3D Video: Acquisition, Coding, and Display

Philipp MERKLE, *Student Member, IEEE*, Karsten MÜLLER, *Senior Member, IEEE*, and
Thomas WIEGAND, *Member, IEEE*

Abstract-- **An overview of the 3D video processing chain is given, highlighting existing and upcoming technologies and standards, addressing dependencies and interactions between acquisition, coding and display, and pointing out requirements, constraints and problems of individual modules.**

I. Introduction

3D video is commonly understood as a type of visual media that provides depth perception of the observed scenery. This is achieved by specific 3D display systems that ensure that a specific different view is projected into each eye of the user. Currently 3D video is entering broad and most probably sustainable mass markets. Cinemas are being continuously upgraded to 3D, content creators in Hollywood and elsewhere are producing more and more movies in 3D and available material is being converted from 2D to 3D. Unlike in previous attempts, technology is now maturated, providing excellent quality.

With more and more 3D cinemas and 3D movies being available, 3D also becomes increasingly interesting for home entertainment or mobile applications. The appropriate 3D video processing chain, from acquisition via coding to display, will be covered in this paper, with main focus on coding and standardization. Standardized media formats for representation, coding and transmission are essential, because interoperability and compatibility are crucial for the success of 3D video. In particular decoupling of content creation from display technology has to be achieved.

Three types of formats need to be distinguished for the 3D video processing chain. First, the production format used during acquisition, second, the transport format used for coding and transmission, and finally, the display format used for output on a display device. As a consequence, a lot of different 3D video formats are available, most of them related to specific display types. The representation format strongly influences the design of the whole 3D processing chain. Having this determined by the display type causes inflexibility of the whole 3D processing chain [1].

The 3D video formats can be divided into two main classes: video-only formats and depth-enhanced formats. Video-only formats are: classical stereo video (CSV) with 2 views, mixed resolution stereo video (MRS) with one view spatially subsampled, and multiview video (MVV) with more than 2 views. Depth-enhanced formats are: Video plus depth (V+D), multiview video plus depth (MVD), and layered depth video (LDV). One basic problem of the video-only formats is that they do not support the adaptation of 3D video content to the actual display conditions. The 3D impression varies with the viewing position, display resolution, distance to the screen, etc. Video-only content produced and optimized for cinema, looks different on a home television or on a mobile device and would require an adaptation of baseline and depth range. However, these requirements are supported by depth-enhanced formats. The depth or disparity information included in the representation allows for adaptation to different displays by rendering or synthesizing virtual views. Therefore current activities concentrate on developing new technologies and standards for depth-enhanced formats. In the following sections, an overview of existing and upcoming developments for 3D video is given, addressing dependencies and interactions between acquisition, coding and display as well as requirements, constraints and problems of the individual modules.

II. 3D Video Acquisition

3D video acquisition consists of two main parts: capturing and post-processing. Setups with 2-3 cameras are typically used for capturing 3D video. Problems and challenges with multi-camera systems are temporal synchronization, geometrical calibration, and color balance between the individual cameras. Furthermore depth sensor enhanced camera setups are rarely used, because of the limited spatial resolution and depth range of available sensors. Future development in this field might enable direct capture of depth-enhanced 3D video.

For coding and display purposes, the captured sequences need to be converted from the production into the transport format by post-processing. Regarding the video-only formats minor adjustments, like color correction, subsampling or color format conversion, might be necessary, while more complex algorithms, like rectification and depth estimation, are required for the depth-enhanced formats. Various algorithms for estimating depth or disparity maps from 2-view [2], 3-view (within MPEG), and *N*-view video [3] have been developed, but they are still error-prone and can be highly complex. Further advancement of depth estimation algorithms is expected, as highly accurate depth maps are mandatory for the success of depth-enhanced 3D video formats.

III. 3D Video Coding

Since 3D video formats consist of at least two video sequences and possibly additional depth data, efficient compression is essential for 3D video. Regarding the different video-only formats, existing standards are already widely established, especially from the H.264 family of coding standards: H.264/AVC [4] simulcast coding, where each video sequence is processed independently, can be applied to CSV, MRS, and MVV. Coding efficiency is limited, as inter-view correlations are not exploited. H.264/AVC Stereo SEI

978-1-4244-4316-1/10/$25.00 ©2010 IEEE

Message coding can be applied to CSV. The two sequences are interlaced and encoded dependently in field coding mode, using inter-field prediction. H.264/MVC [5] was designed for multiview video coding and can therefore be applied to CSV and MVV. The sequences are encoded dependently, using inter-view prediction. The coding efficiency is about as good as for H.264/AVC Stereo SEI Message coding, both outperforming H.264/AVC Simulcast coding.

