MOBILE 3D VIDEO BROADCAST

M. Oguz Bici, Done Bugdayci, Gozde B. Akar

Middle East Technical University
Ankara, Turkey

Atanas Gotchev

Tampere University of Technology
Tampere, Finland

ABSTRACT

In this paper, we present a complete framework of an end-to-end error resilient transmission of 3D video over Digital Video Broadcasting - Handheld (DVB-H) and provide an extensive analysis of coding and transmission parameters. We perform the analysis for different coding and error resilience schemes using different contents coded at different bitrate levels. Throughout the experiments, we investigate the effects of video content type, video bitrate, coding method and unequal protection level for different channel conditions. The results show that Multi-view Coding (MVC) coding outperforms Simulcast and distribution of available bitrate between video quality and Forward Error Correction (FEC) protection is an important factor in different channel conditions.

Index Terms— Error characteristics, stereo, error resilience, DVB-H system, UEP, EEP, Simulcast, MVC

1. INTRODUCTION

Among many of the multimedia applications available for mobile devices, mobile TV has significant priority. Advances in technology contributed to the development of receivers with low power consumption, increased computational power and smart antennas with improved RF performances on significantly reduced dimensions. Last decade witnessed the emerge of various digital TV standards and their attempts to serve for the growing demand for mobile tv technology one of which is DVB-T. Soon after the discovery of poor performance of DVB-T in mobile environments, a new standard, DVB-H, is introduced specific to hand-held devices.

DVB-H is based on the existing DVB-T standard with the introduction of two new elements for mobility: Multi Protocol Encapsulation - Forward Error Correction (MPE-FEC) and time slicing [1]. MPE-FEC capability, which allows the delivery of Forward Error Correction sections in addition to datagrams delivered in MPE sections, is introduced in order to provide service under several different reception conditions. On the other hand time slicing is defined to reduce the average power consumption by enabling the transmission of data in bursts rather than a continuous transmission.

On the other hand, recent advances in 3D displays give a rise to the studies related to 3D TV broadcasting, which refers to the extension of traditional broadcasting to 3D capable displays. With the availability of switchable autostereoscopic displays, it is possible to receive and consume 3D contents on mobile devices. Even though there are mature studies on transmission of monoscopic video over DVB-H [2] [3], and a few studies on 3D video transmission over DVB-H [4] [5], none of these works study the effect of coding or transmission specific error resilient tools on the robustness under different channel conditions.

This paper has two main contributions to the studies related to transmission of 3D video over DVB-H: 1) MPE-FEC protection including unequal protection strategies are introduced for 3D video transmission over DVB-H. 2) An analysis of coding and transmission parameters for robust end-to-end transport of 3D video over DVB-H is performed through extensive simulations in order to find a compromise between compression efficiency and robustness with respect to typical channel conditions.

We concentrate on Simulcast and Multi-View Coding (MVC) coding approaches. In order to study the effect of equal and unequal error protection on these videos, we provide different levels of protection for each view according to its importance. We allocate equal channel resources for different test cases in order to make fair comparison. We achieve this by allocating the same burst duration for the total of two views. During the handling of distorted video, we employ a specific error concealment tool.

The rest of the paper is organized as follows: Section 2 gives the details of the end-to-end 3D DVB-H system and explains error-resilient tools used in the system. Section 3 presents the simulation environment and simulation results with suggestions on proper coding/protection/transmission scenarios. Finally, Section 4 gives the conclusion and future work.

2. SYSTEM OVERVIEW

The building blocks of the overall system used for the analysis of the error resilient tools are given in Figure 1 [4]. The transmission side consists of three main blocks, namely the encoder, streamer and DVB-H link layer encapsulator. Then, the system has an offline block to model the real physical transmission at the channel. The rest is the receiver side blocks which are the transport stream decapsulator, the decoder and a player. For proper display of 3D video, one would also need a multi-view video player and a 3D dedicated display.

There are several representations that can be used for 3D scenes such as multi-view video, V+D, geometry with texture (lightfield), volumetric (voxels) and holographic. Any of these representations can be used depending on the application; however, for mobile transmission applications, we use stereo video representation, a special case of multiview video, which is a very strong candidate for mobile applications.

