ERROR RESILIENT LAYERED STEREOSCOPIC VIDEO STREAMING

A. Serdar Tan[1]*, Anil Aksay*[2]*, Cagdas Bilen*[2]*, Gozde Bozdagi Akar*[2] *and Erdal Arikan*[1]

[1]Electrical and Electronics Engineering Department, Bilkent University, Ankara, Turkey
[2]Electrical and Electronics Engineering Department, Middle East Technical University, Ankara, Turkey

ABSTRACT

In this paper, error resilient stereoscopic video streaming problem is addressed. Two different Forward Error Correction (FEC) codes namely Systematic LT and RS codes are utilized to protect the stereoscopic video data against transmission errors. Initially, the stereoscopic video is categorized in 3 layers with different priorities. Then, a packetization scheme is used to increase the efficiency of error protection. A comparative analysis of RS and LT codes are provided via simulations to observe the optimum packetization and UEP strategies.

Index Terms— Forward error correction, video coding, stereo vision

1. INTRODUCTION

Stereoscopic video transmission has gained considerable interest in the past few years due to the increase in research and advances on 3-D vision. Stereoscopic video is formed by the simultaneous capture of two video sequences corresponding to left and right views of human visual system. The dependency of the left and right views can be used to implement an efficient stereoscopic video codec. Once coded, in order to transmit it over error prone channels, error robust transmission methods are required.

Common error correction approaches for reliable transmission of monoscopic video over packet networks utilize retransmissions as in [1] or FEC methods as in [2], [3] and [4]. Retransmission method brings large latency due to feedback messages that inform the sender about the reliable reception of data. However, large latency is unacceptable for video streaming applications. LT codes are a novel retransmission-free and low-complexity FEC method introduced in [5]. LT codes have gained attention in the video streaming area in recent years [6].

Even though FEC codes are studied in depth for monoscopic video, only a few studies exist for stereoscopic video [7]. In this paper, we use RS and LT codes to protect the stereoscopic video data against transmission errors. We define 3 layers for stereoscopic video to be used for unequal error protection (UEP). We also present a packetization scheme to increase the efficiency of error protection. A comparative analysis of RS and LT codes are provided via simulations to observe the optimum packetization and UEP strategies.

2. STEREOSCOPIC CODEC

In our experiments, multiview video codec based on H.264 [8] is used. This codec uses a modified Decoded Picture Buffer (DPB) to perform both motion and disparity compensation with reduced complexity. For stereoscopic videos, a special mode allows for monoscopic compatible streams, where standard H.264 decoders can decode only left frames and stereoscopic decoder can decode both left & right frames. In monoscopic compatible mode, left frames are predicted from left frames only, whereas right frames can be predicted from both left and right frames. Right frames are always predicted from previous frames, whereas some of the left frames are encoded without prediction (i.e. I-frames). Stereoscopic encoder and decoder structure is given in Fig. 1.

