STEREO VIDEO COMPRESSION FOR MOBILE 3D SERVICES

P. Merkle[(1)], H. Brust[(1)], K. Dix[(1)], K. Müller[(1)], and T. Wiegand[(1)(2)]

(1) Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institut
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany
{merkle/brust/dix/kmueller/wiegand}@hhi.de

(2) Image Communication Chair
Department of Telecommunication Systems
School of EE and CS
Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

This paper presents a study on different techniques for stereo video compression and its optimization for mobile 3D services. Stereo video enables 3D television, but as mobile services are subject to various limitations, including bandwidth, memory, and processing power, efficient compression is required. Three of the currently available MPEG coding standards are applicable for stereo video coding, namely H.264/AVC with and without stereo SEI message and H.264/MVC. These methods are evaluated with respect to the limitations of mobile services. The results clearly indicate that for a certain bitrate inter-view prediction as well as temporal prediction with hierarchical B pictures lead to a significantly increased subjective and objective quality. Although both techniques require more complex processing at the encoder side, their coding efficiency offers the chance to realize 3D stereo at the bitrate of conventional video for mobile services.

Index Terms— 3D video, video coding, stereo, MVC, mobile services.

1. INTRODUCTION

Interest in 3DTV has remarkably increased recently with more and more products and services becoming available for the consumer market. 3DTV is commonly understood as a type of visual media that provides depth perception of the observed scenery and is also referred to as stereo video. Such 3D depth perception can be provided by 3D display systems which ensure that the user sees a specific different view with each eye [1]. Initiated by the recent popularity of 3DTV, extensive activities for developing new technologies and standards for the complete processing chain can be observed, including production, representation, compression, storage, transmission, and display. With the content being produced, 3D video is also an increasingly interesting technology for home user living room applications and beyond for mobile 3D video services.

Among the various 3D video representations, stereo video is the most widely-used because simple format, involving only color pixel video data. Captured by at least two cameras, the resulting video signals may undergo some processing steps like normalization, color correction, rectification, etc., but in contrast to other 3D video formats no scene geometry information is involved. Stereo video can be directly displayed on a 3D display system without any depth-based rendering. In return the 3D impression cannot be modified with stereo video, as the baseline is fixed from capturing, so that the depth perception cannot be adjusted to different display types and sizes. Compared to other 3D video formats like video plus depth the functionality of stereo video is limited, but is reliable and robust, due to the lack of error-prone processing steps such as depth estimation or view synthesis.

In this paper three different H.264-based coding methods for stereo video are evaluated and compared with respect to their applicability for mobile 3D services. Section 2 specifies the coding methods and their adaptation to stereo video. Simulation results and conclusions are given in sections 3 and 4, respectively.

2. STEREO VIDEO CODING

Stereo video consists of a pair of sequences, showing the same scene for the right and the left eye view. Compared to conventional video, stereo video has twice the amount of data to be stored or transmitted. Especially for mobile video services with its bandwidth and memory limitations, very efficient compression of stereo video is required to realize 3D instead of conventional video. However, efficient compression of stereo video takes advantage of the fact that the left and the right view of a stereo pair show the same scene from slightly different perspectives and are therefore highly redundant. Three of the currently available MPEG coding standards are applicable for stereo video coding, namely H.264/AVC with and without stereo SEI message and H.264/MVC, and the following sections describe their characteristics and their adaptation to stereo video.

