**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N13364**

**Jan 2013, Geneva, Swiss**

|  |  |  |
| --- | --- | --- |
| **Source** | **Communication Group** |  |
| **Status** | **Approved** |  |
| **Title** | **White Paper on State of the Art in compression and transmission of 3D Video** |  |

# 3D Video Displays and Applications

The primary usage scenario for 3D video is to support 3D video applications, where 3D depth perception of a visual scene is provided by a 3D display system. There are many types of 3D display systems including classic stereo systems that require special-purpose glasses to more sophisticated multiview auto-stereoscopic displays that do not require glasses [8]. This section provides a summary of displaying technology, a more comprehensive review of 3D display technologies can be found in [13].

Stereoscopic displays are a most commonly used type of 3D video displays. Such systems require two views (stereo video), where a left-eye view is presented to the viewer's left eye, and a right eye view is presented to the viewer's right eye. The 3D display technology and glasses ensure that the appropriate signals are viewed by the correct eye. This is accomplished with either passive polarization or active shutter techniques.

Since depth perception is known to be dependent on such factors as display size and/or viewing distance, same stereoscopic content being viewed in different viewing environment may provide a different level of depth perception [15]. In some cases it can be desirable or even required to adjust depth perception in automatic mean or through an interaction with the end user. The multiview displays have much greater data throughput requirements relative to conventional stereo displays in order to support a given picture resolution, since 3D is achieved by essentially emitting multiple complete video sample arrays in order to form view-dependent pictures. Such displays can be implemented, for example, using conventional high-resolution displays and parallax barriers; other technologies include lenticular overlay sheets and holographic screens. Each view dependent video sample can be thought of as emitting a small number of light rays in a set of discrete viewing directions – typically between eight and a few dozen for an autostereoscopic display. Often these directions are distributed in a horizontal plane, such that parallax effects are limited to the horizontal motion of the observer.

# 3D Video Data Formats

This section describes the various representation formats for 3D video and discusses the merits and limitations of each in the context of stereo and multiview systems.

**Stereoscopic and Multi-View Video**

A 3D scene can be represented in a traditional video-only data format. In such format, video information from each view point is captured at the full spatial resolution with a temporally synchronized camera array. Resulting representation consists of a separate video sequence with full resolution for each viewing point and the required data rate for captured raw data is practically multiplied by the number of captured views. An example of such representation for 2‑view case (stereo) is shown in Fig. 1.



Fig. 1: An example of representation of 3D scene with stereo video at full spatial resolution [10].

**Frame Compatible 3D Video Formats**

Frame compatible formats refer to a class of stereo video formats in which the two stereo views are essentially multiplexed into a single coded frame or sequence of frames, i.e., the left and right views are packed together in the samples of a single video frame. In such a format, half of the coded samples represent the left view and the other half represent the right view. Thus, each coded view has half the resolution of the full coded frame.

There are a variety of options available for how the packing can be performed. For example, each view may have half horizontal resolution or half vertical resolution. The two such half-resolution views can be interleaved in alternating samples of each column or row, respectively, or can be placed next to each other in arrangements known as the *side-by-side* and *top-bottom* packings (see Fig. 2). The top-bottom packing is also sometimes referred to as *over-under* packing. Alternatively, a "checkerboard" (quincunx) sampling may be applied to each view, with the two views interleaved in alternating samples in both the horizontal and vertical dimensions (as also shown in Fig. 2).

Temporal multiplexing is also possible. In this approach, the left and right views would be interleaved as alternating frames or fields of a coded video sequence. These formats are referred to as *frame sequential* and *field sequential*. The frame rate of each view may be reduced so that the amount of data is equivalent to that of a single view.

