

JOINT SOURCE AND CHANNEL CODING FOR H.264 COMPLIANT STEREOSCOPIC VIDEO TRANSMISSION

P.Y. Yip, J.A. Malcolm
School of Computer Science
University of Hertfordshire
AL10 9AB United Kingdom

email: {p.y.a.yip, j.a.malcolm}@herts.ac.uk

W.A.C. Fernando, K.K. Loo, H. Kodikara Arachchi
School of Engineering and Design
Brunel University, UB8 3PH United Kingdom
email: {anil.fernando, jonathan.loo, hemantha.kodikara-
arachchi}@brunel.ac.uk

Abstract

Stereoscopic video coding research has received considerable interest over the past decade as many 3D displays have been developed. Unfortunately, the vast amount of multimedia content needed to transmit or store a stereo image pair or video sequence has hindered its use in commercial applications. As H.264 offers significantly enhanced compression and a “network-friendly” feature, we have used a H.264 compliant stereoscopic video codec [9] to compress stereo video. The data partitioning (DP) mode in NAL unit of the H.264 codec is exploited for the use of joint source and channel coding (JSCC) taking channel qualities and reliabilities into account. In this paper, we propose a framework of using unequal error protection (UEP) based JSCC scheme on the H.264 compliant stereoscopic video transmission for additive white Gaussian noise (AWGN) channel. Different levels of error protection are assigned to different partitions based on their decoding importance. Performance comparisons are made against equal error protection (EEP) schemes. Results from the simulation show that using UEP schemes, the overall quality of the decoded main and auxiliary video sequence were clearly improved in comparison with the EEP scheme at good SNR but EEP schemes outperformed UEP schemes at low SNR values.

Keywords: Stereoscopic Video Transmission, H.264, JSCC.

1. Introduction

Stereoscopic video coding research has received a considerable interest over the past decade as it intends to provide a reality of vision. Various stereoscopic displays have been developed to present a more vivid and depth representation of a scene so users can experience stereo vision [1]. However, to acquire a 3D visual communication technology several other supporting technologies such as 3D image representation, handling, and compression are required for ultimate commercial applications.

The human visual system has a remarkable ability to perceive three dimensional (3D) depth. This phenomenon is primarily related to projecting two slightly different images caused by binocular disparity on the retinas of the eyes. Stereo vision can be created by acquiring two pictures from the same scene at two horizontally separated positions where the left frame is presented to the left eye and the right frame to the

right eye. The human brain can process the differences between these to yield the overall 3D view. Although stereoscopic video is attractive, the amount of video data needed to be transmitted or stored without exploiting inherent redundancy is twice of a monocular image representing the same scene because every 3D image is represented by two 2D image frames forming a stereo image pair. As the two images are projections of the same scene from two nearby points, some considerable amount of redundancy exists between them. Therefore, by properly exploiting this redundancy, the two streams can be compressed and transmitted over a single monocular channel using the same bandwidth without unduly sacrificing the quality of the perceived stereoscopic image. In order to enable reliable and robust transmission of stereoscopic sequence over AWGN channels which are error prone in nature, good video compression techniques and error robust transmission methods are required.

In this paper, we will be using the existing data partitioning (DP) scheme in H.264 to evaluate possible joint source and channel coding (JSCC) for stereoscopic video applications. The rest of the paper is structured as follows. Section 2 discusses related work in stereoscopic compression and transmission. Our proposed H.264 compliant stereoscopic video encoding method [9] will also be presented in Section II. Section 3 provides a brief overview of the H.264 NAL concept and data partitioning. The proposed H.264 data partitioning approach used on the H.264 compliant stereoscopic video sequence will also be discussed. Section 4 describes the simulation model. Simulated results and discussion are presented in section 5. Finally, Section 6 concludes this paper.

