ROBUST SCALABLE VIDEO TRANSMISSION USING ADAPTIVE DOUBLE BINARY TURBO CODE

N. Ramzan and E. Izquierdo

Queen Mary, University of London, Mile End Road, London E1 4NS, UK

E-mail: {naeem.ramzan, ebroul.izquierdo}@elec.qmul.ac.uk

Keywords: Scalable video, double binary turbo code, Joint source and channel coding.

Abstract

In this paper an approach for joint source-channel coding in a spatio-temporal wavelet-based scalable coding environment is presented. The scalable quality layers are protected using a double binary turbo code for robust scalable transmission over wireless channels. In the proposed schema the double binary turbo code is used for bitstream adaptation with unequal error protection. An efficient method to estimate the minimum distance of the double binary turbo codes is also proposed. Experimental results show promising performance of the proposed model as compared to equal error protection at low signal to noise ratio.

1 Introduction

The design of techniques for robust video transmission over heterogeneous and unreliable channels is an area of active research across the world. This is mainly due to its commercial importance in many crucial applications such as video transmission and access over the Internet; multimedia broadcasting and video services over wireless channels. Also, the distortion at the receiver is a function of both lossy source coding technique and the errors introduced in the channel. So, it is quiet unclear how to best allocate bit budget between source and channel coding to get the optimal result at the user end.

In traditional video communications over heterogeneous channels, the video can be processed offline. Compression and storage is tailored to the targeted application according to the available band-width. However, keeping several different versions of encoded video in the servers is not a storage efficient solution. Furthermore, video delivery over heterogeneous and wireless channels meets challenges that are quite different from conventional transmission solutions. Here, dimensions like scalability, robustness and error resilience need to be re-defined to allow for variability according to individual user preferences and terminals.

Scalable coding promises to solve this problem by "encoding once and decoding many". The goal is to develop novel and efficient coding algorithms enabling content organization in a hierarchical manner to allow coding and interactivity at several granularity levels. Basically, scalable coded bit streams can adapt to the application requirements. Thus, problems inherent to channel capacity variations and better quality of services can be overcome. In wireless applications an additional critical problem relates to the very high bit error rate (BER) or frame error rate (FER). The challenge is to design efficient source and channel coding technology, so that the video content can be transmitted and used at different resolutions or quality levels.

Due to Shannon's theorem of separability [10], source and channel coding have been considered and optimized independently. But the Shannon's theorem is based on the assumption that the source and channel codes are of arbitrary large lengths. This assumption does not hold in practical situation due to limitations on computational power and processing delays. So joint source-channel coding (JSCC) has become an active area of research.

In this work, a JSCC technique [1,9], is described. It combines the scalable video coding (SVC) framework reported in [5] with a forward error correction method (FEC) to achieve jointly source-channel scalable coding. Turbo codes (TCs) are one of the most prominent FEC techniques having received great attention since their introduction in 1993 [2]. This is mainly due to their excellent performance at low bit error rates, reasonable complexity, and versatility for encoding blocks with various sizes and rates. The JSCC technique proposed in this paper is based on SVC and double binary turbo coding [4].

The remaining paper is structured as follows: in section 2, double binary TC, the new iterative approach to find minimum Hamming distance (MHD) of TC and scalable video coding is described. The proposed technique for JSCC is explained in section 3. Section 4 presents some selected results. The paper closes with related discussions and conclusions in section 5.

2 Problem formulation

A SVC codec encodes the input video that is stored on the server. The video is adapted with respect to the user requirements, protected by the FEC against channel errors and modulated before transmission. Additive white Gaussian noise (AWGN) and Rayleigh fading channels are considered.

At the receiver side, the received video is demodulated, decoded by FEC, adapted if needed and finally decoded.

The overall distortion D_S caused by both the source coding and channel coding. First we will discuss which factors are important to optimize the performance for source and channel coding and then try to optimize in a combined way. The overall picture of the system is shown in Figure 1.