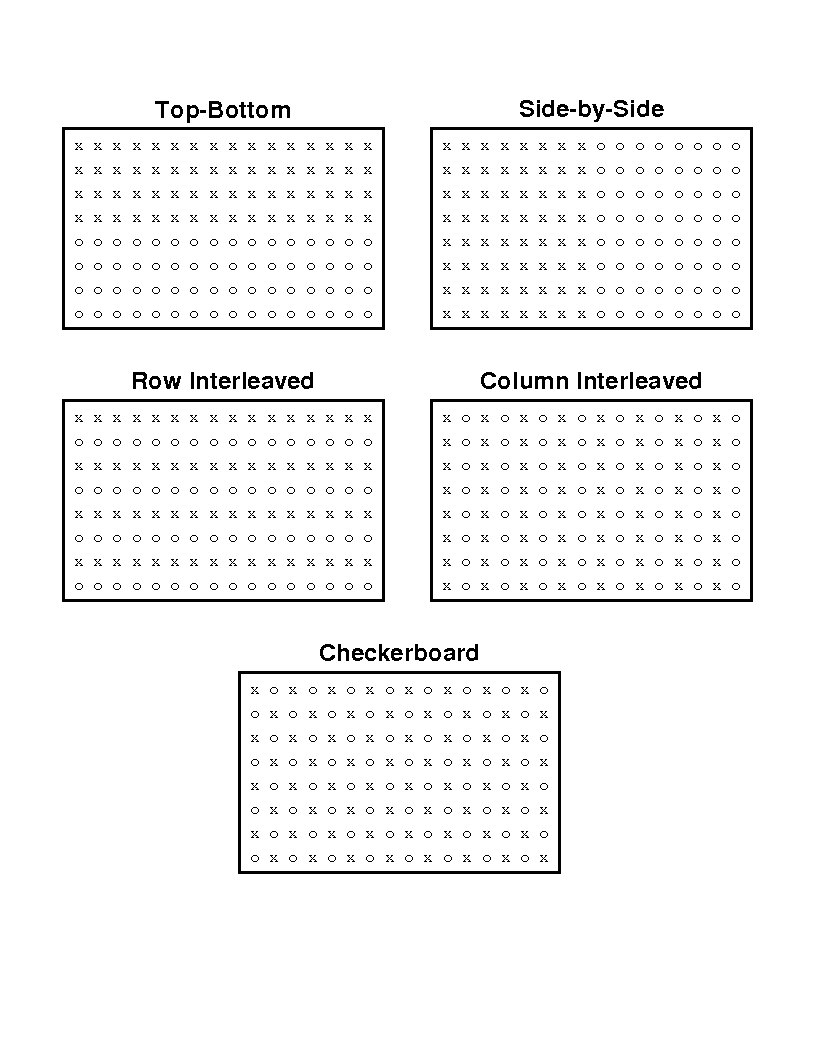


Fig. 2: Common frame-compatible formats where ‘x’ represents the samples from one view and ‘o’ represents the samples from the other view.

The primary benefit of frame-compatible formats is that they facilitate the introduction of stereoscopic services through existing infrastructure and equipment. Representing the stereo video in a way that is maximally compatible with existing encoding, decoding and delivery infrastructure is the major advantage of this format. The video can be compressed with existing encoders, transmitted through existing channels, and decoded by existing receivers [12]. However, legacy devices designed for monoscopic content may not recognize the format and may therefore display the frame-packed video (e.g. both views side by side).

**Service-compatible 3D Video Formats**

Service-compatible 3D video format refers to a class of stereoscopic video format in which one of the two stereoscopic views has full resolution which can be used for legacy 2D devices, whereas the other view may have the same resolution as the base view, or 3/4, 2/3, 1/2 resolution of base view (see Fig. 3). In this format, left and right views are transmitted as an independent video elementary stream. The benefit of the service-compatible 3D video formats is that it can be consumed simultaneously by both legacy 2D device and 3D device with high 3D video quality.