2. Stereoscopic Video

2.1. Stereoscopic Video Compression

In order to provide an efficient and reliable framework for transmitting stereoscopic video sequences, two important factors affecting the overall quality of the reconstructed video need to be considered. First, a good video compression technique is required to reduce the amount redundancy that exists between the main and auxiliary stream before transmission. Besides, we also need to have an error robust scheme which will protect the compressed streams during transmission to reduce errors. Over the years, numerous compression schemes based on intensity or feature have been

devised for stereoscopic video. The first technique proposed by Perkins in [2], is the code sum and difference of the two images in stereo pair. Dinstien *et. al.* then proposed a 3D discrete cosine transform (DCT) coding of stereo image in [3]. This method is equivalent to the sum-difference coding in the transform domain. However, in both of this earlier approaches, the global translation and correlation techniques assumes that objects within the same scene have same disparity values therefore did not provide good overall performance as disparity values increases. Nevertheless, these approaches are not principally efficient as objects within the same scene tend to have different disparity values. Lukacs then constituted in [4] the concept of disparity compensation. In the proposed method, the correspondence between similar areas in a stereo pair using binocular disparity information is established. This method is used to predict the rest of the views from an independently coded view which leads to other disparity compensation based methods. Perkins then formalised that disparity compensation based coding as a conditional coding approach which would be optimal for lossless coding and suboptimal for lossy coding. Puri *et al.* [5] then presented results of MPEG-2 compatible coding. In this method, one view is coded as base layer and another view is coded within enhancement layer of the temporal scalability model of the MPEG-2 standard. Thanapirom *et al.* [6] has proposed a stereoscopic CODEC based on zerotree entropy encoding (ZTE) where this CODEC outperforms MPEG-2 based stereoscopic CODEC.

2.2. Stereoscopic Video Transmission

In terms of stereoscopic transmission, there had not been much literature in this area. Johanson [7] proposed a method of transmitting stereoscopic video over the Internet by extending the transport protocol. Mowafi *et. al.* then proposed another method for stereoscopic transmission over the internet by using a simple method for real time transmission of stereo images over the National Computational Science Alliance Access Grid in [8]. The method is compatible with existing Access Grid software which does not require additional bandwidth or major modification to the coding process.

In this paper, we are proposing to transmit stereoscopic video over AWGN channels with the use of the H.264 technology and JSCC schemes. With the aid of H.264's significant enhanced compression, a H.264 compliant stereoscopic video sequence video will be produced in DP mode and transmitted over an AWGN channel where different levels of error protection based on JSCC scheme will be applied to the partitioned coded video data.

The proposed encoding method applies the H.264 technology to stereoscopic video coding. The main and auxiliary are first fed to the multiplexer to create one combined stream. The output is sent to the modified H.264 based encoder, where the correlation of the auxiliary stream is exploited mainly using motion and disparity correlation. The encoded bit stream in DP mode is fed to the modified H.264 based stereoscopic decoder and to the demultiplexer to separate

the stream back to main and auxiliary streams. As shown in Figure 1, Each macroblock (MB) of a P-picture is predicted by either disparity compensated prediction [9] of the combined stream or by motion compensated prediction from the last but one frame (whichever is smaller). The prediction error (residual) frame of is than also coded in the usual way.

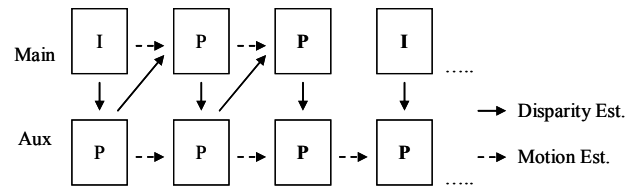


Figure 1 Main and Auxiliary Stream Coding Sequence

3. Network Abstraction Layer (NAL) & Data Partitioning (DP)

There are two distinct layers in the H.264/AVC standard i.e. Video Coding Layer (VCL) and Network Abstraction Layer (NAL). The adaptation of the VCL encoded video data output to variety of communication channels is through the NAL. Basically, the NAL facilitates the delivery of the VCL data to underlying transport layers. Each NAL unit (NALU) makes up a packet where it contains some number of bytes including a header and a payload. The header specifies the type of each NALU and the payload contains related data. In VCL, picture frames are divided into MBs, size of 16x16. An integer number of MBs are further grouped to form a slice. In H.264, each slice can be encoded to fit the size of one (or more) separate NALU that can be independently decodable where texture data can be imported from outside the slice boundaries during motion compensation.

The NAL operates in two modes: Non-partitioned (NP) or data partitioned (DP) mode. In NP mode, a single slice packet contains all the bits belonging to a slice. In DP mode, each packet carries certain syntactical elements of coded video streams. There are 20 video syntactical elements defined in [10]. DP mode is an efficient and important way to make a video bit stream more robust to errors. DP approach allows more important parts of the video data e.g. headers, motion vectors and addressing data be placed ahead of the non-important data which can potentially improve video quality transmission over communication channels. When DP mode is enabled, every slice is divided into 3 separate partitions where each partition can be accommodated in a particular NALU.