The stereo video can be encoded using the following methods: a) Simulcast Coding: This method is used with stereo video representation. Left and right views are compressed individually using the state-of-the-art monoscopic video compression standard H.264/AVC [6]; b) MVC: This method is also used with stereo video as in simulcast. However, this time right view is encoded by exploiting the inter-view dependency using MVC [7] extension of H.264/AVC. In our work, we use IPP prediction structure with simplified prediction scheme in order to take into account the low complexity requirements by mobile devices.

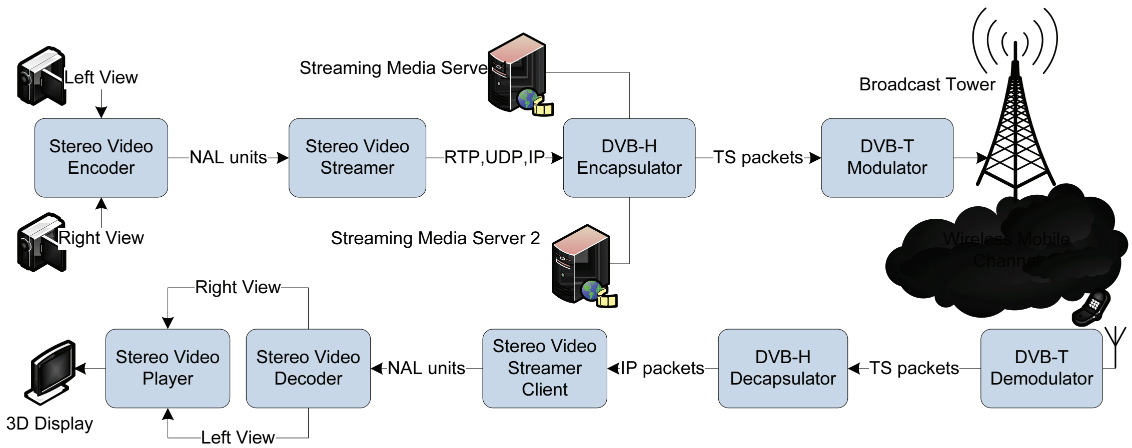


Fig. 1. Block Diagram of the End-to-End Stereo Video Transmission over DVB-H Simulation System.

Regardless of the coding method, the encoder output consists of two separate bit streams compressed and packetized appropriately for transmission. The encoder output is called Network Abstraction Layer (NAL) units (NALU). NALUs are fed to the stereo video streamer to perform RTP/UDP/IP packetization and resulting IP packets are fed to to the DVB-H link layer encapsulator, which is the third block on the transmitter side of the system. The encapsulator firstly creates MPE sections and then MPE-FEC section for forward error correction. We have modified the open source MPE encapsulator software, FATCAPS [8], in order to transmit two views simultaneously by multiplexing and to employ our error resilience schemes for robust transmission. In our system, we implemented seven different error protection modes which are Equal Error Protection (EEP) and Unequal Error Protection (6 different rates UEP5, UEP10, UEP15, UEP20, UEP25 and UEP30) modes. The definition of EEP is to protect the left and right (depth) views equally. This is implemented by adjusting the number of RS columns according to the number of application data columns and the intended FEC rate. We used typical 3/4 FEC rate for the left and right (depth) bursts in EEP mode. Then several unequal protection schemes are derived using this EEP structure. We realize six different UEP schemes by transferring (adding) a ratio of RS columns of the right view burst to the RS columns of the left view burst. The motivation behind unequal protection is that the independent left view is more important than the right or depth view. The right view requires the left view in the decoding process however, left view can be decoded without right view. The ratios of RS columns to transfer are 5%, 10%, 15%, 20%, 25% and 30% respectively. Therefore, UEP5 is closest to the EEP among the UEP schemes where right burst protection weakens as the transfer ratio increases.

The receiver side of the end-to-end system starts with the DVB-H decapsulator block. The unpacking of TS streams, construction of MPE-FEC frame table from sections, erasure decoding and possible recovery of lost data take place in this block. After these, the section payloads (IP datagram) are unpacked to output the NALUs. The next block in the receiver side is the decoder. In this block, the left and right/depth NALU bitstreams are decoded and the video streams for each view are generated. After decoding, the reconstructed left and right views are displayed on a proper 3D compatible display.