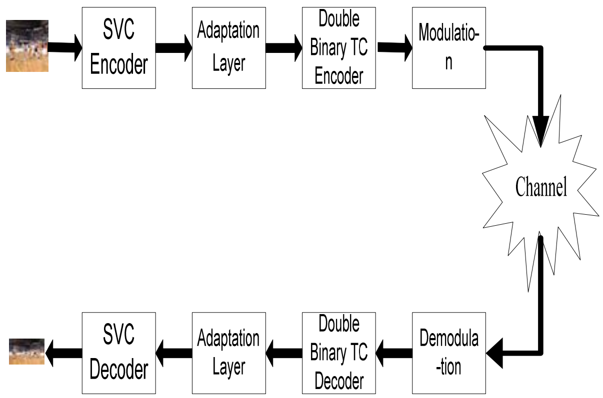

Figure 1: Block diagram of scalable video transmission system.

Double binary TCs are used to adapt the scalable bitstream and its performance is evaluated by measuring the minimum distance.

2.1 Double binary turbo codes

Double binary TCs were introduced in the domain of TCs by Berrou et al.[4]. These codes consist of two binary RSC encoders of rate $2/3$ and an interleaver of length *k*. Each binary RSC encoder encodes pair of data bits and produces one redundancy bit, so desired rate $1/2$ is the natural rate of the double binary TC. In this article, we consider the 8-state double binary TC with generators in octal notation are (15,13) for Y_1. This schema has been adapted by the European Telecommunications Standards Institute (ETSI) for Digital Video Broadcasting with Return Channel via Satellite (DVB-RCS) and Digital Video Broadcasting with Return Channel via Terrestrial (DVB-RCT), as shown in the Figure 2. The tail-biting technique [4] is used to convert the convolutional code to block code that allows any state of the encoder as the initial state. So there is no need to tail bits to derive the encoders to the all-zero state. The used turbo-decoder is composed of two Maximum A Posteriori (MAP) decoders or two Max. Log MAP [8], one for each stream produced by the singular RSC block as shown in Figure 2. The interleaver design [3] is a critical issue since the performance of the TC depends on how well the information bits are scattered by the interleaver. The used interleaver is based on the permutation equation $i = \Pi(j)$ where $j = 0 \ldots N-1$ and N be the number of data couples. The permutation is done on two levels; the first one inside the couple as $(A_j, B_j) = (B_j, A_j)$ if $j \bmod .2 = 0$ and second one between couples i.e.

$i = (P_0 \times j + P) \bmod .N$, with

-If $j \bmod .N = 0$, then $P = 0$.

-If $j \bmod .N = 1$, then $P = N/2 + P_1$.

-If $j \bmod .N = 2$, then $P = P_2$.

-If $j \bmod .N = 3$, then $P = N/2 + P_3$.

Where the values of P_0, P_1, P_2, and P_3 depend on N.

At low error rates or high signal to noise ratio, the performance of the classical TC fluctuates due to the “error floor” [3]. The higher minimum distance [4] can reduce the error floor effect at low error rates. Double binary TCs normally has better performance than classical TC due to larger minimum distance. The minimum distance of TC depends on the interleaver design or how well it shuffles the information bits.

To find the best parameters at certain packet size, we have to evaluate the performance of double binary TC with each set of permutation parameters at given lengths. The simplest method to see the performance of TC at low error rates is by MHD. Five different MHD methods have been introduced in the past [3]. To get better performance for the double binary TC for video communication, we have to evaluate the performance of double binary TC so the new iterative approach is developed to measure the MHD is proposed. This new approach showed good performance with limited complexity.