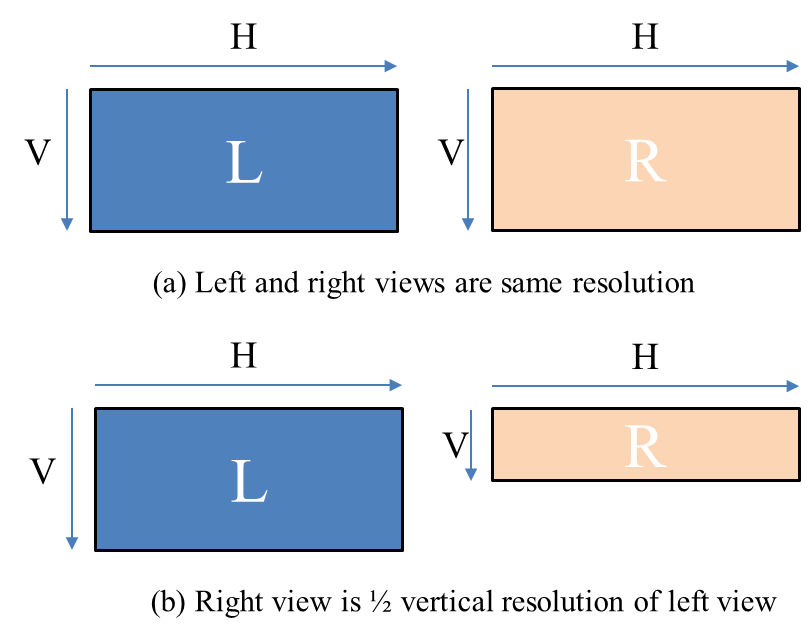


Fig. 3: Example of Service-compatible 3D video format

**Depth-Enhancement Video Formats**

3D scene video representation can be enhanced with supplementary, data e.g. depth information. Such information being available at display side can enable the generation of virtual views through depth-based image rendering techniques [14], and would facilitate a deployment of auto-stereoscopic multiview displays [9][10] and/or stereoscopic displays with adjustable depth perception.

The ISO/IEC 23002-3 (also referred to as MPEG-C Part 3) specifies the representation of auxiliary 2D video and supplemental information, e.g. in the form of a 2D plus depth format (see Fig. 3). In particular, it enables signaling for depth map streams to support 3D video applications.



Fig. 3: Visualization of 2D plus depth format concept.

However, the 2D plus depth format was found to only enable virtual view rendering within a limited viewing angle and not able to handle occlusions and holes resulting from rendering other views. In addition, stereo or multiview signals are not supported by this format, i.e., receivers would be required to generate the second view from the 2D video plus depth data for a stereo display, which is not the convention in existing displays.

To overcome the drawbacks of the 2D plus depth format, a multiview video plus depth format with a limited number of original input views and associated per pixel depth has been introduced, as shown in Fig. 4.



Fig. 4: Multiview video plus depth format for 2 views (MVD2).

Video and range data available from multiple viewing angles allows more sophisticated virtual view rendering algorithms and also provide more information for filling occlusion and/or holes when rendering novel views.

# State of the Art Coding of 3D Video

Several 3D video coding techniques have been designed in recent years, as summarized in this section.

**MPEG-2 Multiview Profile**

The first support for multiview video coding in an international standard was an amendment to the MPEG‑2 video coding standard [1] produced in 1996. This standard exploits inter-view redundancies present in stereo video. Thus, a significant coding gain is achieved in comparison to a simulcast coding approach, where each view is coded independently. MPEG-2 Multiview profile enables coding of two views (stereo) only. In this design, the left view is referred to as the "base view" and its encoding is compatible with that for ordinary single-view decoders. The right view is encoded as an “enhancement view” that uses the pictures of the left view as reference pictures for inter-view prediction.

This profile specifies that the base and enhancement view are coded with an identical set of coding tools defined in MPEG-2. However, motion compensation prediction in enhancement views was extended to exploit inter-view redundancies. For example, a reference picture of the enhanced view could either be a picture from within the enhancement view or a picture from the base view. An example of a prediction structure in the MPEG-2 multiview profile is shown in Fig. 5. The arrows in the figure indicate reference pictures for the predictive encoding of another picture.



Fig. 5: Illustration of inter-view prediction in MPEG-2.

**MPEG-2 Frame-Compatible Signaling**

Recently, there was a request to MPEG standardization to provide a MPEG-2 support for frame compatible 3D data format with Quincunx packing. Extending of existing standard for supporting of additional frame-packing arrangement is conducted within MPEG Video group and FDAM of this extension is expected in October 2012.

**AVC Frame-Compatible Signaling**

The signaling for a complete set of frame-compatible formats has been standardized within the MPEG-4 AVC standard as supplemental enhancement information (SEI) messages. A decoder that understands the SEI message can interpret the format of the decoded video and display the stereo content appropriately.

