Overview of 3D Video Coding Standardization

Jens-Rainer Ohm
Institute of Communication Engineering (IENT)
RWTH Aachen University
52056 Aachen, Germany

***Abstract*—**Several 3D video coding standards are currently developed by the Joint Collaborative Team on 3D Video Extension Development (JCT-3V), a joint working group of ISO/IEC MPEG and ITU-T VCEG. To support advanced applications and displays with wider range and continuous view adaptation, efficient compression of video texture and depth data is targeted. This overview reports on an extension to the Multiview Video Coding (MVC) of the Advanced Video Coding (AVC) standard, such that depth data can be encoded along with each view; an extension of AVC (starting from mono view data), where depth data are also used to achieve efficient compression of additional camera views; an extension of the new High Efficiency Video Coding (HEVC) standard for stereo and multi-view compression; investigations to further extend HEVC for even more efficient joint compression of video texture and depth data. The talk will also summarize on possible future trends in 3D video coding standardization.

Keywords: 3D video coding; AVC; MVC; HEVC; video standardization; MPEG; VCEG; JCT-3V.

I. INTRODUCTION

3D video is intended 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 which require special-purpose glasses, to more sophisticated multiview auto-stereoscopic displays that do not require glasses, up to holographic displays which provide a large continuum of views from various directions. In more advanced displays, it is desirable or even required to adjust depth perception by automatic means or through an interaction with the end user. As a consequence, the data throughput relative to conventional stereo displays becomes much larger, since the 3D impression is achieved by essentially emitting multiple complete video sample arrays in order to form view-dependent pictures. This puts additional challenges to representation formats and compression, which should deliver high quality data with as small amount of bits as possible. One key method to achieve this is the usage of depth or disparity data along with the video texture, which can then be used to generate additional views by synthesis methods known as *image based rendering* [1]. Even though such rendering can be implemented in different ways and is usually outside of the normative scope of standardization, the representation of depth maps and their meaning needs to be specified. In advanced methods, depth maps and their coherency with the video texture can further be exploited for a more compact representation of the overall 3D video. This paper reports about recent developments in this field, which were performed in ISO/IEC MPEG's and ITU-T VCEG's Joint Collaborative Team on 3D Video Coding Extensions Development (JCT-3V).

II. MULTIVIEW VIDEO CODING (MVC) AND ITS EXTENSION BY DEPTH MAP CODING

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 1996 MPEG-2 amendment 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. 1. 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.

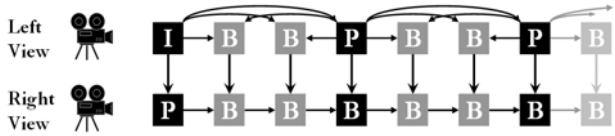

Fig.1 Typical prediction structure of MVC

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 use 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 can be found in [2].

By January 2013, JCT-3V finalized the *MVC-plus-depth* (nicknamed MVC+D) extension of MVC [3], which allows to include encoded depth maps in the video stream. The main target of this work item is to enable 3D enhancements while maintaining MVC stereo compatibility. The target was to encapsulate coded texture and depth maps into a single bitstream with minimal changes to the MVC specification. Macroblock-level changes to the AVC or MVC syntax, semantics and decoding processes were not considered.

The texture data coding is compatible with MVC profiles (MVC High and Stereo High), and a new profile identifier is introduced for coding of multiview depth map data. Multiview texture and depth data are coded independently, and no restrictions on strict association between texture and depth map is imposed. In addition to this, texture and depth map data can be coded at different spatial resolution, were mostly reduced resolution for depth map data is used, see Fig. 2.

