1 Public executive summary

In the area of 3D video representation and coding, recent research and development activities have introduced different formats for specific application areas, as well as coding approaches for efficient compression of multi-view and 3D video data.

Conventional stereo video (CSV) consists of a pair of sequences, showing the same scene for the right and the left eye view. It doubles the amount of data to be stored or transmitted. A derivative of CSV is the mixed resolution format (MRS), which consists of a pair of sequences, showing the same scene for the right and the left eye view. In contrast to CSV one of the two sequences is sub-sampled. Instead of transmitting a 2 view color video (CSV) a single view and an associated depth map can be used. The second view can be rendered using depth-image-based rendering.

A more general representation is to have different multi-view video formats, where a number of input cameras capture a scene from different viewpoints. A specific case under investigation is to use linear N view camera settings, which play a special role for 3D Video applications. Having N views form slightly different viewpoints allows for a 3D impression within a limited range by presenting 2 adjacent of the N views as a stereo pair to the user.

A much more advanced format is multi-view video plus depth (MVD). It is based on the assumption, that the number of views for transmission can be reduced, if additional information about the recorded scene is available. For MVD, this information are per pixel depth maps, which are provided for each remaining view. Layered depth video is a derivative and alternative to MVD. It uses one color video with associated depth map and a background layer with associated depth map. The background layer includes image content which is covered by foreground objects in the main layer.

A strong requirement for 3D video formats is the compatibility to existing stereo displays, which are mainly based on conventional stereo up to know. To combine the good quality of simple stereo video with the extended functionality of depth data availability, the depth-enhanced Stereo format (DES) was created. It contains two views as a stereo pair, together with depth data and possibly occlusion color and depth data for both views.

Currently, a number of different coding methods exist, which are part of international standards. All methods derive from 2D video coding and contain certain structures and coding features, which can be used to code the 3D video representation formats, described above. In H.264/AVC simulcast one coder is applied to N video sequences in a generic way resulting in N encoded bit- or transport-streams. With MPEG-4 AVC Stereo SEI a CSV sequence is interlaced and coded in field coding mode. Dedicated flags tell the decoder that the output has to be deinterlaced into two individual view sequences. Multi-view Video Coding (MVC) is an extension of the Advanced Video Coding (AVC) standard that provides efficient coding of multi-view video. The encoder receives N temporally synchronized video streams and generates one bit stream. The decoder receives the bit stream, decodes and outputs the N video signals. MPEG-C part 3 was specifically developed for stereo video content in the form of one video sequence with color and depth data. The video and the depth sequence are encoded independently, resulting in two bit-streams.

The current state of the art in 3D video coding is MVC for multiple color views, i.e. for the MVV[1] format. The other enhanced formats, like multi-resolution stereo or depth-enhanced formats can currently only be coded, using separate coders, like single view AVC for color and depth in the V+D format. Similar, for the MVD format, separate MVC for color and depth can be applied.

[1] MVV (Multi-view Video): N color input images

The restrictions of currently used multi-view formats and coding approaches show, that further extensions towards a new coding standard are required. These new approaches will incorporate advanced depth-enhanced formats and adapted coding methods, especially for efficient compression of depth data. In particular the linear dependency of coding bit rate from the number of target views has to be broken. Furthermore statistical properties of depth data has to be considered to better code such data. These requirements are addressed in the 3D video ad hoc group where the 2020 3D Media project is strongly involved with WP3.

For advanced 3D audio, the technology developed during the first half of the project has been extensively reported in deliverables D5.5 and D6.2. It makes use of a format that combines different coding and decoding technologies to reproduce different sounds of a 3D soundtrack.

This document discusses how this manner of encoding 3D soundtracks could be incorporated into the MXF[2] container already used for Digital Cinema Package (DCP). Although it would not fit as-is into the MXF specifications, it would be easy to use allowed MXF extensions for the purpose, which would simply require the use of a non (yet) standard codec in the playout system guaranteeing a correct flow.

The MXF container format deals with different "essence" streams encoded with any of a variety of codecs, together with metadata. It has been designed to be platform-agnostic and extendible for future applications by adding new definitions of mappings between data and codecs.The 3D audio format basically requires multiple audio streams and scene-based metadata, therefore, the MXF container is a perfect candidate for the proposed 3D audio format. Such integration would require introducing, and eventually standardising, a new codec within the MXF framework. Moreover, the existence of open-source tools for encoding and decoding MXF files enhances the adoption of the new format and the interoperability among different applications.

In the context of the digital cinema, the DCI standards, and the 2020 3D Media project in particular, choosing MXF as a container for a new 3D audio format seems to be the right choice because the DCI distribution packages (DCP) is already based on the MXF container format, and all the tools in the cinema chain already supports MXF.

[2] MXF (Material Exchange Format): XML based container used for Digital Cinema Package (DCP)