An earlier edition of the standard that was completed in 2004 specified a *stereo video information* (SVI) SEI message that could identify two types of frame-compatible encoding for left and right views: row based interleaving and temporal multiplexing of coded views. The SVI SEI message additionally introduced inter-view restriction flag. So called self-contained flag restricts interview prediction of enhanced view from pictures of base view, thus allowing independent decoding of enhanced layer.

Support for stereo video representation was significantly extended with a new SEI referred to as the frame packing arrangement (FPA). It was specified in an amendment of the MPEG-4 AVC standard [7] and was incorporated into the latest edition [6]. This new SEI message is the current way of signaling frame-compatible stereo video information for all frame packing arrangements, shown in Fig. 2. FPA SEI provides additional functionality, such as mirroring/flipping of the image in either of the side-by-side and top-bottom arrangements and signaling for application of quincunx (checkerboard) sampling to either of the coded views. Finally, the SEI message indicates whether the upper-left sample of a packed frame is for the left or right view and it also supports additional syntax to indicate the precise relative grid alignment positions of the samples of the left and right views, using a precision of one sixteenth of the sample grid spacing between the rows and columns of the decoded video array.

**MVC-Multiview Extensions of AVC**

MVC was developed as a multi-view coding extension for the monoscopic AVC coding standard. MVC provides a compact representation for multiple views of a video scene, providing higher resolution and quality relative to frame-compatible formats. Stereo-paired video for 3D viewing is an important special case of MVC. For higher compression efficiency, the standard enables inter-view prediction in addition to temporal and spatial prediction. The basic concept of inter-view prediction, which was also employed in the MPEG-2 design for multiview video coding, is to exploit both spatial and temporal redundancy for compression. Since the cameras of a multiview scenario typically capture the same scene from nearby viewpoints, substantial inter-view redundancy is present. A sample prediction structure is shown in Fig. 6. Pictures are not only predicted from temporal references, but also from inter-view references. The prediction is selective, such that the best predictor among temporal and inter-view references is automatically chosen in terms of rate-distortion cost on a block basis.

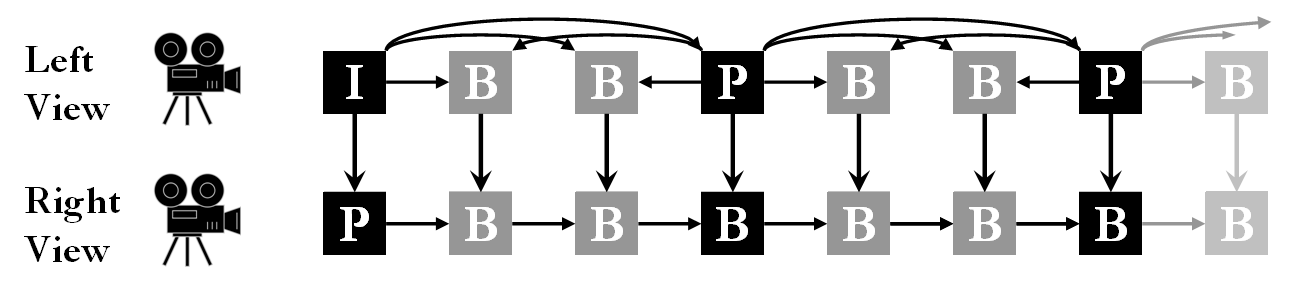


Fig. 6: Illustration of inter-view prediction in MVC.

Another key aspect of the MVC design is that it provides backward compatibility with existing legacy systems such that an MVC bitstream includes a compatible base view. In other words, it is mandatory for the compressed multiview stream to include a base view bitstream, which is coded independently from all other views in a manner compatible with decoders for single-view profile of the standard, such as the High profile. This requirement enables a variety of uses cases that need a 2D version of the content to be easily extracted and decoded. For instance, in television broadcast, the base view could be extracted and decoded by legacy receivers, while newer 3D receivers could decode the complete 3D bitstream including non-base views. MVC makes use of the NAL unit type structure to provide backward compatibility for multiview video. Further details of this design could be found in [8].

