Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2007, Article ID 47517, 12 pages
doi:10.1155/2007/47517

Research Article

Joint Source-Channel Coding for Wavelet-Based Scalable Video Transmission Using an Adaptive Turbo Code

Naeem Ramzan, Shuai Wan, and Ebroul Izquierdo

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

Received 20 August 2006; Revised 18 December 2006; Accepted 5 January 2007

Recommended by James E. Fowler

An efficient approach for joint source and channel coding is presented. The proposed approach exploits the joint optimization of a wavelet-based scalable video coding framework and a forward error correction method based on turbo codes. The scheme minimizes the reconstructed video distortion at the decoder subject to a constraint on the overall transmission bitrate budget. The minimization is achieved by exploiting the source rate distortion characteristics and the statistics of the available codes. Here, the critical problem of estimating the bit error rate probability in error-prone applications is discussed. Aiming at improving the overall performance of the underlying joint source-channel coding, the combination of the packet size, interleaver, and channel coding rate is optimized using Lagrangian optimization. Experimental results show that the proposed approach outperforms conventional forward error correction techniques at all bit error rates. It also significantly improves the performance of end-to-end scalable video transmission at all channel bit rates.

1. INTRODUCTION

The design of robust video transmission techniques over heterogeneous and unreliable channels has been an active research area over the last decade. This is mainly due to its commercial importance in applications such as video transmission and access over the Internet, multimedia broadcasting and video services over wireless channels. In traditional video communications over heterogeneous channels, the video is usually processed offline. Compression and storage are tailored to the targeted application according to the available bandwidth and potential end-user receiver or display characteristics. However, this process requires either transcoding of compressed content or storage of several different versions of the encoded video. None of these alternatives represent an efficient solution. Furthermore, video delivery over error-prone heterogeneous channels meets additional challenges such as bit errors, packet loss, and error propagation in both spatial and temporal domains. This has a significant impact on the decoded video quality after transmission in some cases rendering useless the received content. Consequently, concepts like scalability, robustness, and error resilience need to be reassessed to allow for both efficiency and adaptability according to individual transmission bandwidth, user preferences, and terminals.

Scalable video coding (SVC) promises to partially solve this problem by "encoding once and decoding many." SVC enables content organization in a hierarchical manner to allow decoding and interactivity at several granularity levels. That is, scalable coded bit streams can efficiently adapt to the application requirements. Thus, problems inherent to the diversity of bandwidth in heterogeneous networks and improved quality of services can be tackled. Wavelet-based SVC provides a natural solution for error-prone transmissions with a truncatable bit stream. In addition, channel coding methods can be adaptively used to attach different degrees of protection to different bit-layers according to their relevance in terms of decoded video quality.

Following Shannon's theorem of separability [1], source and channel coding can be considered and optimized independently. However, Shannon's theorem assumes that the source and channel codes are of arbitrary large lengths. This assumption does not hold in practical situations due to limitations on computational power and processing delays. Consequently, joint source-channel coding (JSCC) emerges as the model to overcome the underlying problem in real-world applications. JSCC has been extensively studied in the literature [2–17]. It consists of three basic aspects: finding an optimal distribution of limited resources (such as total transmission rate) between source coder and channel coder [3], designing the source coder to achieve the target source rate, and enhancing the robustness of channel coding [5].

Usually, JSCC applies different degrees of protection to different parts of the bitstream. That means unequal error protection (UEP) is used according to the importance of a given portion of the bitstream. In this context, scalable coding emerges as the natural choice for highly efficient JSCC with UEP, since wavelet-based SVC provides different bitlayers of different importance with respect to decoded video resolution or quality [18]. The impact of applying UEP in base and enhancement layers for fine granularity scalable source coders is discussed in [3–6]. In [12] UEP is applied on progressive data by using Reed Solomon (RS) codes and turbo codes. In these works only the channel coding rate is regarded as adaptive with respect to a progressive bitstream. However, the performance of JSCC not only depends on the channel rate, but also on other parameters inherent to the used channel coder, for example, packet size and interleaver design in turbo coders. These aspects could become critical in the design of efficient JSCC models. Unfortunately, they are less reported in the conventional literature. This important shortcoming of conventional JSCC techniques is addressed in this paper.

The JSCC approach proposed in this paper exploits the joint optimization of the wavelet-based SVC reported in [18] and a forward error correction method (FEC) based on turbo codes [19]. The underlying wavelet-based scalable video coding framework achieves fine granularity scalability using combinations of spatio-temporal transform techniques and 3D bit-plane coding [20]. The spatio-temporal transform consists of 2D wavelet transform and motion compensated temporal filtering (MCTF), which provide spatial and temporal scalabilities, respectively [21]. For the sake of completeness, important characteristics of the used wavelet-based SVC are briefly reviewed in the next section. Regarding channel coding, turbo codes (TC) are one of the most prominent FEC techniques having received great attention since their introduction in [19]. Its popularity is mainly due to its excellent performance at low bit error rates, reasonable complexity, and versatility for encoding packets with various sizes and rates. In this paper, double binary TC (DBTC) [22] is used for FEC rather than the conventional binary TC, as DBTC usually performs better than classical TC in terms of better convergence for iterative decoding, a large minimum distance and low computational cost.

The proposed JSCC scheme minimizes the reconstructed video distortion at the decoder subject to a constraint on the overall transmission bitrate budget. The minimization is achieved by exploiting the source rate distortion (RD) characteristics and the statistics of the available codes. Here, the critical problem of estimating the bit error rate (BER) probability in error-prone applications is also discussed. Regarding the error rate statistics, not only the channel coding rate, but also the interleaver and packet size for TCs are considered in the proposed approach. The aim is to improve the overall performance of the underlying JSCC. In order to optimize the parameter section, an analytical algorithm to evaluate the performance of the channel coder is proposed. It is based on estimating the minimum distance between the zero codeword and any other codeword. It will not escape the reader's notice that so far the problem of finding minimum distance remains an open problem. Solving that problem is crucial to evaluate the performance of DBTCs accurately. An iterative method is proposed to find the minimum distance. Using the proposed technique, the speed and accuracy of approximating the error rate are improved with respect to other techniques from literature, for example, the techniques reported in [23, 24]. At the decoding side, a cyclic redundancy check (CRC) is performed after DBTC decoding. Corrupted bitstream portions, that is, parts of the bitstream failing the CRC, are then removed before source decoding.