3. EXPERIMENTAL RESULTS

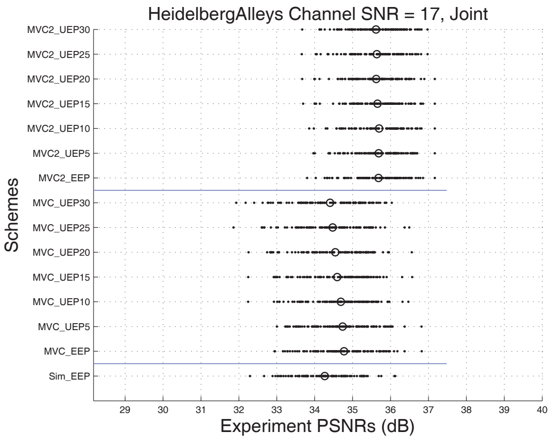
In our experiments, we simulated the physical layer operations and transmission errors due to channel variations using the DVB-H physical layer modeling introduced in [9]. DVB-H has several physical layer parameters that affect the transmission bitrate and RF performance. We worked with a commonly used set of parameters: 16QAM as the modulation scheme, 2/3 convolutional code rate, 1/4 guard interval, 8K FFT mode and 666 MHz carrier frequency which results in a channel capacity of 13.2 Mbps. For the wireless channel simulation, we used the mobile channel model Typical Urban 6 taps (TU6) [9] with 38,9 km/h receiver velocity relative to source (which corresponds to a maximum Doppler frequency = 24 Hz).

Before we started transmission simulation, we chose the QP pairs of the coding schemes according to the following procedure: We first assumed a total target bitrate was given which we chose around 600 Kbps. For the target video bitrate, we found the QP pairs of simulcast and MVC encoded videos and label them as Sim and MVC. This part corresponds to *equal bitrate* case. Then, we found QP pairs of MVC encoded videos which have the same PSNR with simulcast and label them as MVC2. This part corresponds to *equal quality* case.

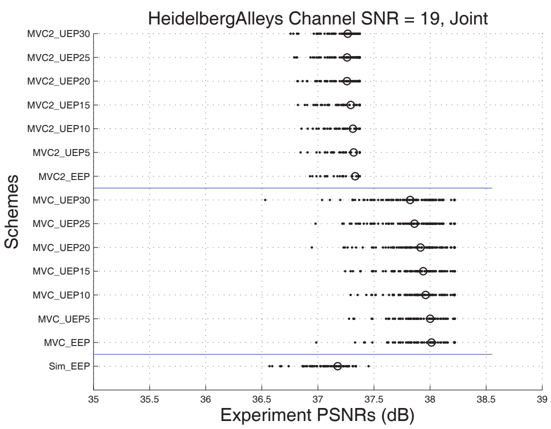
We conducted experiments on the stereo test sequences *Heidelberg Alleys* and *Rhine Valley Moving*. In addition to above mentioned characteristics that model the channel, two different SNR values (17 and 19 dB) are used in the generation of 100 different error patterns. Using these 100 different channel conditions we seek to find the most robust coding method and protection structure for the given DVB-H channel condition. We tested three different coding methods Sim, MVC and MVC2 all of which are explained above previously. Finally, we tested seven protection structures one of which uses equal protection rate for both views and the rest uses unequal protection rates favoring left view with varying percentages.

In the experiments, we evaluated 3D video quality by the objective joint PSNR metric which is calculated as $PSNR_j = 10 \cdot \log_{10}\left(\frac{255^2}{(MSE_l+MSE_r)/2}\right)$ where MSE_l and MSE_r correspond to mean square error between original and distorted left and right sequences respectively.

For the evaluation of the results, we provide the joint PSNR re-



(a) SNR 17 dB



(b) SNR 19 dB

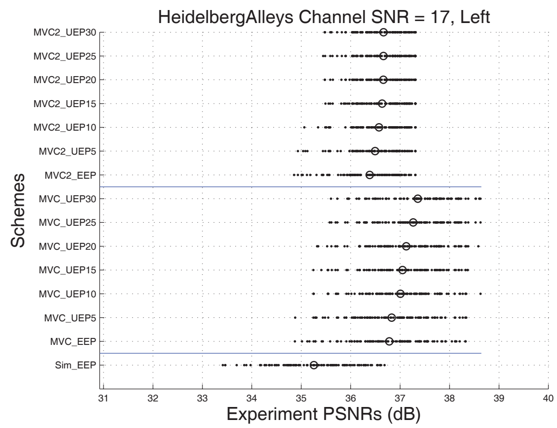
Fig. 2. Joint PSNR distributions of the experiments for Heidelberg Alleys

sults of all 100 experiments conducted for each method as a distribution of resulted video quality. Fixing the channel condition (channel SNR) for any sequence, all the other parameters tested in the study are provided within a plot.Figures 2, 3 and 4 provide the experimental results of both channel SNRs for each content corresponding to Heidelberg Alleys and Rhine Valley Moving.