As with prior video coding standards, profiles determine the subset of coding tools that must be supported by conforming decoders. There are two profiles currently defined by MVC with support for more than one view: the Multiview High profile and the Stereo High profile. Both are based on the High profile of MPEG-4 AVC, with a few differences.

* The Multiview High profile supports multiple views and does not support interlace coding tools.
* The Stereo High profile is limited to two views, but does support interlace coding tools.

For either of these profiles, the base view can be encoded using either the High profile of MPEG-4 AVC, or a more constrained profile known as the Constrained Baseline profile which was added to the standard more recently.

**Hybrid MPEG-2 Video/AVC solutions**

For the terrestrial service-compatible 3DTV broadcasting service based on the ATSC standard, the hybrid MPEG-2 video/AVC solution is used to service of 3DTV broadcasting programs using independent MPEG-2 and AVC codecs. To provide a backward compatibility with legacy DTV receivers, base view is coded using MPEG-2 ‘Main Profile level @ Main level or High level’ and existing MPEG-2 video stream\_type (value of 0x02). The second view is coded using AVC ‘Main Profile level 4.0 or High Profile level 4.0’ and newly defined AVC video stream\_type (value of 0x23) for the service-compatible 3DTV service.

This solution has been adopted as Korean domestic standard in 2011, and is in the ATSC standardization process.

**Scalable stereo / Multiview**

For monoscopic video coding, plans exist to develop a suite of tools for scalable coding for HEVC (which will be executed by JCT-VC, the joint team that is currently defining the base spec for HEVC). Here, the application of particularly view scalability and spatial scalability are considered highly beneficial in the evolution of 3D services, as they allow for backward-compatible extensions for more views, and/or enhancing the resolution of views in a way that decoding by legacy devices is possible.

Therefore, it is likely that tools which JCT-VC will develop for scalable coding of monoscopic video are similarly applicable to frame-compatible stereo and MV-HEVC. If specific additional definitions are necessary for this purpose, this will be conducted by JCT-3V in close coordination with JCT-VC. A timeline for this activity would need to follow the scalable HEVC extension, such that finalization could not be expected earlier than end of 2014 or beginning of 2015.

For AVC, the possibility for a spatial scalable extension of frame-compatible formats (“MFC”) is currently investigated as well. A Call for Proposals on such technology has been launched by the MPEG Requirements group, with responses due by October 2012.

# MPEG-2 systems & file format for carriage and signaling of stereoscopic 3D video

MPEG-2 Systems provides mechanisms to transport stereoscopic 3D video, add time line and other parameters for synchronization of video with other components such as audio and subtitles, buffer management based on the equivalent functions in video compression standards and signaling of various types of stereoscopic 3D compression schemes. MPEG-2 Systems includes program streams and transport streams. Program streams have been used in applications such as VCD and DVD while transport streams have been used in broadcast and distribution applications. The carriage and signaling mechanisms for both program and transport streams are common with some additional functions provided by the transport streams.

MPEG-2 transport is made of constant size packets of 188-bytes each. Each packet is made of a 4-byte header that includes a unique ID for the component that is part of the payload (called Program Identifier PID), an optional ‘adaptation header’ which provides time base data such as ‘Program Clock Reference (PCR)’ and a ‘Program Elementary Stream (PES)’ header which provides timing data for buffer management and synchronization called ‘Decoding Time Stamp (DTS)’ and ‘Presentation Time Stamp (PTS)’. Each video component (called video sequence) is packetized into a succession of 188-byte packets and multiplexed with other appropriate components based on the application.

MPEG-2 Systems also provides signaling mechanism for packets with a unique PID value called ‘Program Association Table (PAT) and ‘Program Map Table (PMT)’. These signaling mechanisms are used by broadcast applications that include multiple programs (also known as channels), each program typically containing video, several audio components in different languages as well as subtitles or closed captioning. The signaling is present in the PMT and it includes the type of video component or components (if the stereoscopic 3D video is coded as two separate video streams), type of audio component as well as others. PMT also conveys other information about the video sequence that may or may not be present in the video stream itself through a mechanism called the ‘Descriptor’. Examples of information conveyed in a descriptor include audio streams language, presence of still pictures in video, use of ‘frame-packing’ arrangement in the video sequence and others. The PAT and PMT packets are part of the transport stream multiplex and these are present a frequent intervals based on the application requirements. Typically broadcast applications require these at frequent interval such as 1 second to assist in quick channel or program acquisition.

