to a player but under which the player has nevertheless to take a decision to retrieve the file or not. These scenarios include progressive file download, adaptive streaming, etc. In such scenarios, the player needs to have sufficient information to determine if it has or not the capabilities of playing the entire content or only a part of the container content, and when multiple container files are provided, to enable a player to choose the most appropriate file(s) to process. The practice to send information about the container content, together with URL(s) to the content and prior to its retrieval, is called hereafter \"in-advance signaling\". |   |
| Explorations | 29 - Compressed representation of neural networks | EXP | 1 | Recently (deep) neural networks (NNs) have become a widely applied method in many application areas, including signal processing and multimedia. Classification methods, feature extractors and encoding methods based on NNs often outperform hand-crafted approaches. In many applications the trained NNs (which may contain large amounts of data) need to be transmitted to other systems or terminal devices (with possibly limited computing capabilities), where they are used for inference and/or are updated with local data. Thus efficient representations for exchanging NNs are required. | To study existing representations of NNs, the state of the art of NN compression methods, and the processing flows of training and deploying NNs to a range of (generic or dedicated) hardware platforms, to identify interfaces where a standard compressed NN representation is needed and the define the requirements for such a representation. |