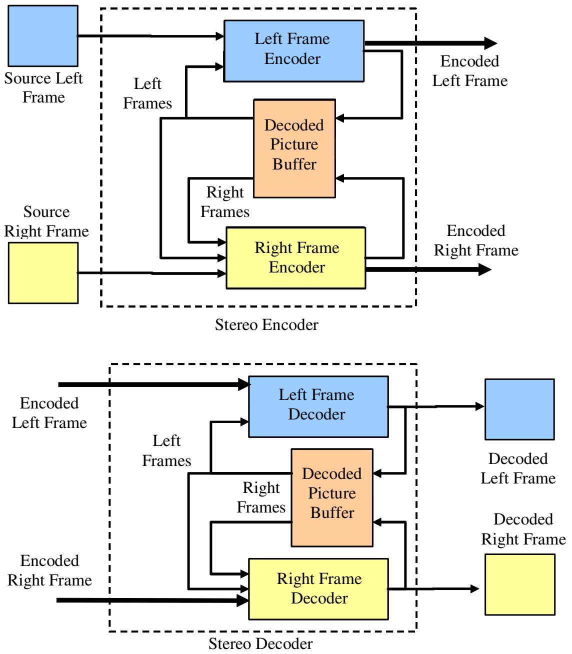

Fig. 1. Stereoscopic Encoder and Decoder Structure

Let $\mathbf{I}_L, \mathbf{P}_L$ and $\mathbf{P}_R$ denote the set of I-frames of left view, P-frames of left views and P-frames of right views respectively. The set of frames can be written in open form as $\mathbf{I}_L = \{\mathrm{I}_{L1}, \mathrm{I}_{L5}, ...\}$, $\mathbf{P}_L = \{\mathrm{P}_{L2}, \mathrm{P}_{L3}, ...\}$, $\mathbf{P}_R = \{\mathrm{P}_{R1}, \mathrm{P}_{R2}, ...\}$, where i denotes the frame number and L and R indicate the frames of left and right video. An illustration is given in Fig. 2 where GOP size is set to 4.

Although this coding scheme is not layered, frames are not equal in importance. We can classify the frames according to their contribution to the overall quality and use them as layers of the video. Since losing an I-frame causes large distortions due to motion / disparity compensation and error propagation, I-frames should be protected the most. Among P-frames, left frames are more important

---

This work was supported by the EC under contract FP6-511568 3DTV and in part by TUBITAK under contract BTT-Turkiye 105E065. A.S.T, A.A and C.B are supported in part by TUBITAK (Scientific and Technical Research Council of Turkey).

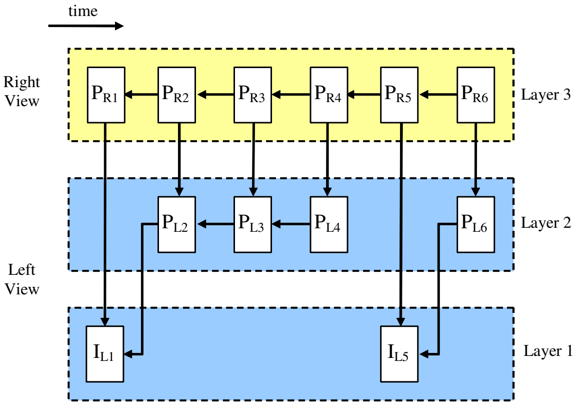

Fig. 2. Layers of stereoscopic video and referencing structure

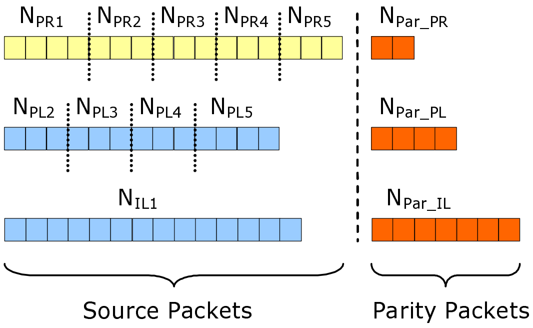

Fig. 3. UEP structure for 3 layers

since they can be encoded without the help of right frames. According to this prioritization of the frames, 3 layers are formed as shown in Fig. 2. UEP protection on the defined layers will be explained in Sec. 4. Note that this protection can be similarly used with any other layered stereoscopic codec.

3. FORWARD ERROR CORRECTION SCHEMES

3.1. Reed-Solomon (RS) Codes

The RS codes [9] are based on the arithmetic of finite fields in GF(2^m). A source block and an encoded block consists of *m* bits and a maximum number of $n = 2^{m-1}$ encoded blocks can be generated for *k* source blocks. The RS code constructs a polynomial whose coefficients are the *m*-bit source blocks. Then the polynomial is sampled at *n* points and these points are transmitted as the encoded blocks. At the decoder the arrival of any *k*-element subset of these *n* encoded blocks is enough to reconstruct the polynomial coefficients which are the source blocks. Thus, an RS encoder generates pre-defined number of encoded packets and decoder can reconstruct the original data from any *k*-element sized subset of encoding symbols. However, the number of the encoding packets is limited and the standard RS coding algorithm requires quadratic time which is not scalable.