The fourth edition of MPEG-2 systems (ISO/IEC 13818-1: 2012) contains the transport and signaling mechanisms for all of the 3D video coding specifications developed by MPEG starting from 1995 to 2012. This section provides some information about the signaling and transport of each of the 3D video standards. 3D video technologies developed by MPEG include the following (in calendar order):

* MPEG-2 ‘multi-view profile’ in ISO/IEC 13818-2.
* MPEG-2 and AVC video with ‘frame packing arrangement’ where each 2D frame carries arrangement of 2 views. These are part of ISO/IEC 13818-2 and ISO/IEC 14496-10.
* 2D video plus a depth map in ISO/IEC 23002-3 also known as MPEG-C.
* Extension to AVC video called Multi-view Coding (MVC) where 2 separate views are coded using view interdependencies.

**MPEG-2 multi-view profile**

The very first 3D video standard was MPEG-2 “Multi-view profile” and this profile covered all video formats through the level mechanism (low, main and high 1440). This profile added the ‘camera parameters’ extension to video to signal additional data. As MPEG-2 systems already included a descriptor called ‘video stream descriptor’ which signaled the MPEG-2 video streams profile, level, frame rate and other parameters, this descriptor was also used to signal 3D video which used multi-view profile when this profile was added to MPEG-2 video. MPEG-2 video receivers that were deployed used this descriptor to determine whether the video was 2D or 3D and chose support based on their capability. No new stream\_type was needed to signal 3D MPEG-2 video component using ‘Multi-view profile’ and this content was also signaled using existing MPEG-2 video stream\_type (value of 0x02).

**Frame packing arrangement**

A novel method of packing the two views of a stereoscopic 3D video (at a lower resolution for each view) into a 2D frame was developed by MPEG so that existing video compression technologies and the tools can be used with no changes for 3D extension. The only additional video signaling included information about use of ‘frame packing’ and the type of packing. This extension is added to both AVC video (using the SEI message for signaling) and MPEG-2 video (using the user\_data extension).

MPEG-2 systems did not allocate a new stream\_type value to signal the use of ‘frame packing’ in the underlying video as existing 2D-only receivers were able to decode these streams without any extra decoding resources. Existing stream\_type values for MPEG-2 video (value of 0x02) and AVC video (value of 0x1B) signaled streams with frame packing arrangement. As the existing MPEG-2 video descriptor did not have any hooks to add extensions, a new descriptor called ‘MPEG-2 stereoscopic video format descriptor’ was added to the MPEG-2 video component to signal use of frame packing in the underlying video so that 3D capable receiving systems could use this information to render the decoded frame onto a 3DTV display. For AVC video with frame packing arrangement, the existing AVC video descriptor was extended (using 1-bit value) to signal use of frame packing arrangement in the underlying video so that 3D capable receiving systems could use this information to render the decoded frame onto a 3DTV display. This 3D video coding scheme is considered as part of ‘frame-compatible’ technologies as 2D-only capable receiving systems may not be able to render meaningful 2D images from a framed packed data. There are, however, private implementations where one of the views from the frame packed image is decoded and upsampled by 2D-only receivers to create meaningful 2D images.

**2D video plus a depth map in ISO/IEC 23002-3 (also known as MPEG-C)**

ISO/IEC 23002-3 used auxiliary information to convey ‘depth map’ and ‘parallax map’ which is used to generate 3D capability using 2D view that was coded using existing MPEG video compression technologies. The auxiliary information can also be compressed using existing MPEG video technologies.

MPEG systems added additional signaling for ISO/IEC 23002-3 auxiliary data component by allocating new stream\_type (0x1E) for use in the PMT. A new descriptor called ‘auxiliary video stream descriptor’ was also added (value of 47) to signal the compression technology used by the auxiliary information (such as MPEG-2 video, ISO/IEC 14496-2 video or AVC video). The PMT included 2D-video component and the auxiliary data component that were signaled in separate PID values as well as the ‘auxiliary video stream descriptor’ which conveyed additional information about the auxiliary information such as compression type. Auxiliary information component is also carried in PES like the 2D-video component with time stamp values (PTS/DTS) that are used to synchronize the 2D and auxiliary stream access units.

