Lower End Mobile Devices by using 3D Video Compression

P. Swetha[1], Ch. Deepika[2], Ch. Sreedhar[3]
[1, 2]Assistant Professor, CSE Department, Vidya Jyothi Institute of Technology, C.B.Post. Aziz Nagar, Telangana, India
Assistant Professor[3], ECE, Nishitha College Engineering & Technology, Telangana, India

Abstract— this paper presents a study on 3D video and depth compression using MPEG standards and its optimization to the mobile services. Video and depth invoke 3D television, but as mobile services are subject to various limitations, including bandwidth, memory, and processing power, efficient compression as well as low complexity view synthesis is required. Two MPEG coding standards are applicable for video and depth coding, namely MPEG-C Part 3 and H.264 Auxiliary Picture Syntax. These methods are evaluated with respect to the limitations of mobile services and the achievable quality for rendering the second stereo view from compressed video and depth. In conclusion video and depth is an interesting alternative to conventional stereo video for mobile 3D services. The results indicate that depth can be compressed at significantly lower bitrates than a secondary video, however at the expense of increased complexity for rendering the second view at the decoder.

Keywords:— Lower end mobile devices, video and depth coding, view synthesis, stereo.

I.　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 attractive technology for home user living room applications and beyond for mobile 3D video services. The video and depth (V+D) representation is an interesting alternative to stereo video for realizing 3D video. It allows to adjust the stereo rendering at the decoder and to optimally adapt the 3D impression for any given display. However, this extended functionality comes at the cost of an increased complexity. Especially for mobile 3D services the rendering of one output view from V+D at the receiver side is crucial. In this paper two different H.264-based coding methods for V+D are evaluated and compared with respect to their applicability for mobile 3D services. Section 2 introduces the V+D format and view synthesis. Section 3 specifies the coding methods and their adaptation to V+D.

II.　V+D LAYOUT

The video and depth format consist of a conventional monoscopic color video and an associated per pixel depth map, which can be regarded as a monochromatic, luminance-only video signal. Based on the fact that depth maps are a 2D representation of the 3D scene surface, a stereo pair can be rendered from the video and depth information via view synthesis, as illustrated in Fig. 1. The depth data is usually generated by depth/disparity estimation from a captured stereo pair. Such algorithms can be highly complex and are still error-prone.
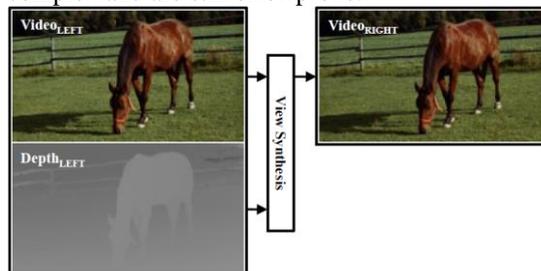

Fig.1. Video and depth format and its application: View synthesis of the second view for stereo output.

For V+D view synthesis is necessary at the receiver for generating the second view of a stereo pair to be presented on stereoscopic displays. The major challenge with V+D to stereo rendering is the visual quality of the synthesized view, as rendering artifacts may result in a wrong and thereby annoying 3D impression in case the left and right view are inconsistent. Rendering stereo views from V+D has been extensively investigated, but many of the existing algorithms are not suitable for mobile applications, as they are either too complex or require special graphic APIs and a corresponding GPU. The proposed view synthesis algorithm was developed for stereo rendering with mobile applications, intended to

realize an appropriate balance between quality and complexity. By using rectified, parallel views, the complexity of the rendering process can be dramatically reduced, allowing line-wise processing of the V+D input data without requiring a z-buffer. Main features of the implemented view synthesis algorithm are rapid pixel shifting, direct YUV processing, and straight hole-filling. Pixel shifting means that, due to rectification, pixel positions between the original and the synthesized view are only horizontally shifted by a value calculated from the corresponding depth value and a constant scaling factor. Direct YUV processing means that, unlike other V+D view synthesis rendering algorithms, no computationally intensive conversion of the YUV input stream to RGB is required, but pixel-shift rendering is directly performed on the YUV input data. Straight hole-filling means to instantly detect and fill gaps resulting from pixel pasteurization of the non-integer shift positions and from conclusions. For the latter no information is available for background areas covered by foreground objects in the original view.