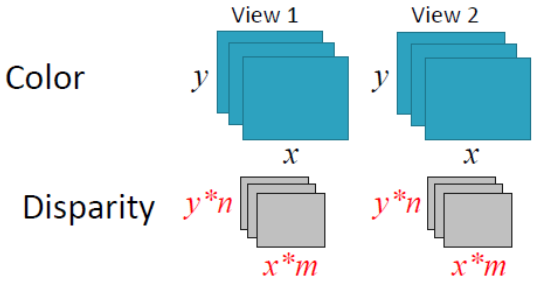

Fig.2 MVC+D data with depth map at reduced resolution

III. AVC COMPATIBLE VIDEO-PLUS-DEPTH EXTENSION

MVC+D codes depth maps independently, and due to the fact that they come in addition to the texture, it is not possible to code the 3D video at lower rates than the stereo video, when the same quality of texture shall be retained. In order to resolve this barrier, another extension of the AVC standard is currently under development, which is targeted to be finalized by end of 2013. It is referred to as 3D-AVC [4]. The specification requires that only the base texture view is compatible with AVC. However, the compatibility of dependent texture views to MVC is provided, as the foreseen profile requires decoders to be capable of decoding MVC and MVC+D streams as well.

Similarly to MVC+D, 3D-AVC supports coding of texture and depth map data at different spatial resolutions, particularly with reduced resolution for depth map comparing to the texture. In addition to this, 3D-AVC exploits inter-component dependencies between texture and depth and introduces joint coding of texture and depth data. Inter-component dependency is exploited in two directions, as shown in Fig.3. The depth map of the base view is coded with information extracted from the coded texture view, whereas dependent texture views are coded with information available from the associated depth, which is coded prior to the texture. An example of the first case is usage of texture motion vectors for depth[1]; an example for the second case is *view synthesis prediction*, where depth-controlled image rendering is performed to achieve a better prediction than would be possible by motion compensated or ordinary disparity compensated prediction of MVC. Beyond that, methods of illumination compensation are used to better predict the texture of the dependent view, which could be different by illumination effects or due to different camera transfer characteristics.

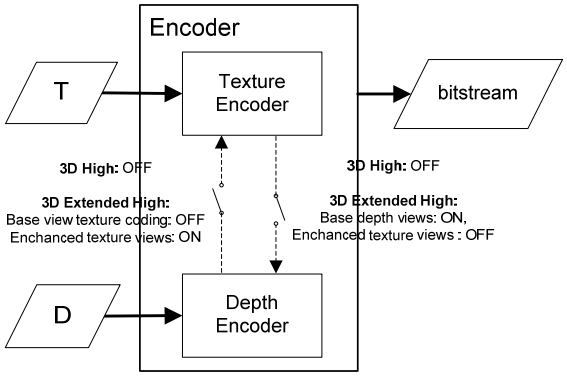

Fig.3 MVC+D data with depth map at reduced resolution

The technology included in the current draft of 3D-AVC [4] provides a significant improvement in coding efficiency in comparison to the MVC+D, recently reported gain is about 25% on average rate savings for the total bitrate, compared to MVC+D at same quality.

IV. 3D VIDEO WITH HEVC-BASED CODING TECHNOLOGY

High-efficiency video coding (HEVC) is a new video coding standard for which the base version (2D) has been finalized by January 2013 [5]. For a more deep insight into the compression tools and architecture of HEVC, the reader is referred to [6]. A primary usage of HEVC is seen in the area of high and ultra-high definition (UHD) video. Many HD displays already provide the capability for stereo rendering, and it can be expected that the increased resolution and display size of UHD displays makes them even more suitable for such purposes. Beyond that, the improved compression capability of HEVC (approximately half bit rate with same quality compared to AVC High profile) makes it attractive for the introduction of stereo; for example, it can be expected that by using mechanisms that exploit the redundancy between views, HEVC would be able to encode full resolution stereo at significantly lower rates than AVC would need just for one (monoscopic) view at the same quality and resolution.

Similar to the AVC-based projects, JCT-3V is performing development study of two 3D video solutions that are using the HEVC coding technology. The first is a multi-view extension of HEVC, so called MV-HEVC and another is a depth enhanced HEVC-based full 3D video codec, 3D-HEVC. This section provides a review on these two developments.