The remaining paper is organized as follows. Section 2 outlines important aspects of the two cornerstones of the proposed JSCC framework: wavelet-based SVC and DBTC. The characteristics of the SVC bitstream are presented and the relevance of fine granularity scalability for efficient JSCC is described. Furthermore, generic aspects of the DBTC are also described in Section 2. Details of the proposed JSCC are presented in Section 3. Specifically, the proposed JSCC distortion estimation approach and the iterative algorithm to find the minimum distance in DBTC are discussed. Selected results from computer simulations are given in Section 4. The paper closes with conclusions and a brief discussion on future research directions in Section 5.

2. SYSTEM OVERVIEW

The proposed framework consists of two main modules as shown in Figure 1: scalable video encoding and UEP encoding. At the sender side, the input video is coded using the wavelet-based scalable coder [18]. The resulting bitstream is adapted according to channel capacities. The adaptation can also be driven by terminal or user requirements when this information is available. The adapted video stream is then passed to the UEP encoding module where it is protected against channel errors. Three main submodules make up the UEP encoding part. The first one performs packetization, interleaver design, and CRC. The second one estimates and allocates bit rates using a rate-distortion optimization. The last UEP encoding submodule is the actual DBTC. After quadrature phase shift keying (QPSK) modulation, the video signal is transmitted over a lossy channel. At the receiver side, the inverse process is carried out. The main processing steps of the decoding are outlined in Figure 1. In this paper additive white Gaussian noise (AWGN) and Rayleigh fading channels are considered. However, the proposed method can be equally applied to other lossy channels. Two critical parts of the framework depicted in Figure 1 are the wavelet-based scalable coder and the DBTC module. For the sake of completeness, these two modules are elaborated in the remaining of this section.

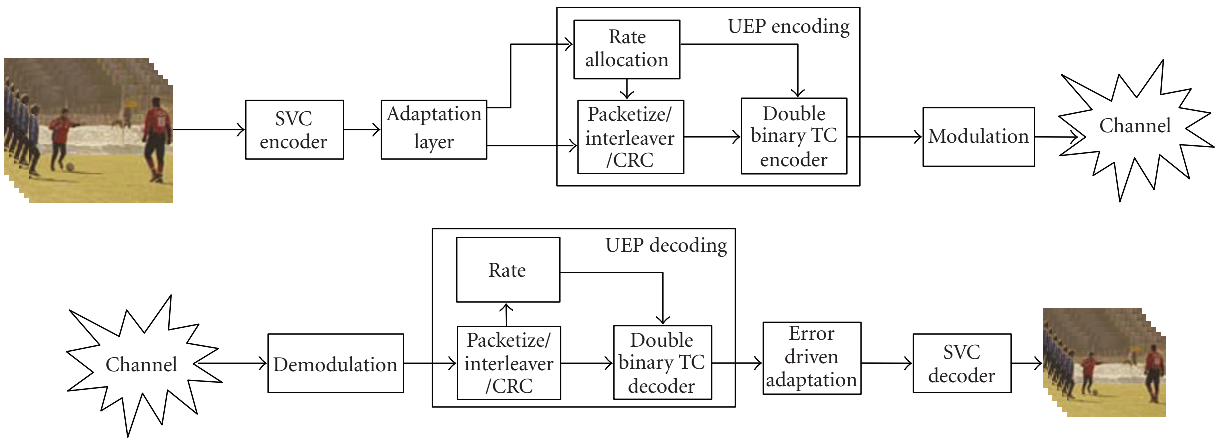

Figure 1: Communication chain for video transmission.

2.1. *Scalable video coding*

The scalable video codec considered in this paper is based on the wavelet transform performed in temporal and spatial domains [18]. In this wavelet-based video coder, temporal and spatial scalability are achieved by applying a 3D wavelet transform on the input frames. In the temporal domain MCTF with flexible choice of wavelet filter is used. In the spatial domain adaptive 2D wavelet transform is applied. The multiresolution structure resulting from MCTF and 2D subband decomposition enables temporal and spatial resolution scalabilities. The MCTF results in motion information and wavelet coefficients that represent the texture of transformed frames. These wavelet coefficients are then bit-plane encoded in order to achieve quality scalability. The used embedded entropy coding leads to fine granular quality scalability on all supported spatial and temporal resolutions. The resulting fine granular quality scalability is used to steer the targeted unequal error protection of the FEC technique in the JSCC, as detailed in the next section.

The main features of the used codec are [20] hierarchical variable size block matching motion estimation, flexible selection of wavelet filters for both spatial and temporal wavelet transform on each level of decomposition, including the 2D adaptive wavelet transform in lifting implementation and embedded zero-tree block entropy coder. For a more detailed description of the complete architecture and features of the wavelet-based scalable coder the reader is referred to [18].

The input video is initially encoded with the maximum required quality. The compressed bitstream features a highly scalable yet simple structure. The smallest entity in the compressed bitstream is called an atom, which can be added or removed from the bitstream. The bitstream is divided into group of pictures (GOPs). Each GOP is composed of a GOP header, the atoms, and allocation table of all atoms. Each atom contains the atom header, motion vectors data, and texture data of a certain subband. The bitstream structure is shown in Figure 2.