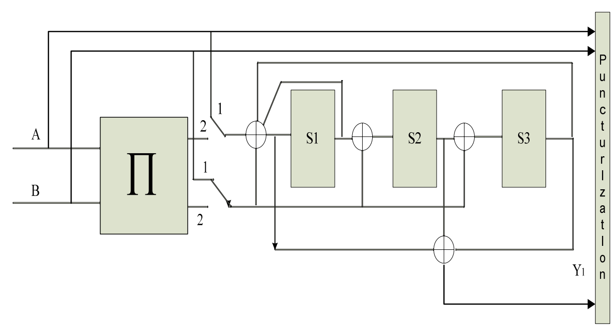

Figure 2: Double binary turbo encoder

2.2 New iterative approach to measure minimum distance of m-Binary TCs

An information frame is denoted by $D = (d_1 \ldots d_i \ldots d_k)$ where d_i is the vector of m-binary data applied at the input of the turbo encoder at time i, and $d_i = (d_{i,1} \ldots d_{i,j} \ldots d_{i,m})$. The output of the turbo encoder is $C = (c_1 \ldots c_i \ldots c_n)$. Here, c_i is a vector of length $m+n$ bits, that is $c_i = (c_{i,1} \ldots c_{i,j} \ldots c_{i,m+n})$, where $c_{i,j}$ is the systematic bit if $j \le m$ and parity bit for $j > m$. This codeword is mapped by the QPSK modulator into the transmitted vector $x = (x_1 \ldots x_i \ldots x_n)$ that is also constituted of n vectors. x_i has length $m+n$ and $x_i = (x_{i,1}...x_{i,j}...x_{i,m+n})$, where $x_{i,j} = 2c_{i,j} - 1$ for $i = 1...m+n$. After transmission over the Gaussian channel, the received vector is

$$R = (r_1...r_i...r_n) \quad (1)$$

with

$$r_i = (r_{i,1}...r_{i,j}...r_{i,m+n}) \quad (2)$$

Assume that an Additive White Gaussian Noise (AWGN) channel is used and the all zero codeword, i.e., $r_q = -1$ for all q, is received. The proposed method estimates the messages corresponding to the all zero codeword when the $i-th$ codeword bit is set equal to p. p takes all values between $2m - d_{\min}/2$ and $2m + d_{\min}/2$, Then iterative decoding is performed until a valid non-zero codeword is obtained. If we find the valid codeword then its Hamming distance HD will be calculated and compare it with MHD. If the new HD is smaller than MHD then we replace the new HD with the MHD otherwise discard the new HD. So we can individuate the MHD and evaluate the performance of the TC. We use this new iterative approach in our proposed scheme to evaluate the performance of different interleaver, code rates and packet lengths.

To get better performance for the double binary TC for video communication, the particular block length is selected that behaves better in the low error rates by calculating the minimum distance using iterative approach for m-binary turbo code.