III.   V+D CODING

Mobile video services with its bandwidth and memory limitations require efficient compression of video and depth for realizing 3D instead of conventional video. Two of the currently available MPEG coding standards are applicable to video and depth, namely MPEG-C Part 3 and H.264/AVC Auxiliary Picture Syntax. The following sections describe their characteristics and their adaptation to video and depth.

A. MPEG-C

According to the ISO/IEC 23002-3 standard [2], "MPEG-C Part 3" is specified as a representation format for depth maps which allows encoding them as conventional 2D sequences and additional parameters for interpreting the decoded depth values at the receiver side. Note, that the MPEG-C Part 3 video and depth standard does not specify the transport and compression techniques.
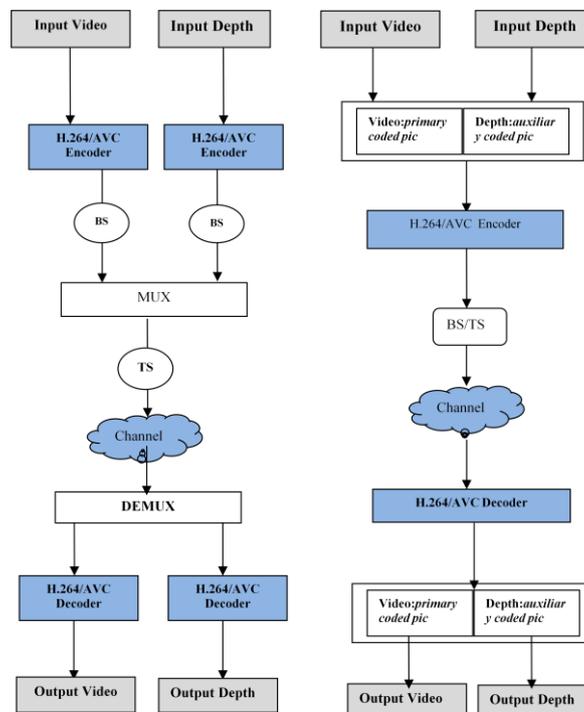

Fig. 2. Schematic block diagrams for MPEG-C Part 3 (left) and H.264 Auxiliary Picture Syntax (right) coding with video and depth format data.

The overview diagram in Fig. 1 illustrates the coding procedure of MPEG-C Part 3 in combination with the H.264/MPEG-4 AVC codec [3]. The video and the depth sequence are encoded independently, resulting in two bitstreams (BS). For transmission these two bit-streams are interleaved frame-by-frame in the multiplexer (MUX), resulting in one transport-stream (TS), that may contain additional depth map parameters as auxiliary information.

After transmission over the channel the demultiplexer (DEMUX) separates this stream into the two bit-streams. These are decoded independently, resulting in the distorted video and the distorted depth sequence for one of the two views of a stereo pair.

B.   H.264 Auxiliary Picture Syntax:

According to the H.264/MPEG-4 AVC standard (H.264) [3], "Auxiliary Picture Syntax" specifies extra monochrome pictures sent along with the main video stream. An auxiliary coded picture supplements the primary coded picture and may be used in combination with other data not specified by H.264 in the display process, for example for such purposes as alpha blend compositing or video and depth stereo rendering. Auxiliary coded pictures have the same syntactic and semantic restrictions as a monochrome redundant coded picture, must contain the same number of macro-blocks as the primary coded picture and have no normative effect on the decoding process.

The overview diagram in Fig. 1 illustrates the coding procedure of H.264 Auxiliary Picture Syntax for V+D format data. The H.264 codec is applied to both sequences simultaneously but independently (with the video being the primary coded picture and depth the auxiliary coded picture), resulting in one encoded bit- or transport-stream (BS/TS). After transmission over the channel this stream is decoded, again simultaneously but independently for primary and auxiliary coded pictures, resulting in the distorted video and the distorted depth sequence for one of the two views of a stereo pair.