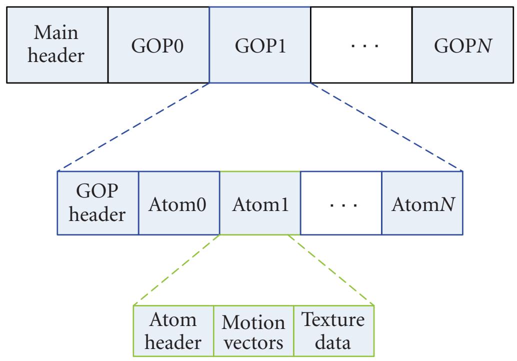

Figure 2: A detailed description of used scalable bitstream.

For the sake of visualization and simplicity, the bitstream can be represented in a 3D space with coordinates q = Quality, t = Temporal resolution, and s = Spatial resolution, as shown in Figure 3. There exists a base layer in each domain that is referred to as 0th layer and cannot be removed from the bitstream. Therefore, in the example shown on Figure 3, 3 quality, 3 temporal, and 3 spatial layers are depicted. Each atom has its coordinates in (q, t, s) space.

2.2. *Double binary turbo codes*

Double binary TCs were introduced by Douillard and Berrou in [22]. These codes consist of two binary recursive systematic convolutional (RSC) encoders of rate 2/3 and an interleaver of length k. Each binary RSC encoder encodes a pair of data bits and produces one redundancy bit. Thus, 1/2 is the natural rate of a DBTC. In this article, the 8-state DBTC with generator polynomials (15,13) in octal notation is considered. It is well known that due to its excellent performance, this DBTC has been widely adopted by the European Telecommunications Standards Institute (ETSI) for Digital Video Broadcasting (DVB). The architecture of DBTC encoder is shown in Figure 4.

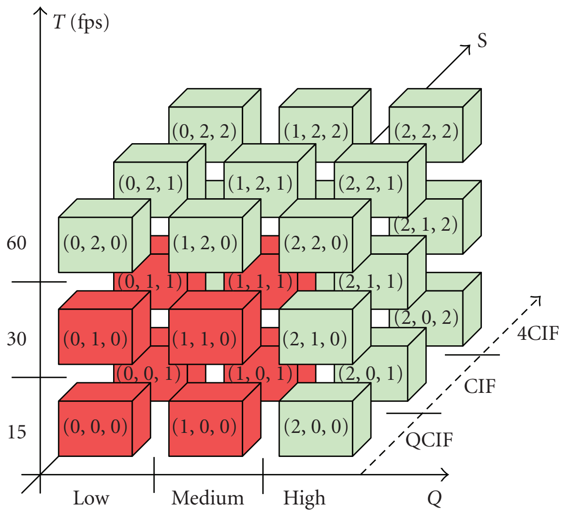

Figure 3: 3D representation of a scalable video bitstream.

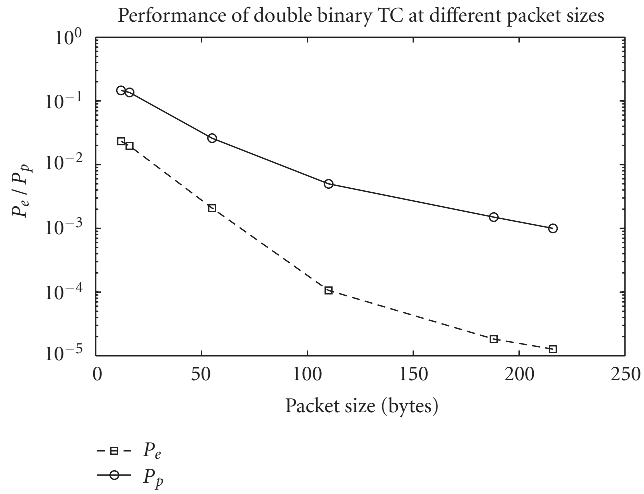

Figure 5: Performance of DBTC at different packet sizes with rate $R_1 = 1/2$.

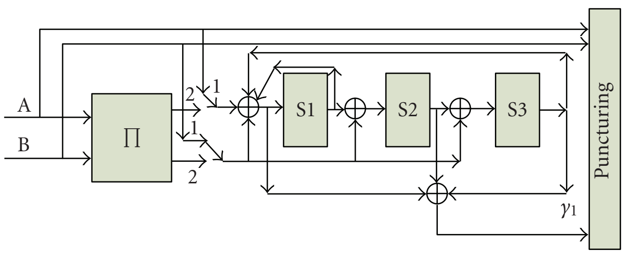

Figure 4: Double binary turbo encoder.

The turbo decoder is usually composed of two Maximum A Posteriori (MAP) or Max-log-MAP decoders [25], one for each stream produced by the singular RSC block as shown in Figure 4. Since the iterative process is similar for both MAP and Max-log-MAP algorithm, and explained in [22, 25].

In this iterative process the interleaver design is critical since the performance of the TC depends on how well the information bits are scattered by the interleaver. Permutations of almost regular permutation (ARP) and di-thered relative prime (DRP) interleavers are elaborated in [26, 27], respectively. A comparison of DVB standard interleaver and DRP interleaver has been performed and reported in [24]. According to this analysis DRP is more stable at high signal-to-noise ratio E_b/N_o, while DVB is comparatively more steady for low E_b/N_o. Therefore, how to adaptively select according to source-channel condition is critical for the overall performance of JSCC.

Furthermore, the performance of the DBTC is also significantly influenced by its packet size. For example, the performance of DBTC with different packet sizes at channel rate $R_1 = 1/2$ at SNR $= 1.2$ dB for 1000 packets is illustrated in Figure 5, where P_e is bit error probability, P_P is the packet error probability. Generally speaking, the performance of DBTC improves as the packet size increases for a given channel rate. However, the best tradeoff of packet size is also crucial to the overall performance.