[1] Note that in the most recent version of 3D-AVC, texture-dependent coding of depth has been removed, in order to achieve a better compatibility with MVD+D

To achieve higher compression efficiency than simulcast coding of multiview texture data, HEVC-based multi-view coding is currently developed, which is conceptually similar to the MVC extension of AVC that was introduced in section II. MV-HEVC targets to utilize a redundancy between different texture views captured from the same scene, by re-purposing the existing motion-compensated prediction as disparity-compensated inter-view prediction. A base view is defined, which can be decoded by a (monoscopic) HEVC main profile legacy decoder. The whole approach is simply defined by extending the high-level syntax appropriately, by rearrangement of decoded picture buffers to store the reference pictures as needed, without any changes to the core of the coding layer below the level of coded tree blocks (CTB). Furthermore, high-level syntax elements are added to identify which pictures belong to each view, as well as allowing standalone extraction of the base view. Under the current draft specification, MV-HEVC does not include coding of depth data. If however justified by market requirements, carriage of depth data could be included again in a similar fashion as in AVC's MVC depth extension (i.e. carrying depth maps as an independent monochrome channel without change of low-level coding tools). The specification of the multiview extension is planned to be finalized by early 2014, such that HEVC will soon support more efficient compression of stereo content than would be achievable by so-called frame compatible formats, which place the pictures from different views into a monoscopic frame (e.g. left/right, top/bottom), but cannot take benefits from inter-view redundancy, and are not backward compatible with monoscopic decoders by simply extracting the base view substream. Furthermore, as no low-level coding tools are changed in MV-HEVC, it will be very easy to re-purpose existing decoders for the stereo/multiview case. The most recent specification of the HEVC multiview extension can be found in [7].

With the advancement of UHD display technology and the capability to present stereoscopic views with higher resolution, the demand for higher compression capability is again expected to arise together with advanced display features supported by depth maps. Therefore, joint compression of video texture and depth maps is becoming even more attractive, which is currently explored in JCT-3V by an extended codec concept nicknamed as 3D-HEVC.

Similarly to 3D-AVC concept described in section III, the 3D-HEVC design exploits inter-component dependencies between texture and depth and introduces joint coding of texture and depth data. However, the concept is slightly different in that the depth map of a dependent view is not allowed to be utilized when coding the texture of the dependent view, i.e. the coding order is *texture first* for all views. Alternatively, the depth map of the base view can be used to perform view synthesis prediction in the dependent view, which requires some additional tricks since the corresponding areas of the two views are not co-located. On the other hand, texture-to-depth dependency, such as usage of texture motion.

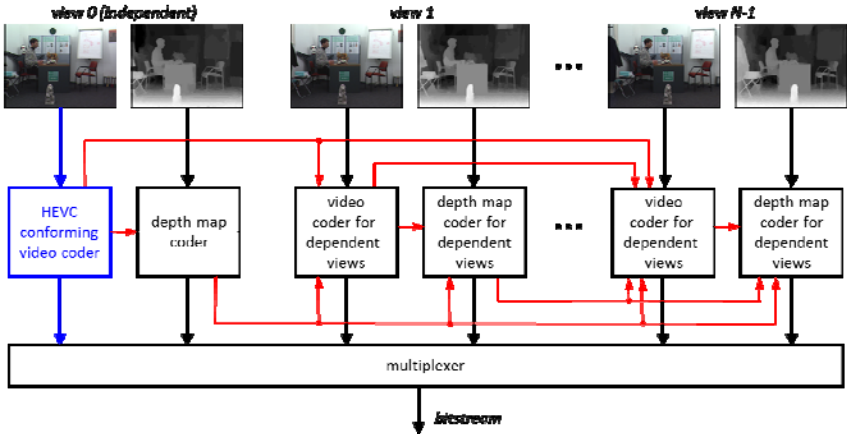

Fig.4 High level flowchart of the 3D-HEVC encoder

vectors for the depth maps in the dependent views. An example flowchart of 3D-HEVC coding is shown in Fig. 4.