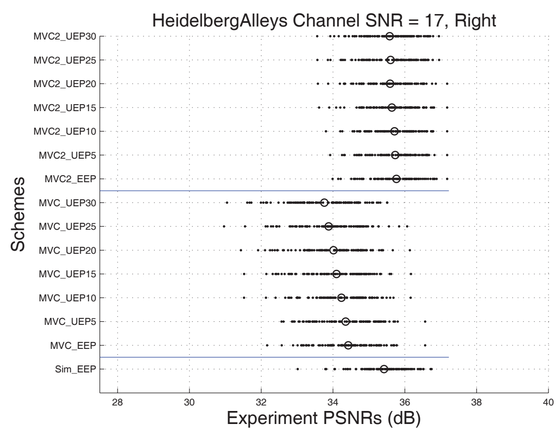
From figures we observe that transmission scenarios using interview dependency in encoding is always superior to simulcast. In Figure 3(a), which provides us the PSNR distributions of the experiments for left view at channel SNR 17 dB, we observe that MVC performs better than MVC2. Therefore, even though there are more packet losses for MVC, the received packets result in a higher PSNR. For left view, the increased error protection rate of MVC2 does not account for the PSNR increase achieved in higher bitrate coding. Here, it is significant to note that we keep the total burst duration constant and realize equal quality encoding in this case by reducing the quality of left view and keeping right view quality constant in comparison to MVC. Extra bandwidth is used for increasing the protection rates of both left and right view bursts.

Loss rates for both left and right views are higher in case of MVC. However the increased loss rate of MVC affects left and right view differently. Without transmission errors left view PSNR in MVC is higher than in MVC2 and even there are more losses the received packets result in a higher PSNR due high quality encoding and frame copy concealment. The video quality of right view is the same for both MVC and MVC2 without transmission errors. Therefore the increased loss rate requires a drop in PSNR. Furthermore, in addition to increased loss rate, right view suffers from increased error propagation in prediction since left view loss rate is increased. This results in a decreased joint PSNR for MVC coding as can be seen in Figure 2(a), making MVC2 more robust for transmission under channel SNR 17 dB.

When we consider the results for channel SNR 19 in Figure 2(b), we observe that MVC has a much better performance than the other too. Channel SNR 19 has much lower loss rates and due to the used error protection left view loss rates for both MVC and MVC2 is near zero. Loss rates of right view is similarly around zero for MVC2 but considerably higher for MVC. Still this increased loss rate is insignificant due to the superior performance of left view for MVC and the joint PSNR results agree with this.



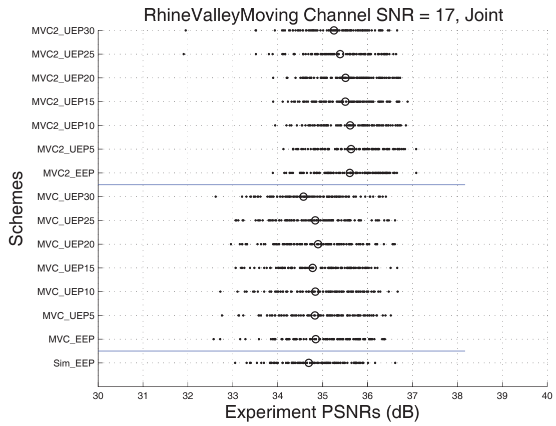
(a) Left View



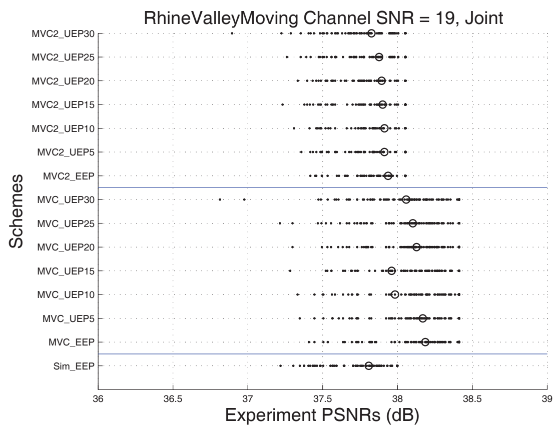
(b) Right View

Fig. 3. Left and Right View PSNR distributions of the experiments for Heidelberg Alleys