To find the optimum parameters, the performance of DBTC needs to be evaluated for each set of permutation parameters. Unfortunately, at low error rates the performance of turbo coders fluctuate significantly even when very large interleaver lengths are used. This fact renders an unfeasible exhaustive evaluation of the permutation parameters in practical applications. As a consequence, the development of effective tools to estimate turbo coder's performances at low error rates becomes acute. Two methods to estimate the performance of TCs by minimum distance ($d_{\min}$) have been proposed recently in [23, 24]. Although these techniques differ in several aspects, they present an important common feature: at low error rates, the TC performance is approximated by

$$P_p \le \left(\frac{1}{2}\right) n(d_{\min}) \operatorname{erfc}\left(\sqrt{d_{\min} R_1 \left(\frac{E_b}{N_o}\right)}\right),$$
$$P_e \le \left(\frac{1}{2}\right)\left(\frac{w_{\min}}{k}\right) \operatorname{erfc}\left(\sqrt{d_{\min} R_1 \left(\frac{E_b}{N_o}\right)}\right). \tag{1}$$

In (1), $R_1 = k/n$ is the rate of the code, E_b is the energy per information bit, N_o is the one-sided noise spectral density, $d_{\min}$ is the minimum distance between the zero codeword and any other codeword, $n(d_{\min})$ is its multiplicity, $w_{\min}$ is the sum of the Hamming weights of the input sequences generating the codewords with Hamming weight $d_{\min}$, and erfc(x) is the complementary error function. Since the parameters R_1 and E_b/N_o in (1) are either known or can be fixed, estimating the code performance becomes equivalent to estimate the minimum Hamming distance between codewords.

Observe that on the one hand the algorithm to find $d_{\min}$ proposed in [23] (error impulse method) is quite efficient but it may converge to a wrong $d_{\min}$. On the other hand, the double error impulse method introduced in [24] gives more accurate results at the expense of time efficiency. Based on this observation a new iterative approach to measure minimum distance of m-Binary TC is proposed and used in the JSCC framework described in this paper. Using the proposed method, the performance of a TC is effectively evaluated by considering different rates R_1, packet sizes, and interleavers. Hence, the bit error probability and packet error probability are being estimated for each available rate, packet size, and interleaver at given channel conditions with accuracy and less complexity. Then the best combination will be selected using RD optimization. The new iterative method to find $d_{\min}$ and RD optimization will be proposed in detail in Section 3.

3. JOINT SOURCE-CHANNEL CODING

The objective of JSCC is to jointly optimize the overall system performance subject to a constraint on the overall transmission bitrate budget. As mentioned before, a more effective error resilient video transmission can be achieved if different channel coding rates are applied to different bitstream layers, that is, quality layers generated by the SVC encoding process. Furthermore, the parameters for FEC should be jointly optimized taking into account available and relevant source coding information. For instance, when DBTC is considered, there are at least the three main aspects that can be optimized to achieve better performance in terms of bit error probability, speed and power: channel code rate; packet size and how the input is interleaved before being fed into the second encoder. An ideal selection of these parameters should lead to minimum overall combined source-channel distortion. Observe that the packet size should be carefully chosen since it influences the bit error probability. To determine optimal channel rate, packet size, and interleaver, the overall RD characteristics should also be considered during channel encoding under given channel conditions.

3.1. Rate distortion optimization for JSCC

In the proposed JSCC framework, DBTC encoding is used for FEC before BPSK/QPSK modulation. CRC bits are added in the packetization of DBTC in order to check the error status during channel decoding at the receiver side. Effective selection of the channel coding parameters leads to a minimum overall end-to-end distortion, that is, maximum system PSNR, at a given channel bit rate. The underlying problem can be formulated as

$$\min D_{s+c} \quad \text{subject to } R_{s+c} \le R_{\max} \tag{2}$$

or

$$\max(\text{PSNR})_{s+c} \quad \text{subject to } R_{s+c} \le R_{\max} \tag{3}$$

for

$$R_{s+c} = \frac{R_{\text{SVC}}}{R_{\text{TC}}}, \tag{4}$$

where D_{s+c} is the expected distortion at decoder, R_{s+c} is the overall system rate, R_{SVC} is the rate of the SVC coder for all quality layers, R_{TC} is the channel coder rate and $R_{\max}$ is the given channel capacity. Here the index notation $s+c$ stands for combined source-channel information.

The constrained optimization problem (2)–(4) can be solved by applying unconstrained Lagrangian optimization. Accordingly, JSCC aims at minimizing the following Lagrangian cost function J_{s+c}:

$$J_{s+c} = D_{s+c} + \lambda \cdot R_{s+c}, \tag{5}$$

with λ the Lagrangian parameter. In the proposed framework the value of λ is computed using the method proposed in [3]. Since quality scalability is considered in this paper, R_{s+c} in (5) is defined as the total bit rate over all quality layers:

$$R_{s+c} = \sum_{i=0}^{Q} R_{s+c,i}. \tag{6}$$

To estimate D_{s+c} in (5), let $D_{s,i}$ be the source coding distortion for layer i at the encoder. Since the wavelet transform is unitary, the energy is supposed to be unaltered after wavelet transform. Therefore the source coding distortion can be easily obtained in wavelet domain. Assuming that the enhancement quality layer i is correctly received, the source channel distortion at the decoder side becomes $D_{s+c,i} = D_{s,i}$. On the other hand, if any error happens in layer i, the bits in this layer and in the higher layers will be discarded. Therefore, assuming that all layers h, for $h < i$, are correctly received and the first corrupted layer is $h = i$, the jointly source-channel distortion at any layer $h = i, i+1, \dots, Q$, at the receiver side becomes $D_{s+c,h} = D_{s,i-1}$. Then, the overall distortion is given by

$$D_{s+c} = \sum_{i=0}^{Q} p_i \cdot D_{s,i}, \tag{7}$$

where p_i is the probability that the ith quality layer is corrupted or lost while the jth layers are all correctly received for $j = 0, 1, 2, \dots, i-1$. Finally, p_i can be formulated as

$$p_i = \left(\prod_{j=0}^{i-1} (1 - pl_j) \right) \cdot pl_i, \tag{8}$$

where pl_i is the probability of the ith quality layer being corrupted or lost. pl_i can be regarded as the layer loss rate.