Here, we have used the DP approach in the NAL unit with joint source and channel coding schemes (JSCC) [12] on H.264 compliant stereoscopic sequence to cope with varying channel qualities. Our H.264 compliant stereoscopic video is divided into 3 partitions based on its importance in the decoding process. Partition A contains the most important information which includes disparity compensated prediction vectors of both channels that is needed to decode picture frames thus needs the highest level of error protection. Partition C requires

least protection since information in partition A and B can be used to decode data in partition C.

4. Simulation Model

The following simulations investigate the impact of channel errors to the overall quality of the reconstructed H.264 compliant stereoscopic sequence when different protection levels were assigned to partition C over a constant bit rate AWGN channel of 3Mbps. Based on literature in H.264 standard, information in partition B is less important than A and information in partition C is the least important. However, assigning an appropriate protection level to partition C partition will maximise the overall reconstructed video quality.

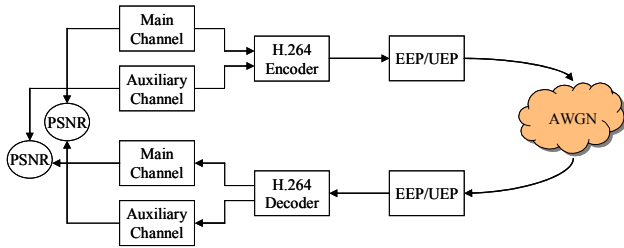


Figure 2 Simulation Model

Figure 2 shows the overall simulation model. Stereoscopic video sequences are compressed by H.264 encoder in DP mode to give H.264 compliant stereoscopic coded video stream. The coded video stream is protected by a rate compatible punctured turbo codes (RCPT) depending on the UEP and EEP scheme. The RCPT code with transfer function (1, 13/15) was used as it gave the best performance through our simulation. Here we proposed two UEP and two EEP channel coding schemes as shown in Table 1. Video quality was compared between the four schemes proposed for both main and auxiliary channels.

Coding Scheme	Channel Code Rate		
	Partition A	Partition B	Partition C
UEP1	Rate 1/3	Rate 1/2	Rate 1/2
UEP2	Rate 1/3	Rate 1/2	Rate 8/9
EEP1	Rate 1/3	Rate 1/3	Rate 1/3
EEP2	Rate 1/2	Rate 1/2	Rate 1/2

Table 1 Summary of Channel Coding Schemes

First, we look at video quality when partition C mainly containing motion and disparity compensated error residuals of both the main and auxiliary channels was protected with different channel code rates: a strong RCPT code rate 1/2 and a weak RCPT code rate 8/9 in UEP2. Similar settings were used for EEP1 and EEP2 schemes, except EEP1 simulation uses RCPT code rate 1/3 and EEP2 uses RCPT code rate 1/2. The selection of channel code rates used in these simulations adapts with ITU-T recommendations on 3G channel coding aspects [13].

5. Simulation Results and Discussion

The characteristics of the two different sequences (*Booksale* and *Crowd*) H.264-compliant stereoscopic used are in Table 2 were examined in the simulations but despite the different characteristics of these, the same conclusions can be drawn for each case. The simulation results obtained from the *Booksale* sequence for main and auxiliary channels using all 4 schemes are shown in Figure 3 and 4 respectively. The following subsections provide the analysis on the performance of UEP, EEP and comparison between UEP and EEP schemes.

Video Input Format	640*240, 160 frames
Frame Rate	30fps (<i>Booksale</i> , <i>Crowd</i>)
Coding Method	CABAC
I-frame Period	6 (IPPPPI), No B-frame used
Channel Model	AWGN
Channel Bitrate	3Mbps (<i>Booksale</i> , <i>Crowd</i>)