Regarding the depth-enhanced 3D video formats, the situation is different, as standardized coding algorithms are not available yet. Motivated by evolving market needs, MPEG has started an activity to develop a generic 3D video standard within the 3DVC ad-hoc group. One objective is to enable stereo devices to cope with varying display types and sizes, and different viewing preferences. This includes the ability to vary the baseline distance for stereo video. A second target is to facilitate support for high-quality auto-stereoscopic displays. Providing all the necessary views for these displays is not possible, so that the new format aims to enable the generation of many high-quality views from a limited amount of input data, e.g. the video data of 2-3 cameras and additional auxiliary information such as depth or disparity maps. A key feature of this new 3D video data format is to decouple the content creation from the display requirements, while still working within the constraints imposed by production and transmission. Furthermore, compared to the existing coding formats, the 3DV format aims to enhance 3D rendering capabilities beyond V+D, while not incurring a substantial rate increase. Simultaneously, at an equivalent or improved rendering capability, this new format should substantially reduce the rate requirements relative to sending multiple views directly with an MVC or simulcast coding format.

New and optimized coding algorithms are required for such a 3D video format. The depth images that come with the depth-enhanced formats represent the scene surface, so that their characteristics differ significantly from video data. Video codecs like H.264 are highly optimized to the statistical properties and human perception of video data, but depth images are never displayed directly and depth coding artifacts lead to geometry distortions in the 3D scene representation which propagate into rendered virtual views. A promising approach that meets these specific requirements is Platelet-based depth coding [6]. Overall the development of advanced 3D video coding algorithms needs to optimize the performance with respect to both video and geometry distortions and in consideration of the quality of rendered virtual views. This requires addressing joint compression and rendering algorithms as well as appropriate quality metrics.

IV. 3D Video Display

Currently, various types of 3D displays are available and under development [7]. Most of them use classical 2-view stereo with one view for each eye and some kind of glasses (polarization, shutter, anaglyph) to filter the corresponding view. Different input formats and interleaving patterns are used in various solutions. Then there are so called auto-stereoscopic displays which do not require glasses. Here, 2 or more views are displayed at the same time and a lenticular sheet or parallax barrier element in front of the light emitters ensures correct view separation for the viewer's eyes. Such displays use even other input formats, interleaving and in most cases more than 2 views. Beyond that development of holographic displays showed promising progress recently.

For display purposes the transport format data needs to be converted into the display format as a final step in the 3D video processing chain. The video-only formats can be directly displayed, but as already mentioned the whole processing chain, especially capture, needs to be adapted to the particular display system. Depth enhanced formats are much more flexible in this regard, because virtual views can be generated via depth-image-based rendering (DIBR). Within practical limits DIBR allows to render several virtual views at arbitrary positions, so that various display formats can be supported. Although the basic principle of 3D image warping is well-defined, various algorithms for DIBR have been developed [8], but they are still error-prone and can be highly complex.

V. Conclusion

Efficient algorithms, standards, and technologies for end-to-end systems with video-only formats already exist, but it is not decided yet, which one will be the most widely accepted and supported. As an important next step, a MPEG coding standard for 2-3 view depth-enhanced formats and support for 2-view and multiview displays is under development. Further development will complete and optimize the processing chain for depth-enhanced formats.

References

[1] A. Vetro, S. Yea, and A. Smolic, “Towards a 3D Video Format for Auto-Stereoscopic Displays”, *Proc. SPIE Conference on Applications of Digital Image Processing XXXI*, Vol. 7073, September 2008.

[2] N. Atzpadin, P. Kauff, and O. Schreer. “Stereo Analysis by Hybrid Recursive Matching for Real-Time Immersive Video Conferencing”, *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 14, no. 3, pp. 321-334, March 2004.

[3] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-Quality Video View Interpolation Using a Layered Representation”, *ACM SIGGRAPH and ACM Trans. on Graphics*, Los Angeles, CA, USA, August 2004.

[4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 13, no. 7, pp. 560-576, July 2003.

[5] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, “The Emerging MVC Standard for 3D Video Services”, *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, No. 1, January 2009.

[6] P. Merkle, Y. Morvan, A. Smolic, K. Mueller, P. H. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, pp. 73-88, January 2009.

[7] 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.

[8] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, “Intermediate View Interpolation based on Multi-View Video plus Depth for Advanced 3D Video Systems”, *Proc. IEEE Internat. Conf. on Image Processing (ICIP'08)*, San Diego, CA, USA, Oct. 2008.