According to (8), the performance of the system depends on the layer loss rate, which in turn depends on the DBTC rate, the packet size, and the interleaver. Once the channel condition and the channel rate are determined, the corresponding loss rate pl_i can be estimated by applying an iterative algorithm to estimate minimum distance between the zero code word and any other codeword $d_{\min}$ in the DBTC. Assuming that $d_{\min}$ is available, pl_i can be estimated as

$$pl_i \propto \left(\frac{1}{d_{\min}} \right). \tag{9}$$

Using (9), p_i can be evaluated from (8). As a consequence the problem of finding p_i boils down to find $d_{\min}$. An accurate and efficient algorithm in finding $d_{\min}$ is given in the following section.

Table 1: Minimum distance of DBTC at different code rates and packet sizes by different methods.

Rate of DBTC	Packet size of DBTC (bytes)	$d_{\min}$ by error impulse method	$d_{\min}$ by double error impulse method	$d_{\min}$ by proposed method
1/3	188	31	33	33
1/2	188	19	19	19
2/3	188	13	12	12
3/4	188	9	9	9
1/2	53	18	17	18
1/2	110	16	16	16
1/2	212	19	20	20

3.2. Determine minimum distance

Let $D = (d_1 \cdots d_x \cdots d_z)$ denote an information frame, where $d_x = (d_{x,1} \cdots d_{x,y} \cdots d_{x,m})$ is the vector of m-binary data applied at the input of the turbo encoder at time x. The output of the turbo encoder is $C = (c_1 \cdots c_x \cdots c_n)$. Here, c_x is a vector of length $m + n$ bits. That is, $c_x = (c_{x,1} \cdots c_{x,y} \cdots c_{x,m+n})$, where $c_{x,y}$ is the systematic bit if $y \le m$ and the parity bit if $y > m$. The codeword is mapped by the QPSK modulator into the transmitted vector $w = (w_1 \cdots w_x \cdots w_n)$. Each vector w_x has length $m+n$, that is, $w_x = (w_{x,1} \cdots w_{x,y} \cdots w_{x,m+n})$, where $w_{x,y} = 2c_{x,y} - 1$ for $x = 1 \cdots m + n$. After transmission over the lossy channel, the received vector is

$$\begin{aligned} R_r &= (r_1 \cdots r_x \cdots r_n) \\ \text{with } r_x &= (r_{x,1} \cdots r_{x,y} \cdots r_{x,m+n}). \end{aligned} \tag{10}$$

To describe the iterative technique to estimate $d_{\min}$, let us assume that the all zero codeword, that is, $r_q = -1$ for all q, is received. Initially, $d_{\min}$ is set equal to a large default value. The proposed method estimates the messages corresponding to the all zero codeword when the xth codeword bit is set equal to u. Here, u takes all values between $2m - d_{\min}/2$ and $2m + d_{\min}/2$. Then iterative decoding is performed until a valid nonzero codeword is obtained. The Hamming distance (HD) of a valid codeword is calculated and compared to $d_{\min}$. If the new HD is smaller than $d_{\min}$, then the new HD is assigned to $d_{\min}$, otherwise the newly estimated HD is discarded and the value of u is increased. This process is then repeated until the new $d_{\min}$ is found or an upper limit $2m + d_{\min}/2$ of iterations u is reached. So $d_{\min}$ can be individuated at given interleaver, rate, and packet size.

A thorough experimental evaluation has been conducted to show that the proposed technique to estimate $d_{\min}$ is as accurate as the precise double error impulse method presented in [24], with a much faster process. In fact, the proposed method is as fast as the error impulse method introduced in [23], however with a better precision. Selected results of this evaluation are given in Table 1. In most of the cases the proposed method produces the same result as double impulse method [24] while it appears to be more robust than error impulse method [23]. As an example, Table 2 shows the comparison of different interleavers at rate 1/3 for packet size 188 bytes. The results from Table 2 indicate that the performances of ARP, DVB, and DRP are comparably good, whereas the S-random interleaver performs much worse for double binary TC. Therefore, only ARP, DVB, and DRP interleavers are considered in the proposed JSCC.

Table 2: Minimum distance of different interleavers at rate = 1/3 for packet size 188 bytes by the proposed method.

Interleaver	$d_{\min}$ by proposed method
S-random	18
DVB	33
ARP	34
DRP	36

This iterative approach to measure $d_{\min}$ is used to evaluate the performance of different interleavers, code rates, and packet lengths and hence to estimate the lost probability of the ith layer p_i in (8). Using the proposed method in the determination of $d_{\min}$, the estimated end-to-end distortion can be computed. Substitute corresponding distortion and rate into (5), the Lagrangian cost for each combination of channel rate, packet size, and interleaver is computed and compared. The combination leading to the minimum cost will be selected for each quality layer. As described in Section 2, the scalable video coding produces an atomic bitstream where the source distortion, coding bit rates for each quality layer are readily available after coding. In addition, the minimum distance for each packet size and interleaver can be precomputed and stored instead of computing it for each parameter combination. Therefore, it is easy for JSCC to obtain the Lagrangian cost for each parameter combination. Since a finite set of a few quality layers, channel rates, packet sizes, and interleavers is considered, the corresponding computation complexity falls into a practical implementation. However, if many quality layers are encoded in a fine granularity bitstream, or much more components are to be optimized, this exhaustive computation may render the system impractical because of a huge complexity. In this way, dynamic programming could be used during optimization to reduce the complexity. As one of the options, source-channel bit budget can be firstly optimally allocated along the quality layers using dynamic programming. The other parameters for channel coding (packet size and interleaver) can be optimized for each quality layer given a certain channel rate.