Table 2 Simulation Parameters

5.1. Performance of UEP Schemes

This section analyses the performance of 2 UEP schemes. Partition A and partition B is protected by rates 1/3 and 1/2 RCPT codes respectively in both schemes. The only difference is that partition C which mainly contains inter-coded elements like motion and disparity compensated error residuals of both the main and auxiliary channels is protected by RCPT code rate 1/2 in UEP1 and 8/9 in UEP2. Theoretically, partition C uses information from partition A and B to decode. In UEP1, partition C used the same code rate of partition B and it is observed that it is overprotected. In UEP2, partition C is protected with a weak RCPT code rate 8/9. From simulations of both the sequences, we could see that the PSNR values of the reconstructed video reduced considerably at lower SNR region. At SNR 2.5dB, PSNR using UEP1 scheme for both H.264 compliant stereoscopic sequences is approximately 27dB in main channel and 18dB in auxiliary channel. Both sequences using UEP2 scheme yield PSNR of approximately 25dB in main channel and 16dB in auxiliary channel at similar SNR. PSNR for both channels using UEP1 in both sequences improved rapidly and reaches its maximal values at 3.5dB. On the other hand, PSNR values of UEP2 increases rather slowly and maximized at higher SNR of 6.0dB for both sequences in both channels.

5.2. Performance of EEP Schemes

This section looks at the performance of two EEP schemes. All partitions in both EEP schemes are protected by the same RCPT code rates. EEP1 uses a RCPT code rate of 1/3 and EEP2 uses a code rate of 1/2. For each of the PSNR curve in both schemes, it is observed that the PSNR of the both video sequences reaches its maximal values at approximately 2.5dB and 4.5dB in both channels for channel rate 1/3 and 1/2 respectively. Maximal PSNR value of EEP2 is always higher than EEP1 allowing subjective quality to improve considerably. EEP2 produced higher PSNR in high SNR

region and EEP1 improves PSNR remarkably at low SNR region. As we are considering a constant bit rate (CBR) of 3Mbps, channel rates are fixed for both schemes. Thus, if more redundancy is added to compress H.264 compliant stereoscopic video data, the number of information bits will be reduced. With EEP2 using a RCPT code rate of 1/2, fewer error correction bits are required which explains why EEP2 yields a higher PSNR than EEP1.

5.3. Performance Comparison Between EEP and UEP

In earlier sub-sections, we have investigated the performance of EEP and UEP schemes separately. Here, we compare the performance between these schemes. From Figure 6 and 7, we observed that EEP1 improves quality at low SNR region and UEP2 gives maximal PSNR values at high SNR. In the main channel, UEP1 outperforms EEP1 at 3dB and EEP2 outperforms EEP1 at 3.6dB and UEP2 supersedes all schemes at 6dB. Similar conditions were observed in the auxiliary channel but at different SNR values. Therefore, a channel adaptive coding (CAC) approach can be adopted to switch between schemes to reduce the amount of redundant bits required as SNR increase to give an overall increase in the H.264 compliant stereoscopic video quality output.

6. Conclusions

In this paper, we used the DP approach within the NAL on H.264-compliant stereoscopic video sequences. We have considered UEP and EEP channel coding schemes to protect the H.264-compliant stereoscopic video transmission over a constant bit rate AWGN channel. The UEP schemes used were designed based on the DP feature introduced in H.264/AVC standard. It is observed from the simulations that if motion and disparity estimation error residuals of the main and auxiliary channels in partition C are protected properly, the quality of the transmitted video sequences will improve. The comparison between UEP and EEP approach shows that EEP outperformed UEP at low SNR and the performance slightly decreases at higher SNR region compared to UEP schemes. The adoption of the three partition approach of H.264/AVC here in this paper was originally used to exploit the potential advantage of UEP principle, however this approach did not promise good video quality at low SNR however adopting the use of channel adaptive coding (CAC) method to switch between these schemes will significantly improve the overall video quality for both channels over a constant bit rate transmission channel.

References

[1] W. Woo, "Rate-Distortion based Dependent Coding for Stereo Image/Video: Disparity Estimation and Dependent Bit Allocation", *PhD thesis*, USC, LA, CA, USA, 1998.
 [2] M.G. Perkins, "Data Compression of Stereo pairs", *IEEE Transaction on Communications*, Vol.40, No.4, pp. 684-696, 1992.
 [3] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov, and A. Henik, "On Stereo Image Coding," *Proceedings of the 9th International Conference on Pattern Recognition*, Rome, Italy, pp. 357-359, 1988.