3.2. Fountain Codes

A novel approach that provides retransmission free reliability, low latency and loss rate adaptability is fountain coding which is first mentioned in [10]. Fountain codes are well-suited for lossy packet networks. An ideal fountain encoder can generate potentially infinitely many encoding symbols from the original data consisting of *k* symbols in linear time and decoder can reconstruct the original data from any *any k*-element subset of received encoding packets in linear time.

3.2.1. Luby Transform (LT) Codes

LT codes [5] are the first practical realization of fountain codes. The input packets to the encoder are called *input symbols* and the encoded packets are called *output symbols*. The encoding and decoding of LT codes is detailed in [5]. LT codes are asymptotically optimal codes, namely the number of input symbols *k* has to be large for satisfactory performance. LT decoder can reconstruct all input symbols with high probability if $k(1+\varepsilon)$ output symbols are received, where ε is the ratio of overhead and tends to 0 as *k* increases. The original LT coder did not perform well for our case, thus we used a modified version of LT codes as described in the following section.

3.2.2. Systematic LT Codes

In the systematic coding schemes first the original then the parity data is transmitted. Original LT coder is non-systematic, namely the generated output symbols do not include input symbols. However, the access to original data is beneficial in some cases such as video transmission where 100% reliability is not obliged. In systematic case, even if the decoder can not recover any lost source symbols it still has some received parts of source data and error concealment techniques can be applied for the lost symbols.

Raptor codes [11] are another type of fountain codes which use the combination of an outer fixed-rate FEC code and an inner LT code. A systematization method for raptor codes has been recently proposed in [12]. In our work we applied a similar systematization procedure to original LT coding scheme. The resulting systematic LT codes yield better performance compared to original LT codes for video transmission applications.

4. UEP METHOD FOR GENERATING THE PARITY SYMBOLS

The sequence of generated frames is given as $[\mathrm{I}_{L1}, \mathrm{P}_{R1}, \mathrm{P}_{L2}, \mathrm{P}_{R2}, \mathrm{P}_{L3}, \mathrm{P}_{R3}, ..., \mathrm{I}_{L(N+1)}, \mathrm{P}_{R(N+1)}, \mathrm{P}_{L(N+2)}, \mathrm{P}_{R(N+2)}, ...]$, where N is the GOP size. The common way of protection against errors is to apply FEC to the fixed-sized NALU packets of each frame separately. In our work we treat each NALU packet as an input symbol. In the case of protecting each frame information individually we obtain small numbers of input symbols which is far from the optimal region of LT codes. Thus, in order to increase the number of input symbols we concatenate the consecutive frames of $\mathbf{P}_L$ and $\mathbf{P}_R$. Denoting the number of concatenated frames as N_{conc} and assuming that $N_{conc} = 5$ we obtain the frame groups as $\{[\mathrm{I}_{L1}], [\mathrm{P}_{R1}, \mathrm{P}_{R2}, \mathrm{P}_{R3}, \mathrm{P}_{R4}, \mathrm{P}_{R5}], [\mathrm{P}_{L2}, \mathrm{P}_{L3}, \mathrm{P}_{L4}, \mathrm{P}_{L5}], [\mathrm{P}_{R6}, \mathrm{P}_{R7}, \mathrm{P}_{R8}, \mathrm{P}_{R9}, \mathrm{P}_{R10}], [\mathrm{P}_{L6}, \mathrm{P}_{L7}, \mathrm{P}_{L8}, \mathrm{P}_{L9}, \mathrm{P}_{L10}], ...\}$. Error protection is applied to the concatenated packets of the corresponding grouped frames in square brackets.

In order to define the priorities of layers we use p_1, p_2, p_3 to represent the ratio of protection for layer-1,2 and 3 respectively. Thus, the ratio of the number of inserted parity packets to layers is calculated as $(p_1 : p_2 : p_3)$. In Fig. 3 we present an illustration of UEP structure based on the frame grouping method. Each square in Fig. 3 represents a fixed-sized NALU packet. $\mathrm{N}_{\mathrm{PR}i}$, $\mathrm{N}_{\mathrm{PL}i}$ and $\mathrm{N}_{\mathrm{IL}i}$ denote