IV.   THE PROVISIONAL OUTCOME

After The simulations for the two different V+D coding approaches, namely MPEG-C Part 3 and H.264 Auxiliary Picture Syntax, have been configured with respect to realistic simulation conditions for mobile applications. For the experiments we used the JM 14.2 implementation of the H.264/MPEG-4 AVC codec, with the following configuration: an intra period of 16 frames for random access and error robustness, a search range of 32, two temporal prediction structures (GOPsize = 1 for simple IPP… prediction, GOPsize = 16 for complex prediction with hierarchical B pictures), and four different qualities (QP = 24, 30, 36, 42). The coding approaches were evaluated with four representative video and depth 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 one texture video and the associated monochromatic depth sequence of the left view of a stereo pair, each with a resolution of 480×270 pixels, 5-10 seconds length, and a frame rate of 30 fps. For our simulations the video and the depth sequence are encoded and decoded independently.

The results of the described coding experiments are presented in Fig. 3 (left), comparing the objective quality in terms of RD-performance. The gains that can be achieved with hierarchical B pictures for the individual sequences differ considerably, depending on factors like scene content and depth complexity as well as temporal variation. At the same quality between almost zero and up to 50% of the bitrate can be saved with hierarchical B pictures.

As described in section 2, the V+D data of the left view is used to render the right view via view synthesis. Together with MPEG-C Part 3 compression different qualities (and thereby bitrate ratios) of coded left video and coded left depth can be combined. The influence of such combinations on the RD-performance of the rendered right view has been evaluated. Fig. 3 (right) shows the results for all possible combinations of the four video and the four depth coding qualities. The right view is rendered from compressed left view V+D, using the right view rendered from original left view V+D as a reference for PSNR calculation.

The curves combine points of constant color bitrate and points of constant depth bitrate. Apparently curves of constant color bitrate and thereby quality are steeper. In most cases increasing the depth bitrate has a stronger influence on the overall quality than increasing color bitrate. It can be concluded that good depth quality is essential for good overall quality. An envelope curve would indicate the optimum RD-performance and thus the optimum bitrate distribution between video and depth for a given total bitrate. In contrast to MPEG-C Part 3, the aforementioned optimization is not applicable to H.264 Auxiliary Picture Syntax coding, as different coding settings for video and depth are not supported.

Informal subjective expert viewing has been carried out for the simulation results on a stereoscopic display. This lead to the conclusion that the objective RD results are confirmed, as for the same bitrate a higher quality is achieved by using hierarchical B pictures for temporal prediction or in return a lower bitrate is necessary to achieve the same subjective quality.