[4] M.E. Lukacs, "Predictive Coding of Multi-viewpoint Image Sets", *Proceedings of the IEEE-IECEJ-ASJ International Conference on Acoustics, Speech, and Signal Processing*, Japan, pp. 521-524, 1986.
 [5] A. Puri, R. V. Kollarits B.G. Haskell, "Stereoscopic Video Compression Using Temporal Scalability", *Proceedings of SPIE on Visual Communications and Image Processing*, Taiwan, Vol.2501, May, pp. 745-756, 1995.
 [6] S. Thanapirom, W.A.C. Fernando, E.A. Edirisinghe, "A zerotree stereo video encoder," *Proceedings of the 2003 International Symposium on Circuits and Systems*, Vol.2, 2003.
 [7] M. Johanson, "Stereoscopic video transmission over the Internet", *Proceedings of the 2nd IEEE Workshop on Internet Applications*, pp. 12-19, 2001.
 [8] M. Mowafi, J. Nan, T.P. Caudell, S.Wei, M.Wu, "Real-time transmission of stereo images over the access grid", *Proceedings of the IEEE International Symposium on Circuits and Systems*. Vol.5, pp. 26-29 2002.
 [9] A.B.B.Adikari, W.A.C. Fernando, H. Kodikara Arachchi, K.K.Loo, "H.264 Compliant Stereoscopic Video CODEC for Low Bit Rate Applications", to appear in *Proceedings of IEEE Canada Conference on Electrical and Computer Engineering*, 2005.
 [10] S. Wenger, T. Stockhammer, "An Overview on the H.26L NAL Concept", JVT-B028, JVT of ISO/IEC MPEG & ITU-T VCEG (SG16/Q6), Second Meeting, Geneva, 2002.
 [11] G. Sullivan, T. Wiegand, and T. Stockhammer, "Using the Draft H.26L Video Coding Standard for Mobile Applications", *IEEE International Conference on Image Processing*, Thessaloniki, Greece, 2001
 [12] J.L.Massey, "Joint source and channel coding", in J. K. Skwirzynski, editor, *Communication Systems and Random Process Theory*, The Netherlands, Sijthoff and Nordoff, pp. 279-293, 1978.
 [13] O.F. Acikel and W.E. Ryan "Punctured Turbo-Codes for BPSK/QPSK Channels", *IEEE Transactions on Communications*, Vol.47, No.9, pp. 1315-1323, 1999.

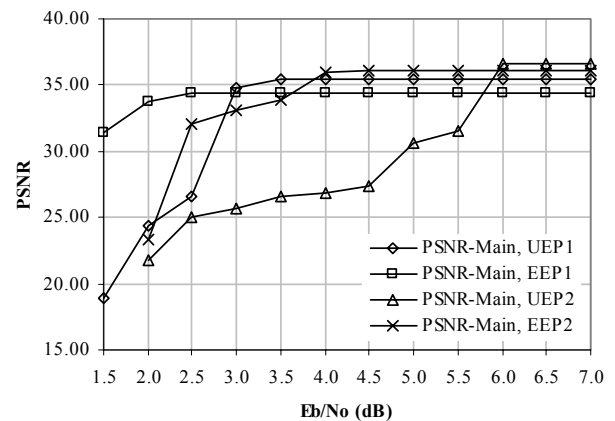


Figure 3 Main channel PSNR-Y performance of "Booksale" sequence at 30fps with channel bit rate of 3Mbps

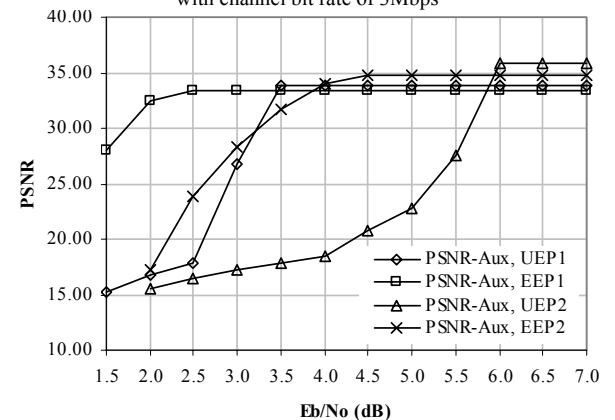


Figure 4 Auxiliary channel PSNR-Y performance of "Booksale" sequence at 30fps with channel bit rate of 3Mbps