After JSCC, the received codeword at the receiver side is demodulated and then decoded by DBTC decoder. The early stopping (ES) technique (CRC check) is used at each half turbo iteration. If the packet of information passes the CRC, then the iterative turbo decoding process is stopped. Otherwise, the iterative decoding process is stopped after six turbo iterations. This ES-based approach enables a significant decrease of channel decoding time. In the DBTC decoder if a packet remains corrupted after six turbo iterations, then the corresponding atoms in the bitstream are labeled as corrupted. If an atom (q_i, t_i, s_i) is corrupted after channel decoding or fails to qualify the CRC checks, then all the atoms which have higher index than i are removed by the error driven adaptation module outlined in Figure 1. Finally, SVC decoding is performed to evaluate the overall performance of the system.

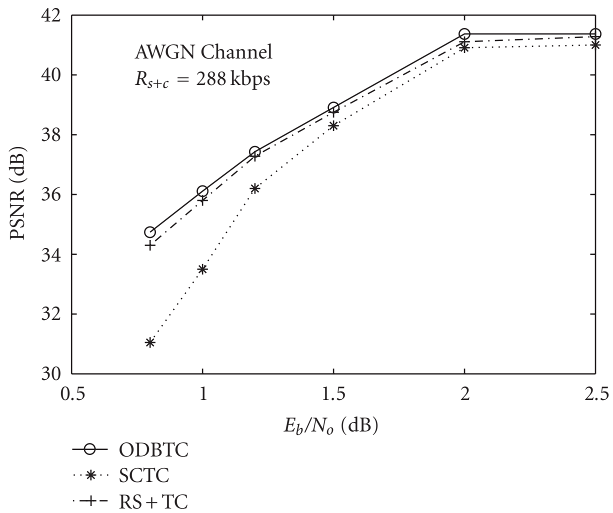

Figure 6: Average PSNR for City QCIF sequence at 15 fps at different signal-to-noise ratio (E_b/N_o) for AWGN channel.

4. EXPERIMENTAL RESULTS

The performance of the proposed JSCC framework has been extensively evaluated using the wavelet-based SVC codec [18]. For the proposed JSCC UEP optimal channel rate, packet size and interleaver for DBTC were estimated and used as described in this paper. The proposed technique is denoted as “ODBTC.” In this paper, DVB, ARP, and DRP interleavers, channel rates (1/3, 2/5, 1/2, 2/3, 3/4, 4/5, and 6/7) and packet sizes (16, 55, 110, 188, 216) in bytes are considered for ODBTC. Max-log-MAP algorithm produces approximately the same result as the MAP algorithm for DBTC, as reported in [22]. That means, the decoding complexity can be decreased without any significant loss of performance for DBTC by using Max-log-MAP algorithm. For this reason, the Max-log-MAP algorithm is used in ODBTC. Two other advanced JSCC techniques were integrated into the same SVC codec for comparison. The first technique used serial concatenated convolutional codes of fixed packet size of 768 bytes and pseudo random interleaver [15]. It is denoted as “SCTC.” Since product code was regarded as one of the most advanced in JSCC, the technique using product code proposed in [12] was used for the second comparison. This product code used RS codes as outer code and turbo codes as inner code [12], so it is denoted by “RS + TC” in this paper. It is noticeable that this scheme was initially targeting wavelet-based image transmission. Nevertheless it is very straightforward to extend them to video transmission by replacing the image subbands using quality layers of scalable video in RS + TC. The corresponding parameters in [12] were adopted for video in RS + TC in this paper.

After QPSK modulation, the protected bitstreams were transmitted over error-prone channels. Both AWGN and Rayleigh fading channels were used in the experimental evaluation. For each channel emulator, 50 simulation runs were performed, each one using a different error pattern. The decoding bit rates and sequences for signal-to-noise ratio (SNR) scalability defined in [28] were used in the experimental setting. For the sake of conciseness the results reported in this paper include only certain decoding bit rates and test sequences: City at QCIF resolution and Soccer at CIF resolution and several frame rates. Without loss of generality, the $t + 2D$ scenario for wavelet-based scalable coding was used in all reported experiments. The average PSNR of the decoded video at various BER was taken as objective distortion measure. The PSNR values were averaged over all decoded frames. The overall PSNR for a single frame was computed by

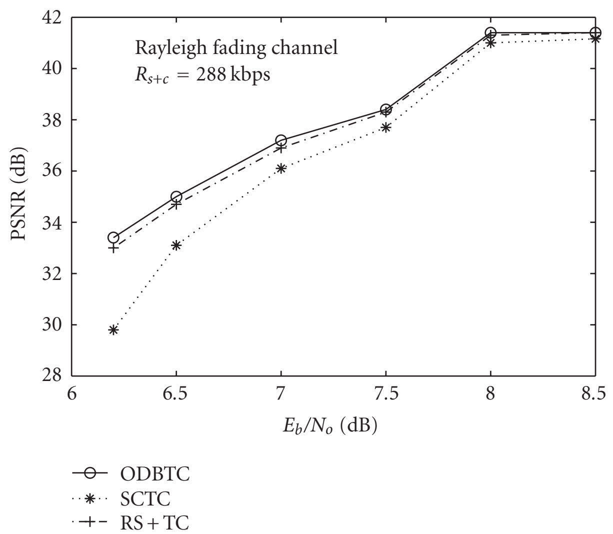

Figure 7: Average PSNR for City QCIF sequence at 15 fps at different signal-to-noise ratio (E_b/N_o) for Rayleigh fading channel.

$$\mathrm{PSNR} = \frac{(\mathrm{PSNR\_}Y + \mathrm{PSNR\_}U/4 + \mathrm{PSNR\_}V/4)}{1.5}, \tag{11}$$

where PSNR_Y, PSNR_U, and PSNR_V denote the PSNR values of the Y, U, and V components, respectively.