This 3D video scheme is considered as part of ‘service-compatible’ technologies as 2D-only capable decoders can decode and render the known 2D-video component in the program and ignore the auxiliary information component while decoders that support this technology use both the 2D-video component and auxiliary data component to render 3D video.

**Stereoscopic Video Application Format in ISO/IEC 23000-11**

In today’s technological arena, three-dimensional content services are considered as one of the most promising applications in the market. The market of applying stereoscopic video contents on digital devices is getting expanded and matured in movie, broadcasting and communications sectors. There are already various types of digital devices such as laptops, mobile phones and digital SMART 3DTVs available for capturing and displaying stereoscopic video contents in the market. However, these stereoscopic contents have difficulties in storage, interchange, management, editing, and presentation due to the lack of a common file format, which is considered as the hurdle for immersive 3D market. Thus, MPEG has completed the development of a new application standard format called as “Stereoscopic Video AF”, which provides the file format for an interoperable storage format of stereoscopic video and associated audio, images, and metadata in mobile and fixed high quality 3DTV environment.

ISO/IEC 23000-11 is used as a common file format for playback and storage of stereoscopic contents on various 3D devices. This standard provides various features such as i) Playback and storage of stereoscopic contents including various stereoscopic composition types; ii) Playback and storage of stereo-monoscopic mixed contents; iii) User interaction by scene representation for stereoscopic contents; iv) Supporting the compatibility with legacy file format such as ISO base media file format; v) Supporting the visual safety information for stereoscopic contents.

**Multi-view Coding (MVC) over MPEG-2 systems**

This extension was added to AVC video standard to support 3D stereoscopic video as well as multiple-view video systems. The existing 2D-video component is called ‘base view’ and is fully compatible with AVC video technology, while the additional views are compressed using inter-view dependencies for better compression.

MPEG-2 systems added several changes for carriage and signaling of MVC video to support different application use cases. The base view is signaled using the existing AVC video stream type value of 0x1B while the MVC coded additional views are signaled using a new stream type value of 0x20. Systems usage allows applications to concatenate views (in increasing order) into a single video stream component (called MVC video sub-bitstream) as well as use of many of these video sub-bitstreams in the same program. MVC video sub-bitstream is also carried in PES with time stamps (PTS and DTS) that are used to synchronize these views with the base view video component. A new descriptor called ‘MVC extension descriptor’ was added for use by the sub-bitstream component and this signals information such as view-values in these components as well as structure of the access units to assist MVC decoder systems. MPEG systems extended the buffer management model (called system target decoder STD) that was needed to aid in the re-assembly of base video and video sub-bitstreams to assist the MVC video decoder system and render 3D or multi-view video.

For stereoscopic 3D video applications MVC video includes the base view and one additional view (in a video sub-bitstream), the PMT signals two video components one with stream type value of 0x1B for base view and the additional view using stream type value of 0x20. The MVC extension descriptor is also associated with the additional view to convey more information.

MVC coding scheme is also part of the ‘service-compatible’ technologies as the base view is always decodable by an existing AVC 2d-only capable receiver using the signaling available in the PMT. MPEG systems recently added the capability to signal association between the 2 views and left or right eye to assist in 3D display rendering. This signaling was done through an extension to the MVC extension descriptor.

**Stereoscopic 3D video using simulcast of independently compressed views**

This application use case is only supported through the MPEG systems specification as the independent view compression uses existing MPEG video compression technologies such as MPEG-2 video or AVC video (and not MVC video or ISO/IEC 23002-3 technology). This use case was needed by applications where there was a limited network data bandwidth (such as terrestrial broadcast) and there were regulations that mandated MPEG-2 video coding for the base view.

This MPEG-2 systems signaling scheme supports delivery of stereoscopic 3D video content where the base view video stream (which is usually 2D compatible) and the additional view video stream are coded independently using either MPEG-2 or AVC video or any combination thereof. Additional signaling is provided through descriptors to enable 3D capable receivers combine these independently compressed views to present decoded data on a 3DTV display system.