For the comparison of error protection strategies, we observe that unequal protection methods results in a PSNR increase for left



(a) SNR 17 dB



(b) SNR 19 dB

Fig. 4. Joint PSNR distributions of the experiments for Rhine Valley Moving

view and decrease for right view. However for joint PSNR results there is no significant improvement and equal error protection strategy provides the best performance. This is mainly due to prediction structure and layering used in the experiments. The dependency of right view on left view is valid for only the intra-coded frames of left view. However left view burst contains also the frames that are not reference for right view. Therefore increasing left view protection is not very efficient this way. To improve the unequal methods performance, the dependency of right view on left may be increased by using a prediction structure with more references or the frames maybe layered such that left view frames that are reference for right view are within a different burst.

4. CONCLUSION

In this paper, we have provided an extensive analysis of coding and transmission parameters for robust end-to-end transport of 3D video over DVB-H. We perform the analysis for different coding and error resilience schemes using two contents. Simulation results are evaluated for the combination of each coding method and error protection mode. The analysis of different coding methods shows that MVC outperforms Simulcast most of the cases. Another important observation about MVC modes is the importance of allocation of FEC rate and payload for a given bitrate. When the channel SNR is greater than 17 dB, MVC (coded with higher quality) is better than MVC2 (formed using more FEC). However, MVC2 has higher PSNR values when the packet loss rate increases. Analyzing different error protection methods shows that EEP has better results than other FEC modes. In order to exploit unequal strategies more effectively, other prediction structures with increased inter view dependency and further layering maybe considered. As future work, we mainly plan to deal with transmission of hierarchically encoded content, new UEP schemes with different layering structures and carry out an extensive study about finding optimized coding and transmission parameters.

5. ACKNOWLEDGMENT

MOBILE3DTV project has received funding from the EC's ICT programme in the context of the 7th Framework Programme under grant agreement n 216503. M. Oguz Bici is partially supported by The Scientific and Technological Research Council of Turkey (TUBITAK). The test videos were obtained from http://www.3dtv.at website. We would like to thank Anil Aksay for his help during the set up of the system.

6. REFERENCES

[1] G. Faria, JA Henriksson, E. Stare, and P. Talmola, "DVB-H: Digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.

[2] M.M. Hannuksela, V.K.M. Vadakital, and S. Jumisko-Pyykko, "Comparison of error protection methods for audio-video broadcast over DVB-H," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[3] D. Gómez-Barquero, D. Gozálvez, and N. Cardona, "Application layer fec for mobile tv delivery in ip datacast over dvb-h systems," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 396–406, 2009.

[4] A. Aksay, M. O. Bici, D. Bugdayci, A. Tikanmaki, A. Gotchev, and G. B. Akar, "A Study on the Effect of MPE-FEC for 3D Video Broadcasting over DVB-H," in *MobiMedia'09*, 2009.

[5] S. Cho, N. Hur, J. Kim, K. Yun, and S.I. Lee, "Carriage of 3D audio-visual services by T-DMB," *Electronics and Telecommunications Research Institute, Republic of Korea, in Proc ICME*, 2006.

[6] "ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC, Advanced Video Coding for Generic Audiovisual Services," March 2009.

[7] A. Vetro et. al., "Joint Draft 8.0 on Multiview Video Coding," *Joint Video Team, Doc. JVT-AB204*, 2008.

[8] "ETSI, Digital Video Broadcasting (DVB): DVB-H implementation guidelines," 2009, TR 102 377 V1.3.1.

[9] M. Oksanen, A. Tikanmaki, A. Gotchev, and I. Defee, "Delivery of 3D Video over DVB-H: Building the Channel," in *NEM-Summit'08*, 2008.