A summary of PSNR results is shown in Figures 6 to 8. These results show that the proposed UEP ODBTC consistently outperforms SCTC and achieving PSNR gains at all

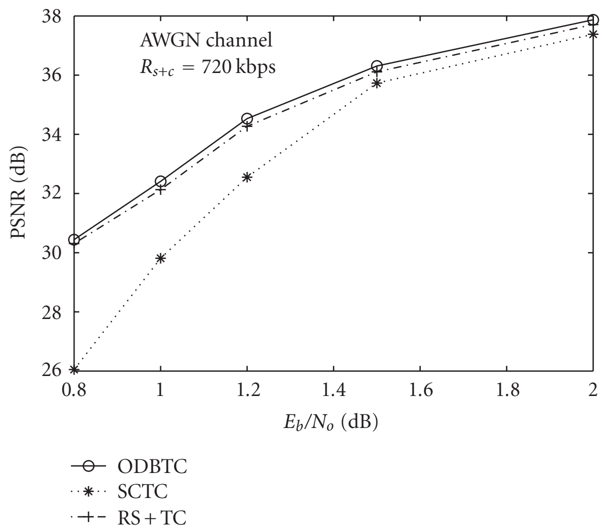

FIGURE 8: Average PSNR for Soccer CIF sequence at 30 fps at different signal-to-noise ratio (E_b/N_o) for AWGN channel.

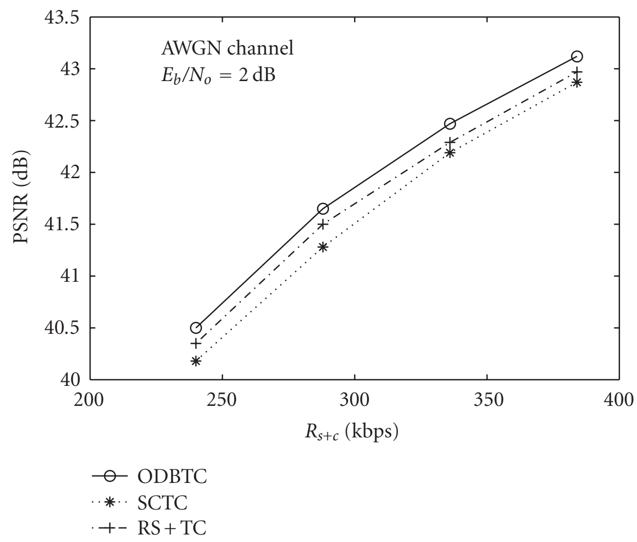

FIGURE 9: PSNR performance of City QCIF at 15 fps at different bit rates.

signal-to-noise ratios (E_b/N_o) for both AWGN and Rayleigh fading channels. Specifically, for the sequence City up to 3 dB can be gained by SCTC when low E_b/N_o or high channel errors are considered for both AWGN channel and Rayleigh fading channel. A similar behaviour for AWGN is reported for sequence Soccer in Figure 8. It can be observed that the proposed scheme achieves the best performance among different channel conditions. As the channel errors increase or E_b/N_o decreases, a gap between the proposed scheme and SCTC becomes larger. The performance of RS + TC is almost comparable to ODBTC, with a slight PSNR degradation in most of the cases. However, it should be noticed that RS + TC uses product code where a much larger complexity will be introduced by encoding and decoding of RS codes and TC together.

A summary of PSNR results is shown in Figures 9 and 10 at different decoded bit rates, for City QCIF 15 fps at 288 kbps and Soccer CIF 30 fps at 720 kbps. These results show that for the considered channel conditions, the proposed ODBTC consistently outperforms the SCTC, achieving PSNR gains at all tested bit-rates. Specifically, for the sequence City up to 1 dB can be gained for Rayleigh fading channel at 7 dB, while up to 0.3 dB over SCTC, when low channel errors for AWGN channel are considered. RS + TC performs better than SCTC, but comparable to ODBTC. At high SNR, the gap is widened up to 0.4 dB.

Figures 11 and 12 show the PSNR _Y performance versus frame number of the compared methods for the same test conditions. As an observation the proposed ODBTC consistently displays a higher PSNR compared to the SCTC, while its performance is slightly better than RS + TC.

These results also confirm the consistent better performance of the proposed technique ODBTC for both AWGN and Rayleigh fading channels. Figure 11 shows comparison results for the City sequence at 288 kbps at an E_b/N_o = 1.7 dB. It will not escape the reader's notice that ODBTC has a higher PSNR fluctuation than the other two techniques. The observed PSNR fluctuation is inherent to scalable video coding for certain sequences and bit rates. After transmission, corrupted quality layers have to be discarded due to channel errors, resulting in a rather smooth but blurred sequence. However, when error protection is effective, more quality layers will be recovered and the resulting sequence is very close to the one at the original bit rate. From a different point of view, this fluctuation also serves to some extent to appreciate the better error protection of the proposed approach. Considering PSNR values, it can be seen that our proposed scheme shows better PSNR in every frame at low error rate. More quality layers will be recovered and the resulting sequence is very close to the one at the original bit rate. Furthermore, the performance is even better at higher error rate (E_b/N_o = 7.2 dB) for Rayleigh fading channel, as shown in Figure 12 for the Soccer CIF sequence at 720 kbps.

Selected results of subjective quality improvements are also given in Figure 13. Here, a comparison of reconstructed 90th frame of City QCIF at 15 fps and 288 kbps is displayed. Again, the three different approaches in the low E_b/N_o at 1 dB are considered. The original, reconstructed without FEC, 90th frame of the same sequence is shown at the top-right of Figure 13. It can be observed that the image quality obtained by the proposed UEP scheme is much better than the one obtained with the SCTC and a slightly better than the RS + TC.