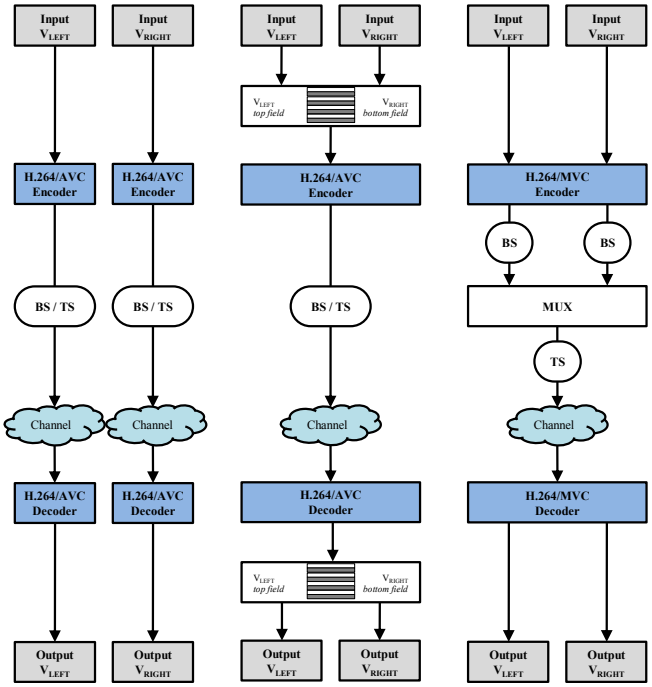

Fig. 1. Schematic block diagrams for H.264/AVC Simulcast (left), H.264/AVC Stereo SEI Message (middle) and H.264/MVC (right) coding with stereo video format data.

2.1. H.264/AVC Simulcast

According to the H.264/AVC standard [2], "H.264 Simulcast" is specified as the individual application of an H.264/AVC conforming coder to several video sequences in a generic way. H.264/AVC is the latest video coding standard of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). H.264/AVC has recently become the most widely accepted video coding standard and covers all common video applications ranging from mobile services and videoconferencing to IPTV, HDTV, and HD video storage.

For stereo video the overview diagram in Fig. 1 illustrates the coding procedure of H.264/AVC Simulcast with the left and right view of a stereo pair. The H.264/AVC encoder is applied to each of the two input sequences independently, resulting in two encoded bit- or transport-streams (BS/TS). After transmission over the channel the two streams are decoded independently, resulting in the distorted video sequences of the stereo pair.

2.2. H.264/AVC Stereo SEI Message

According to the H.264/AVC standard [2], the "Stereo video information SEI message" is specified as follows: This SEI message provides the decoder with an indication that the entire coded video sequence consists of pairs of pictures forming stereo-view content. It defines six flags to control the mapping of frames or fields of the coded video sequence to the left and right view of a stereo pair at the decoder. The display process itself is not specified in the standard. For H.264/AVC Stereo SEI Message coding the encoder is operating in field coding mode, exploiting inter-view dependencies of stereo video by inter-field prediction.

The overview diagram in Fig. 1 illustrates the coding procedure of H.264/AVC Stereo SEI Message for the left and right view video sequence of a stereo pair. These two sequences are interlaced line-by-line into one sequence, where the top field contains the left and the bottom field the right view. The H.264/AVC coder is applied to the interlaced sequence in field coding mode, resulting in one encoded bit- or transport-stream (BS/TS). After transmission over the channel this stream is decoded, resulting in the distorted interlaced sequence. For output this sequence is deinterlaced to the two individual view sequences.

2.3. H.264/MVC

According to the H.264/MVC standard [3], "Multiview Video Coding" is specified as an extension to the family of H.264 standards. For MVC, the single-view concepts of H.264/AVC are extended, so that a current picture in the coding process can have temporal as well as inter-view reference pictures for motion-compensated prediction [4]. To meet additional requirements for 3D video MVC includes a number of new techniques for improved coding efficiency, reduced decoding complexity, and new functionalities for multiview operations [5].

The overview diagram in Fig. 1 illustrates the coding procedure of H.264/MVC for the left and right view video sequence of a stereo pair. The H.264/MVC coder is applied to both sequences simultaneously for inter-view predictive coding. Fig. 2 highlights how combined temporal and inter-view prediction with H.264/MVC is applied to stereo video format data. The resulting two dependent encoded bit-streams (BS) may contain the camera parameters as auxiliary information. For transmission both bit-streams are interleaved frame-by-frame in the multiplexer (MUX), resulting in one MVC-compliant transport-stream (TS). After transmission over the channel this stream is decoded (and thereby demultiplexed), resulting in the distorted video sequences of the stereo pair.

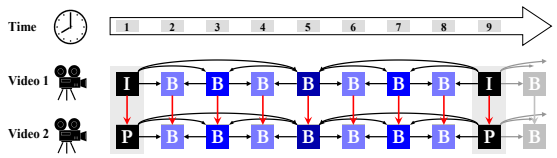

Fig. 2. MVC coding with stereo video: inter-view prediction (red arrows) combined with hierarchical B pictures for temporal prediction (black arrows).