p_e	R	N_{conc}	PSNR-RS	PSNR-LT	PSNR-No Protection
0.05	0.05	5	33.441	33.079	32.067
0.05	0.05	25	33.644	33.442	
0.05	0.10	5	34.442	33.786	
0.05	0.10	25	34.968	34.644	
0.10	0.10	5	31.406	31.191	30.061
0.10	0.10	25	31.671	31.463	
0.10	0.20	5	33.308	32.539	
0.10	0.20	25	34.060	33.684	
0.20	0.10	5	28.137	28.327	27.650
0.20	0.10	25	27.989	28.047	
0.20	0.20	5	29.499	29.394	
0.20	0.20	25	29.604	29.427	

Table 1. Average PSNR (dB) for different UEP ratios

the number of NALU packets in frames $\mathrm{P}_{Ri}, \mathrm{P}_{Li}$ and I_{Li} respectively. The parity packets are obtained by either LT or RS encoding applied to the corresponding grouped source packets. $\mathrm{N}_{\mathrm{Par\_PL}}$, $\mathrm{N}_{\mathrm{Par\_PR}}$ and $\mathrm{N}_{\mathrm{Par\_IL}}$ denote the number of inserted parity packets. Let R denote the fraction of inserted parity packets and R_i denote the fraction of inserted parity packets reserved for layer-i. Then, the channel protection is distributed to the layers such as: $R_i = R(p_i / \sum_j p_j)$. For example, the number of parity packets for layer-3 can be calculated as: $\mathrm{N}_{\mathrm{Par\_PR}} = R_3(\mathrm{N}_{\mathrm{IL1}} + \sum_i(\mathrm{N}_{\mathrm{PR}i} + \mathrm{N}_{\mathrm{PL}i}))$.

5. EXPERIMENTAL RESULTS

The proposed scheme for transmission of stereo H.264 /AVC streams is evaluated based on the ITU-VCEG loss patterns [13] and loss simulator [14]. As mentioned above systematic LT codes and RS codes are used based on their suitability for our case as explained in Sec. 3.2.2. The encoded packets are generated according to the UEP method given in Sec. 4. Since LT codes are probabilistic codes, loss simulation is repeated 25 times by changing the initial point of the loss pattern each time.

In our simulations we compared the performance of stereo video transmission with LT and RS codes. The channel protection results are also compared with the protection free case. All the channel protection we use is systematic. In case of unrecovered losses, stereoscopic video decoder performs an efficient error concealment algorithm for both block and frame losses using motion vector projection and boundary matching. The results are provided for stereoscopic video pair *Rena* (Camera 38, 39) (640 × 480, first 450 frames). I-frames are inserted every 25 frames. NALU packet size is fixed to 250 bytes. Video is encoded with 586 *Kbps* bitrate. We denote average packet loss probability as p_e.

The reconstruction quality measure is PSNR. PSNR value of a stereo-pair is calculated according to the following formula, where D_l and D_r represent the mean-squared error in the left and right frames [15]. Reconstruction quality of the video without any loss is 36.556 *dB*.

$$PSNR_{pair} = 10 \log_{10} \frac{255^2}{(D_l + D_r)/2}$$

In our simulations, we have set p_1 equal to 1 and varied p_2 and p_3 ratios. We keep p_1 constant with highest possible ratio, since layer 1 consists of most important packets. According to the UEP allocation explained in Sec. 4, we have calculated PSNR values for several UEP ratios. In Table 1, we have given the average PSNR values over different UEP ratios. In Fig. 4, comparison of RS and LT coding schemes with $R = 0.1$ but varying p_e is presented.

It can be seen that LT protection provides better results where channel protection rate is less than packet loss rate. This is due to the fact that systematic RS coding can only reconstruct all lost input symbols if at least k output symbols arrived whereas LT can still reconstruct some of the lost input symbols even though less than k output symbols arrived. When protection rate is greater or equal to packet loss rate, RS coding performs better due to the overhead of LT coding.