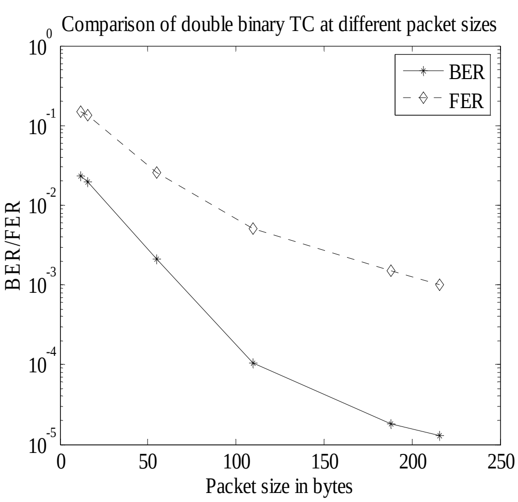

Figure 3: Performance of DBTC at different packet sizes

2.3 Scalable video coding

Our scalable video codec [5] is based on a motion compensated temporal filtering (MCTF) and of spatial wavelet transform based on wavelets, producing a set of spatio-temporal subbands and $3D$ embedded entropy coding. The multiresolution structure resulting from MCTF and *2D* or $t+2D$ subband decomposition enables temporal and spatial resolution scalabilities, respectively. Spatio-temporal decomposition followed by a *2D* or $3D$ embedded entropy coding leads to fine granular quality scalability on all supported spatial and temporal resolutions. The input video is encoded, producing the bitstream of the maximum required quality - which, if the application requires, can be up to quasi-lossless (resulting in imperceptible quality loss). The main features of the codec are: hierarchical variable size block matching motion estimation, scalable motion vectors, flexible selection of wavelet filters for both spatial and temporal wavelet transform on each level of decomposition, *2D* wavelet transform in lifting implementation and embedded zerotree block entropy coder. The used scalable video bitstream consists of packets of data called atoms. An atom represents the smallest entity that can be added or removed from the bitstream. For easier interpretation, the bitstream can be represented in a $3D$ $(q-t-s)$ space (q = Quality, t = Temporal resolution, s = Spatial resolution), as shown in Figure 4. In this figure, $Q = T = S = 2$. Q, T, S are the number of refinement layers in quality, temporal and spatial domain, respectively. There exists a base layer in each domain that is referred as 0^{th} layer except from refinement layers and cannot be removed from the bitstream. Therefore, in the example shown on Figure 4 we have 3 quality, 3 temporal and 3 spatial layers. Each atom has its coordinates in $q-t-s$ space, which are denoted by (q,t,s). High level description of an actual scalable video bitstream is illustrated in [7,11].

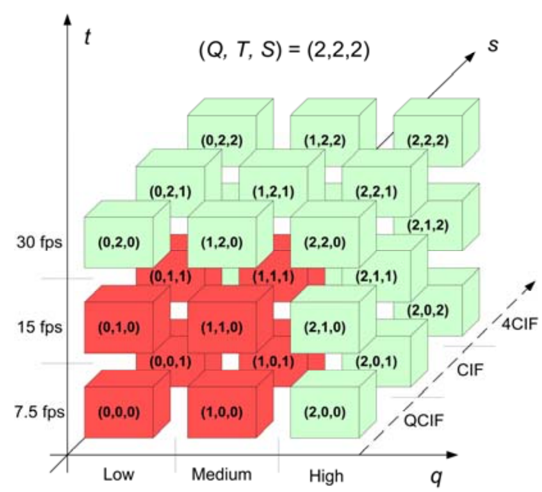

Figure 4: 3D representation of a scalable video bitstream

As there are several progression orders for multi-domain scalable bitstream, specifically for our case of quality, temporal and spatial scalability there are 3! = 6 orders. These are: QTS, QST, TQS, TSQ, SQT, STQ, where the convention is that the first letter refers to the domain which progresses most slowly, while the last to the one which progresses most quickly.

3 Proposed technique

In the video transmission system, we utilize JSCC for effective selection of a source coding rate and a channel coding that will guide to the least amount of overall end to end distortion D_S or maximize the PSNR of total system at a given total rate R_B. The problem we address is as follows. Given R_S is the overall system rate, R_{SVC} is the rate of the SVC coder for all scalable layers, R_{CRC} is the rate after packetization and R_{TC} is the channel coder rate. We have to optimize total bits such that

$$\min D_S \text{ subject to } R_S \leq R_B \quad (3)$$

or

$$\max(PSNR)_S \text{ subject to } R_S \leq R_B \quad (4)$$

where

$$R_S = R_{SVC} / R_{TC} \quad (5)$$

To optimize the equations 1 and 2, the atomic structure of scalable bitstream is adapted regarding the domain that progress most slowly. For example using the default setting of the SVC, the quality layers progress most slowly. Then we packetize each quality layer using packets of equal size. During the packetizing of the scalable bitstream, cyclic redundancy check (CRC) bits are added to check the error free delivery of packets at decoder side. Each quality layer is protected by FEC.

We use the double binary TC as FEC to adapt scalable video bitstream. Double binary TC's bit rate is twice at the decoder output as compared to the binary TC because it decodes two bits at the same number of turbo iterations. Double binary TC is less sensitive to puncturization so we can easily use unequal error protection for scalable bitstream with CRC bits.

The lowest quality layer and bitstream header is adapted with lowest channel rate that means with highest protection and the remaining layers are adapted with respect to its importance. The available code rates are ($1/3$, $2/5$, $1/2$, $2/3$, $3/4$, $4/5$ and $6/7$) [6]. The interleaver design and packet size of DBTC are also considered by using the dmin. So higher dmin DBTC is used for important layers and vice versa.