V.   SIMULATION RESULTS

Fig.3. Provisional outcome: RD-comparison for coding of the left view video (V_L) and depth (D_L) data and temporal prediction with and without hierarchical B pictures (left), sample pictures (middle) and view synthesis of the right view for different combinations of coded left view video and depth (right).
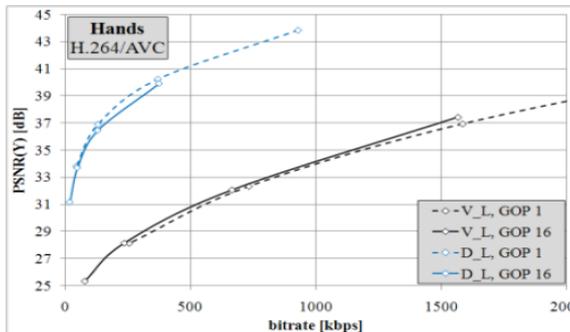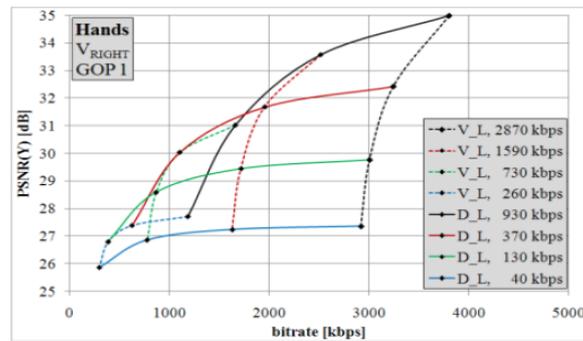
1) Results for vehicle car:
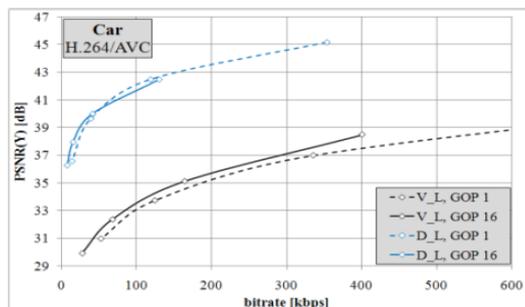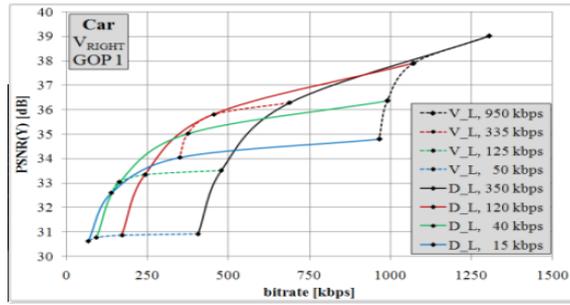
2) Results for small picture:
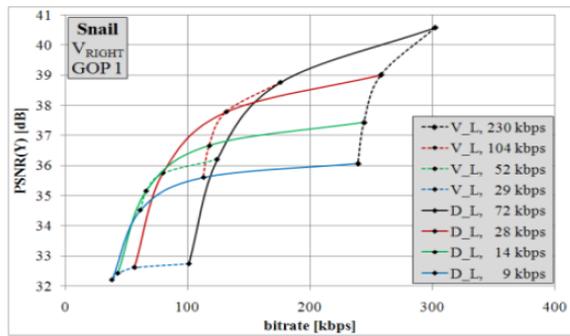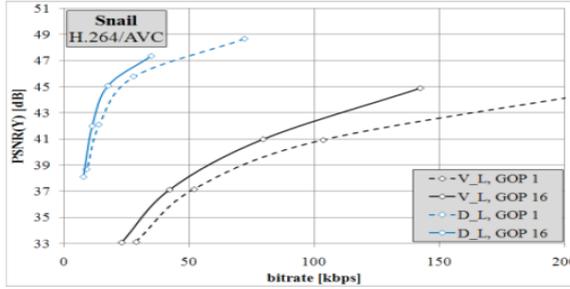
3) Results for vehicle image:

ACKNOWLEDGMENT

REFERENCES

[1]   A. Smolic, H. Kimata, and A. Vetro, "Development of MPEG standards for 3D and free viewpoint video," in *Three-Dimensional TV, Video, and Display IV*, vol. 6016 of *Proceedings of SPIE*, Boston, Mass, USA, October 2005.

[2]   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.

[3]   ISO/IEC JTC1/SC29/WG11, "ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information", Doc. N8259, Klagenfurt, Austria, July 2007.

[4]   C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291 of *Proceedings of SPIE*, pp. 93–104, San Jose, Calif, USA, May 2004.

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

BIBILOGRAPHY

P.Swetha is from HYDERABAD, TELANGANA. Completed M.TECH in CSE with specialization CSE from Vidya Vikas Institute of Technology affiliated to JNTUH in 2009.Currently .she is working as an Assistant professor in CSE department at Vidya Jyothi Institute of Technology, Hyderabad from 2014. Her areas of research interests include Data Mining, Networking, Android & Network security,Cloud Computing.

CH.Deepika is from HYDERABAD, TELANGANA. Completed M.TECH in CSE with specialization (COMPUTER SCIENCE & ENGINEERING) from Hyderabad Institute of Technology & Management affiliated by JNTUH.Currently working as Assistant professor in CSE department in Vidya Jyothi Institute of Technology, Hyderabad from 2014. Her Areas of research interests include Data Mining, Networking , Android & Network security.

CH. SREEDHAR is from HYDERABAD, ANDHRAPRADESH. Completed M.TECH in ECE with specialization (VLSI SYSTEMS DESIGN) from Kshatriya college of engineering affiliated by JNTUH in 2011and B.TECH in ECE from Aizza College of Engineering & technology affiliated by JNTUH in 2006. Currently he is working as an Assistant professor in ECE department at nishitha of Engineering & Technology, Hyderabad from 2007. His areas of research interests include Wireless&Mobilecommunications, Digitalsignalprocessing, Image processing, Telecommunications, communication systems, Signal processing, embedded systems, network security