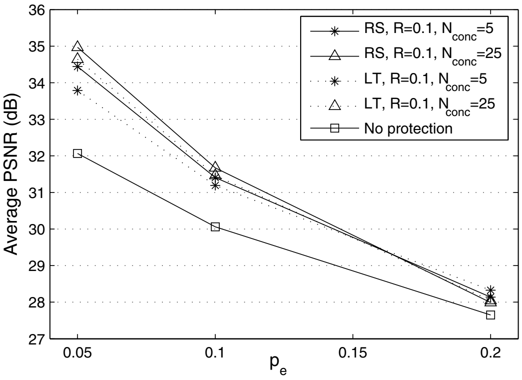

Fig. 4. Average PSNR (dB) for 10% protection and varying loss

In Fig. 4, we also provided the results where no channel protection is applied. As the average packet loss increases, the gap between the protected and unprotected case decreases. This shows the importance of estimating packet loss probability or using adaptive protection rate. Since LT codes are low-complexity codes and can provide potentially infinitely many parity packets, LT coding provides better real-time adaptation compared to RS coding.

Variation of PSNR values with different protection ratios are given in Fig. 5, 6, 7 and 8. In Fig. 5 and 6, channel protection is sufficient to protect most of the packets. In these cases, optimal layer protection ratio tends to protect layer 1 and 2 (left frames) instead of layer 3 (right frames). Even though this may seem like optimum UEP strategy favors monoscopic stream, this is not the case, since right frames are coded using left frames. Thus, an increase in the quality of left frames results in an increase in the quality of right frames indirectly. However, as seen from Fig. 7 and 8 if the packet loss increases beyond the capabilities of channel protection, then optimal layer protection ratio tends to protect only the most important layer (I-frame). This is due to the fact that losing an I-frame causes the highest quality distortion.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we provided the performance comparison of RS and LT codes in a packet loss environment for stereoscopic video streaming. We have defined a layered stereoscopic video structure and applied packetization and UEP method to these layers. The simulation results yield the optimum operation region of UEP for the defined distortion of stereo-pair. Results also show that in matching channel protection rates, RS coding performs better than LT coding with the complexity disadvantage of RS coding. However, LT coding provides an efficient solution for adaptive systems due to its low-complexity and capability of generating potentially limitless parity symbols.

Future studies will include insertion of additional layers and addition of real-time loss-rate adaptation which will lead to a more efficient real-time error resilient stereoscopic streaming system.