3. EXPERIMENTAL RESULTS

The simulations for the three different stereo video coding approaches, namely H.264/AVC Simulcast, H.264/AVC Stereo SEI Message, and H.264/MVC, have been configured with respect to comparable and realistic simulation conditions for mobile applications. For the experiments we used the different coder implementations and configurations presented in Table 1. Each coding approach was evaluated with four representative stereo test data sets that cover different types and levels of scene content complexity and temporal variation (see examples in Fig. 3). The test data sets consist of a left and right view sequence, each with a resolution of 480×270 pixels, 5-10 seconds length, and a frame rate of 30 fps.

	H.264/AVC Simulcast	**H.264/AVC Stereo SEI**	**H.264/MVC**
Coder Implementation	JM 14.2	JM 14.2	JMVM 7.0
Standard	H.264/AVC	H.264/AVC	H.264/MVC
Coding Mode	normal	field coding	inter-view prediction
Quantization Parameter	24 30 36 42	24 30 36 42	24 30 36 42
GOP Size	1 (IPP...) 16 (HBP)	1 (IPP...) 16 (HBP)	2 (IBP...) 16 (HBP)
Intra Period	16	16	16
Search Range	32	32	96
Symbol Mode	CABAC	CABAC	CABAC

Table 1. Coding settings for experimental simulations (HBP = temporal prediction with hierarchical B pictures).

The results of these coding experiments are presented in Fig. 3 (left and right), comparing the objective quality in terms of RD-performance. This comparison clearly indicates that the overall RD-performance of both Stereo SEI Message and MVC is better than Simulcast coding, while Stereo SEI even performs better than MVC in some cases. Comparison of the RD-performance gains between Simulcast and Stereo SEI (using equal coding conditions) shows that for the same quality up to 35% of the total bitrate can be saved with interview prediction and up to 60% with hierarchical B pictures. However, in some cases as for the "Hands" sequence the gain is negligible.

Informal subjective expert viewing has been carried out for the simulation results of the three different stereo video coding approaches on a stereoscopic display. This lead to the conclusion that the objective RD performance can be confirmed subjectively, as for the same bitrate a lower quality is achieved by Simulcast coding than by Stereo SEI or MVC coding. In return for these two approaches a lower bitrate is necessary to achieve the same subjective quality as Simulcast.

From a complexity point of view the three coding approaches are comparable, as they are all based on H.264 with the same basic operations. Independent of the particular approach, hierarchical B pictures as well as inter-view prediction with complex prediction operations mean a considerable increase of memory requirements and delay.

4. CONCLUSIONS

This paper investigated the stereo video representation format and appropriate coding standards for 3D mobile applications. Simulations were carried out with realistic coding settings, e.g. intra period of 16 for random access and error robustness. A typical set of test sequences was used targeting display resolutions in mobile devices and covering different types of content. As for any type of video coding, the same amount of raw input data leads to very different RD-performance.

The experimental results showed that the required bitrate for achieving acceptable quality mainly depends on the complexity of the sequence content. The coding gain from inter-view prediction (Stereo SEI & MVC) varies largely, leading to a significant reduction of bitrate (up to 35% in our experiments) for some sequences, but to negligible gains for others. For temporal prediction hierarchical B pictures can be adopted to each of the evaluated coding methods, leading to significant coding gains (up to 60% in our experiments). Not using hierarchical B pictures not only results in considerably higher bitrates for the same objective quality, but even in a worse subjective quality. Again, the gain from using hierarchical B pictures differs largely for individual sequences, depending on the complexity of the sequence content. However, the higher RD performance with hierarchical B pictures is achieved for the price of increased complexity and memory requirements.

In summary the presented study shows that stereo video compression for mobile 3D services is a challenging task. Although advanced coding techniques like inter-view prediction and hierarchical B pictures require more complex processing, they lead to significant coding gains. Therefore we expect that further development and optimization of these promising approaches will enable to realize 3D stereo at the bitrate of conventional video for mobile services.

5. ACKNOWLEDGEMENTS

This work was supported by EC within FP7 (Grant 216503 with the acronym MOBILE3DTV). We would like to thank KUK Filmproduktion GmbH, Munich, Germany, for providing the *Hands*, *Car*, *Horse* and *Snail* stereo video data (http://www.kuk-film.de/).

6. REFERENCES

[1] 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.

[2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", November 2007.