In addition, the 3D-HEVC development investigates development of more sophisticated and possibly simplified (lower complexity) depth map coding in the sub-CTB level of the core codec. This is motivated by the fact that the structure of depth maps significantly deviates from video texture, in that they usually show much more constancy (flat areas or gradual changes) and significant discontinuities at object edges. Therefore, specific coding tools (entitled as depth modelling modes – DMM) are defined, which allow to characterize the depth within a block by an edge (whose position could also be derived from the texture) and the depth values on both sides. Furthermore, not the whole depth range may be present in a given depth picture, which can be exploited by coding the depth via a depth lookup table (DLT).

The newest test model of 3D-HEVC can be found in [8]. Similar to the requirements for 3D-AVC, it is anticipated that the market adoption of such an approach, which unlike the case of MV-HEVC requires re-design of decoders, would only be justified when about 25% of total bit-rate reduction in comparison to MV-HEVC is achieved. Therefore, the launch of an HEVC extension for closely integrated video plus depth solutions will hardly happen before 2014 or 2015.

CONCLUSIONS AND OUTLOOK

In this overview, new developments in international standardization for 3D video compression were presented. Currently, the main focus is on video-plus depth formats that would enable delivery of high-quality data for feeding the view generation needed for advanced 3D displays. However, the range of views that can be supported by depth based image rendering is usually relatively narrow, as artifacts caused by geometric distortion or wrong rendering of occluded areas significantly increase when the synthesized view is farther away from any available original view. This could eventually be resolved when true 3D models of objects or scenes were used, and rendering would be performed more in the fashion of 3D computer graphics. This however also would require reliable and complete 3D structure analysis of the video scene, which may be difficult to achieve. Alternative approaches,

which rather introduce a larger amount of geometrical distortions, but do not produce artifacts in case of occlusions, are warping formats or global-depth based formats, which are currently also investigated in JCT-3V [9].

ACKNOWLEDGMENT

Figures and some text elements of this paper were extracted from the MPEG and JCT-3V documents [10] and [11], that the author of this invited overview talk had edited jointly with Anthony Vetro, Dmitri Rusanovskyy and Karsten Müller.

REFERENCES

JCT-3V and -VC documents listed below are publicly available from the web sites http://phenix.it-sudparis.eu/jct3v/ and http://phenix.it-sudparis.eu/jctvc/, respectively.

[1] E. Izquierdo, J.-R. Ohm : "Image-based rendering and 3D modeling : A complete framework," Signal Processing : Image Communication 15 (2000), pp. 817-858

[2] 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.

[3] “MVC Extension for Inclusion of Depth Maps Draft Text”, document JCT3V-C1001, Geneva, January 2013.

[4] JCT3V number of 3D AVC DAM .

[5] “High Efficiency Video Coding (HEVC) text specification draft 10”, document JCTVC-L1001, Geneva, January 2013.

[6] G. J. Sullivan, J.-R. Ohm, W.-J.Han and T. Wiegand: “Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE Trans. Circ. Sys. Video Tech. (22), no. 12, pp. 1649-1668, Dec 2012

[7] “MV-HEVC Draft Text 4”, document JCT3V-D1004, Incheon, April 2013

[8] “3D-HEVC Test Model 4”, document JCT3V-D1005, Incheon, April 201

[9] “JCT3V AHG Report: Alternative Depth Formats”, document JCT3V-D0008, Incheon, April 2013

[10] “White Paper on State of the Art in 3D Video”, ISO/IEC JTC1/SC29/WG11 (MPEG) document N12142, Stockholm, July 2012

[11] “Work Plan in 3D Standards Development”, document JCT3V-B1006, Shanghai, October 2012