At the decoder side if a packet is error-corrupted, the CRC check after channel decoding then we point out the corresponding atom in the bitstream. Let an atom (q_x, t_x, s_x) be corrupted after channel decoding or fails to qualify, then CRC checks and remove all the atoms from the bitstream that are corrupted

At low to moderate Eb/No, the proposed technique provides better results, but at moderate to high Eb/No, the equal error protection (EEP) technique is better. To optimize the performance of proposed technique, we use adaptive double binary TC, we use EEP for moderate to high Eb/No and UEP is used when channel condition is poor. For this feedback information from the channel is used.

4 Experimental evaluation

Several simulations were performed using the SVC codec, double binary TC together with QPSK/BPSK or higher order modulation on AWGN and Rayleigh fading channels. Selected results for the BUS CIF sequence at 30fps are reported in this section.

We found the best channel rates according to different quality layers with CRC bits are reported in Table 1.

Quality layers and $R_{SVC} + R_{CRC}$	Q1 100 kbps	Q2 80 kbps	Q3 70 kbps	Q4 75 kbps
R_{TC} for EEP	1/2	1/2	1/2	1/2
R_{TC} for UEP	1/3	1/2	2/3	6/7

Table 1: R_{TC} for different quality layers

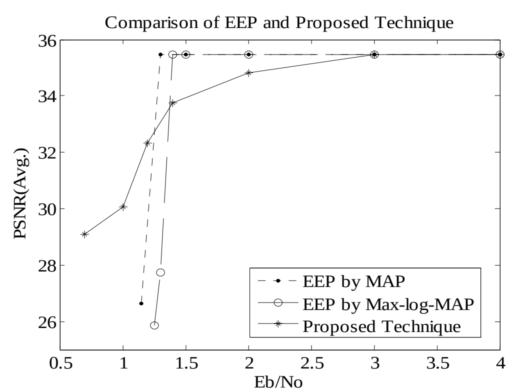

Figure 5: Proposed technique vs Equal Error Protection by MAP and Max-log-MAP decoding algorithm for channel coding

Figure 5 and Figure 6 show the average PSNR against the signal to noise ratio Eb/No (Eb represents energy per bit and No is the one sided noise spectral density) for scalable video transmission over an AWGN channel and Rayleigh fading channel respectively. R_S for EEP is 650 kbps. EEP is applied by using MAP and Max-log-MAP decoding algorithm for channel coding and rate of channel coding R_{TC} is set to $1/2$.

The results show that there is only 0.15db gain in MAP with respect to Max-log-MAP algorithm for double binary code but the complexity of MAP is too high with respect to Max-log-MAP. So we use Max-log-MAP for the proposed technique of JSCC because it improves efficiency.

As the error protection is applied on a quality layer wise, the system rate R_S is 650 kbps as well. Each quality layer is adapted with different rate of double binary TC R_{TC} as reported in Table 1.

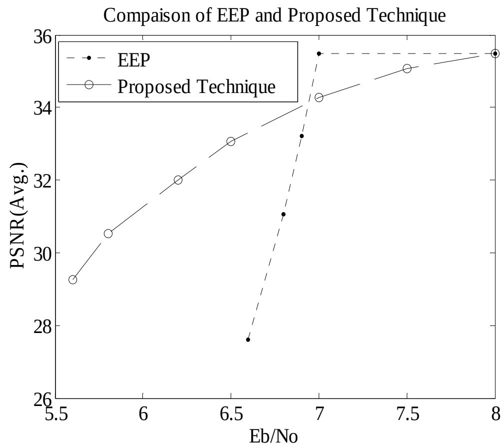

Figure 6: Comparison of EEP and Proposed Technique over Rayleigh fading channel

The proposed technique shows graceful degradation of the PSNR at low Eb/No. But at moderate Eb/No region, the performance of the proposed technique is worst than the EEP. This problem can be overcome by using the adaptive double binary TC by knowing the channel condition and the result is shown in Figure 7.