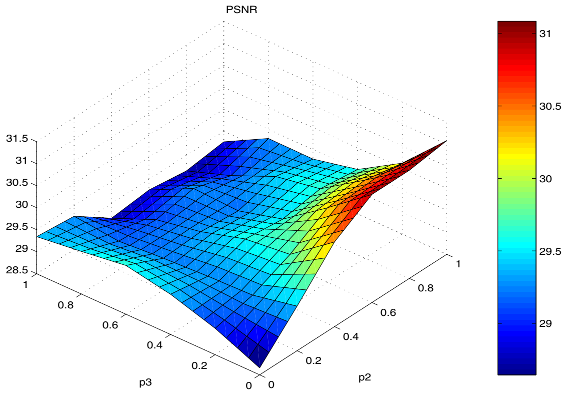

Fig. 5. RS coding, $p_e = 0.20$, $R = 0.20$, $N_{conc} = 5$

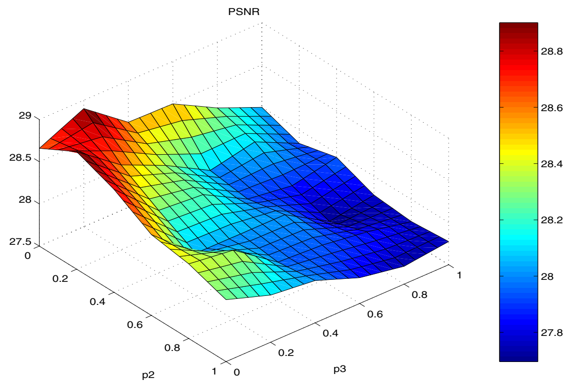

Fig. 7. RS coding, $p_e = 0.20$, $R = 0.10$, $N_{conc} = 5$

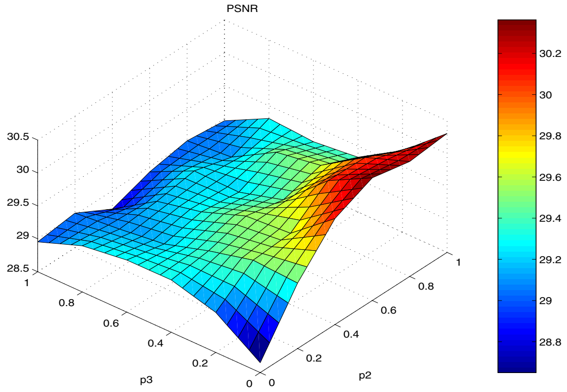

Fig. 6. LT coding, $p_e = 0.20$, $R = 0.20$, $N_{conc} = 5$

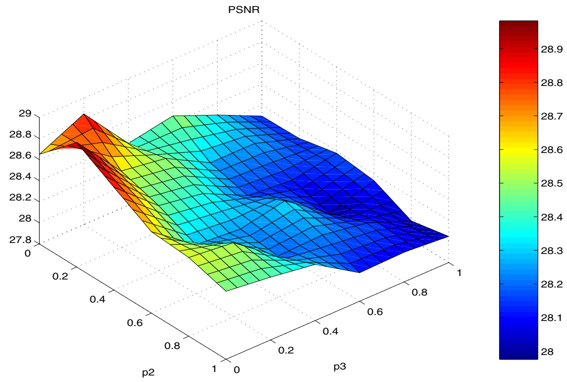

Fig. 8. LT coding, $p_e = 0.20$, $R = 0.10$, $N_{conc} = 5$

7. REFERENCES

[1] GJ Conklin, GS Greenbaum, KO Lillevold, AF Lippman, YA Reznik, R.N. Inc, and WA Seattle, "Video coding for streaming media delivery on the Internet," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 3, pp. 269–281, 2001.

[2] M. Link B. Girod, K. Stuhlmuller and U. Horn, "Packet loss resilient internet video streaming," in *Proc. SPIE Visual Commun. Image Processing*, 1999.

[3] Hua Cai, Bing Zeng, Guobin Shen, Zixiang Xiong, and Shipeng Li, "Error-resilient unequal error protection of fine granularity scalable video bitstreams," *EURASIP Journal on Applied Signal Processing*, vol. 2006, 2006.

[4] J.W. Pei, Y. Modestino, "H.263+ packet video over wireless IP networks using rate-compatible punctured turbo (rcpt) codes with joint source-channel coding," in *Proc. of the IEEE ICIP*, 2002.

[5] M. Luby, "LT codes," in *Proc. of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2002, pp. 271–282.

[6] J. Chakareski J.-P. Wagner and P. Frossard, "Streaming of scalable video from multiple servers using rateless codes," in *Proc. IEEE Conf. Image Proc. (ICME)*, Toronto, Canada, July 2006.

[7] P.Y. Malcolm J.A. Fernando W.A.C. Loo K.K. Arachchi H.K. Yip, "Joint source and channel coding for h.264 compliant stereoscopic video transmission," in *Canadian Conf. on Electrical and Computer Engineering*, May 2005.

[8] C. Bilen, A. Aksay, and G. Bozdagi Akar, "A multi-view video codec based on H.264," in *Proc. IEEE Conf. Image Proc. (ICIP)*, Oct. 8-11, Atlanta, USA, 2006.

[9] G. Solomon L.S. Reed, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, pp. 300–304, June 2001.

[10] M. Mitzenmacher J. Byers, M. Luby and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *Proceedings of ACM Sigcomm*, 1998.

[11] A. Shokrollahi, "Raptor codes," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, pp. 2551–2567, 2006.

[12] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor codes for reliable download delivery in wireless broadcast systems," in *Proc. of the IEEE CCNC*, 2006.

[13] S.Wenger, "Error patterns for internet experiments," in *VCEG Q15-I-16r1*, 2002.

[14] Y.-K. Wang Yi Guo, Houqiang Li, "SVC/AVC loss simulator donation," in *JVT-Q069*, 2005.

[15] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 898–903, Oct. 2002.