Stream\_type values already exist to signal MPEG-2 video (0x02) and AVC video (0x1B) and these are used to signal the base layer of service compatible components (in the PMT) with no changes. The base layer is also specified as the full resolution 2D compatible layer. The compression scheme for the second view is signaled using two new stream\_type values, one for second view compressed using MPEG-2 video (0x22) and one for second view compressed using AVC video (0x23).

In addition, two descriptors(‘stereoscopic program info descriptor’ and ‘stereoscopic video info descriptor’) are specified to signal additional information that assists in the identification of services at program level as well base and additional view components for service-compatible 3DTV services.

The ‘stereoscopic program info descriptor’ provides information at a program level regarding the identification of 2D-only (monoscopic), frame-compatible stereoscopic 3D as well as service-compatible stereoscopic 3D services.

The ‘stereoscopic video info descriptor’ provides information to 3D receivers such as whether this view is the base view, association of this view to left or right eye for display and any up sampling factors needed if the additional view is compressed at a lower resolution.

This signaling scheme is also considered part of “service compatible” technologies as the base view is decodable by a 2D-only capable receiver. The signaling scheme is explicitly defined for applications that use ‘simulcast’ of the two stereoscopic views and does not include scalable or temporal enhancements.

# References

1. ITU-T and ISO/IEC JTC 1, "Final Draft Amendment 3", Amendment 3 to ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N1366, Sept. 1996.
2. A. Puri, R. V. Kollarits, and B. G. Haskell. "Stereoscopic video compression using temporal scalability", *Proc. SPIE Conf. Visual Communications and Image Processing*, vol. 2501, pp. 745–756, 1995.
3. X. Chen and A. Luthra, "MPEG-2 multi-view profile and its application in 3DTV", *Proc. SPIE IS&T Multimedia Hardware Architectures*, San Diego, USA, Vol. 3021, pp. 212-223, February 1997.
4. J.-R. Ohm, "Stereo/Multiview Video Encoding Using the MPEG Family of Standards", *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems VI*, San Jose, CA, Jan. 1999.
5. G. J. Sullivan, "Standards-based approaches to 3D and multiview video coding", *Proc. SPIE Conf. Applications of Digital Image Processing XXXII*, San Diego, CA, Aug. 2009.
6. ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services", ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.
7. G. J. Sullivan, A. M. Tourapis, T. Yamakage, C. S. Lim, eds., "Draft AVC amendment text to specify Constrained Baseline profile, Stereo High profile, and frame packing SEI message", Joint Video Team (JVT) Doc. JVT-AE204, London, United Kingdom, July 2009.
8. A. Vetro, T. Wiegand, G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard", Proceedings of the IEEE, Vol. 99, Issue 4, pp.626-642, April 2011.
9. K. Müller, P. Merkle, T. Wiegand, “3-D Video Representation Using Depth Maps,” Proceedings of the IEEE, Vol. 99, Issue 4, pp.643-656, April 2011.
10. A. Vetro, A.M. Tourapis, K. Muller, T. Chen,, "3D-TV Content Storage and Transmission", IEEE Transactions on Broadcasting, Vol. 57, Issue 2, Part 2, pp. 384-394, June 2011.
11. T. Schierl, S. Narasimhan, “Transport and Storage Systems for 3-D Video Using MPEG-2 Systems, RTP, and ISO File Format,” Proceedings of the IEEE, Vol. 99, Issue 4, pp.671-683, April 2011.
12. A. Vetro, "Frame Compatible Formats for 3D Video Distribution", In Proc. ICIP2010, 2010.
13. J. Konrad and M. Halle, “3-D Displays and Signal Processing – An Answer to 3-D Ills?”, *IEEE Signal Processing Magazine*, vol. 24, no. 6, Nov. 2007.
14. P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, “Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability”, *Signal Processing: Image Communication. Special Issue on 3DTV*, Feb. 2007.
15. R. Patterson, L. Moe, T. Hewitt,"Factors that affect depth perception in stereoscopic displays", Human Factors, 34(6):655–667, 1992.