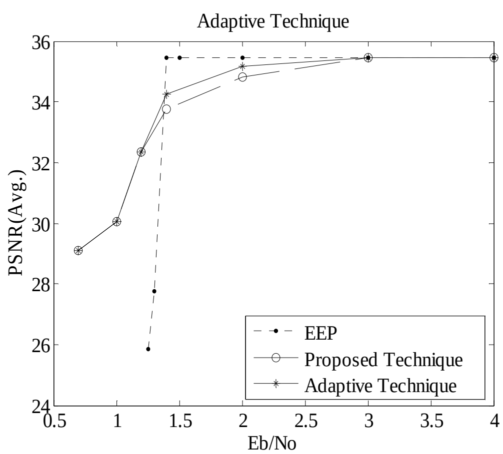

Figure 7: Adaptive Technique by using the feedback information from channel

5 Conclusions

In this paper, we have described an efficient JSCC scheme that combines SVC and double binary TC. The results clearly show that the proposed technique provides a more graceful pattern of quality degradation as compared to EEP at low signal to noise ratio. In the moderate region of Eb/No the adaptive double binary TC enhance the performance of the system. In future work, we will investigate the best trade off between BER and PSNR with respect to the packet size.

Acknowledgements

This research was partially supported by the European Commission under contract FP6-001765 aceMedia.

References

[1] Banister, B.A.; Belzer, B.; Fischer, T.R. “Robust video transmission over binary symmetric channels with packet erasures”, Proceedings. DCC 2002 pp. 162 – 171, April 2002

[2] C. Berrou, A. Glavieux, P. Thitimajshima, **“**Near Shannon limit error-correcting coding and decoding: Turbocodes.”, 1993. ICC 93. Geneva. Technical Program, Conference Record, IEEE *International Conference onCommunications*, vol.2, pp. 1064 -1070 May 1993.

[3] C. Berrou, Y. Saouter, C. Douillard, S. Kerouédan and M. Jézéquel, “Designing good permutations for turbo codes: towards a single model,” in *Proc. IEEE Int. Conf. Communications,* Paris, France, June 2004.

[4] C. Douillard, and C. Berrou. “Turbo Codes With Rate-m/(m + 1) Constituent Convolutional Codes,” *IEEE Trans. On Comm.* Vol. 53, No. 10, 10 Oct 2005.

[5] Marta Mrak, Nikola Sprljan, Toni Zgaljic, Naeem Ramzan, Shuai Wan and Ebroul Izquierdo, “*Performance evidence of software proposal for Wavelet Video Coding Exploration group”*, ISO/IEC JTC1/SC29/WG11/ MPEG2006/M13146, 76th MPEG Meeting, Montreux, Switzerland, April 2006.

[6] Nikola Sprljan, Marta Mrak, Ebroul Izquierdo, "A Fast Error Protection Scheme for Transmission of Embedded Coded Images over Unreliable Channels and Fixed Packet Size", Proc. 30th IEEE International Conference on Acoustics, Speech, and Signal Processing 2005, ICASSP 2005, Philadelphia, PA, USA, 19-23 March 2005, vol.3, pp. 741-744.

[7] N. Sprljan, M. Mrak, G. C. K. Abhayaratne, E. Izquierdo. “A scalable coding framework for efficient video adaptation”, *Proc. Int. Work. on Image Analysis for Multimedia Interactive Services* (*WIAMIS*), (April 2005).

[8] P. Robertson, P. Hoeher, and E. Villeburn. Optimal and Suboptimal Maximum a Posterioi Algorithms suitable for Turbo Decoding. *Euro Tr, Telecomm.* 8: 119-125, Mar.-Apr. 1997.

[9] Q. Zhang, W Zhu and Y. Zhang, “Channel-Adaptive Allocation for Scalable Video Transmission over 3 G Wireless Networks”, *IEEE Trans. on Circuits and Systems for Video Technology* Vol. 14, No. 8, 1049-1063, Dec. 2004.

[10] Sergio Verdu, “Fifty Years of Shannon Theory,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2057–2077, Jan. 1995.

[11] T. Zgaljic, N. Sprljan and E. Izquierdo, “Bitstream syntax description based adaptation of scalable video”, *EWIMT 2005*.