[3] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding", Doc. N9978, Hannover, Germany, July 2008.

[4] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 11, pp. 1461-1473, November 2007.

[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.

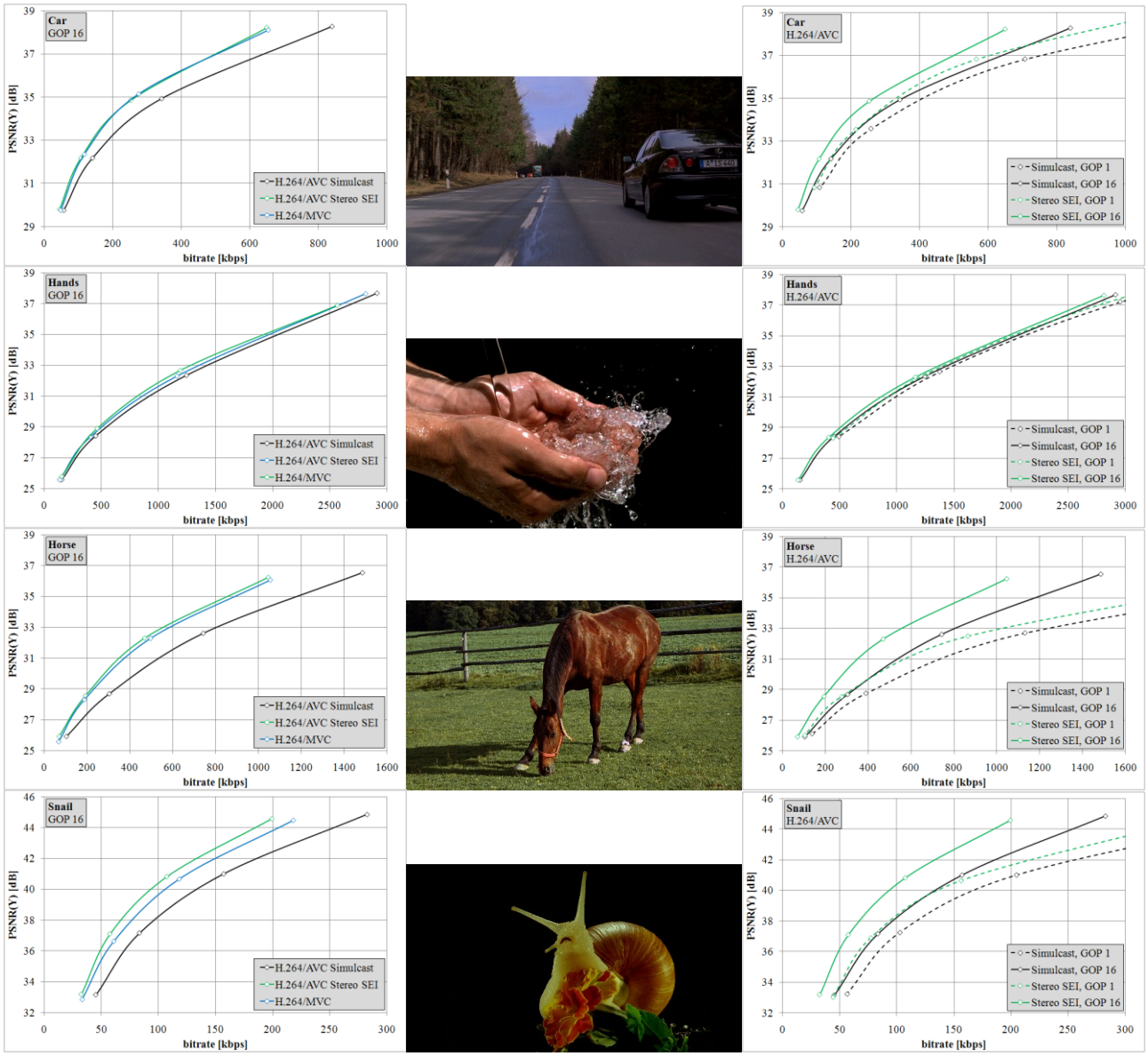

Fig. 3. Experimental results: RD-comparison (total bitrate for both views vs. average PSNR relative to the original sequences) for the three different simulations on stereo video coding approaches (left), sample pictures (middle) and RD-comparison for temporal prediction with and without hierarchical B pictures (right).