The superior performance of the proposed ODBTC has been demonstrated in the previous experiments. Extensive experiments have been conducted to evaluate the gain of each individual parameter in the proposed method. Here two techniques are evaluated and compared with ODBTC: UEP-A and UEP-B. For UEP-A, the DBTC used fixed packet size of 188 bytes and DVB interleaver. In this case only the channel rates were adapted to quality layers using RD optimization. For UEP-B the interleaver design as well as channel coding

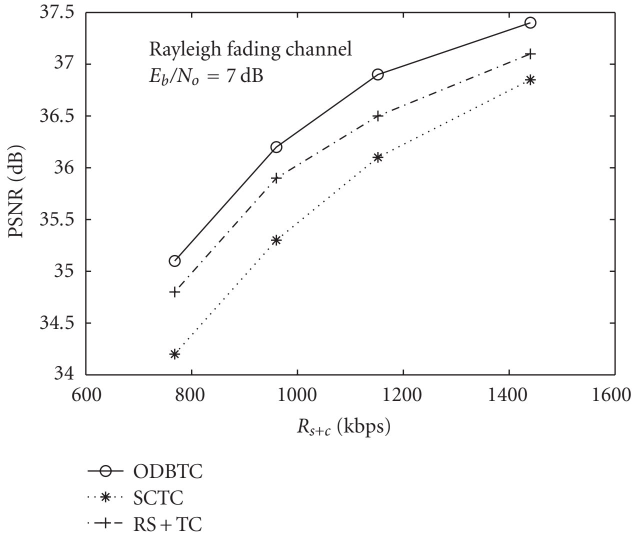

Figure 10: PSNR performance of Soccer CIF at 30 fps at different bit rates.

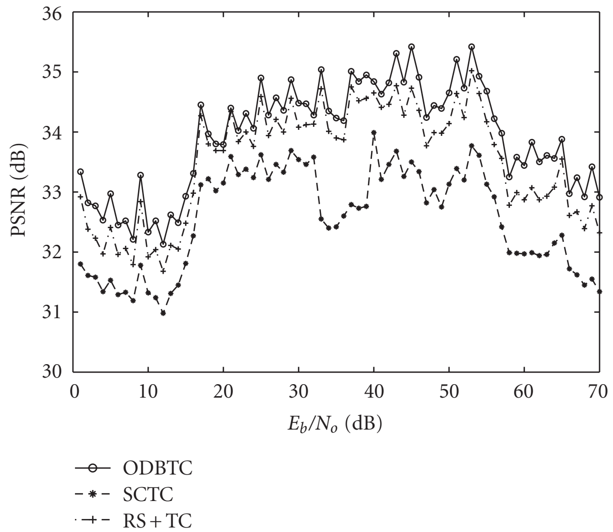

Figure 12: PSNR _Y performance for different frames of Soccer CIF sequence at 720 kbps at $E_b/N_o = 7.2$ dB for Rayleigh fading channel.

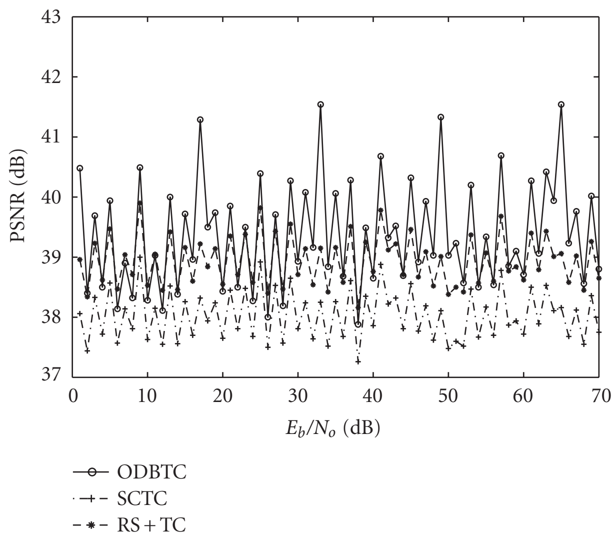

Figure 11: PSNR _Y performance for different frames of City QCIF sequence at 288 kbps at $E_b/N_o = 1.7$ dB for AWGN channel.

rate were optimized together, using fixed packet size of 188 bytes. The compared results indicate that at high E_b/N_o, the major gain is from interleaver design but at low E_b/N_o, the gain is from choosing different packet sizes, as shown in Figure 14.

In addition, the performance gain of using RS codes as outer code is also evaluated. RS codes were integrated to the proposed ODBTC to recover the turbo code packets that fail the CRC test after maximum number of turbo iterations, which was fixed to 6. Here RS code was used as the outer code while DBTC as the inner code. The DBTC was first optimized using the proposed method, and RS codes were further implemented using RD optimization proposed in [12]. The results are reported in Figures 15 and 16 for AWGN and Rayleigh fading channels, respectively. It can be concluded that using RS codes as the outer code improves the performance of ODBTC. However, the gain is marginal for bit error channels considered in this paper. Specifically, only 0.3 dB at high E_b/N_o and about 0.05 dB at low E_b/N_o advantages can be obtained. Actually, RS codes are very effective for burst errors. Therefore, using RS codes as outer code is very useful when the inner code has bursty erroneous paths, for example, RCPC codes [29]. However, the error pattern of DBTC is more complicated and rather randomly distributed. Accordingly, the advantage of RS codes is not so effective for DBTC codes as well as TC [29]. Therefore the gain from RS codes together with DBTC is marginal, because the error pattern of DBTC is more complicated and rather randomly distributed like TC [29]. However, the complexity of introducing RS codes is not neglectable. Consequently, ODBTC is proposed in this paper considering the applied channel condition and system complexity. Apparently, when packet loss or burst error is considered, more significant performance gain can be expected using RS codes as the outer code.

5. CONCLUSION

In this paper, an efficient approach for joint source and channel coding is presented. The proposed approach exploits the joint optimization of the wavelet-based SVC and a forward error correction method based on turbo codes. UEP is used to minimize the end-to-end distortion by considering the channel rate, packet size of turbo code and interleaver at given channel conditions and limited complexity. To efficiently optimize the channel coding parameters, an iterative approach is proposed to estimate the minimum distance of

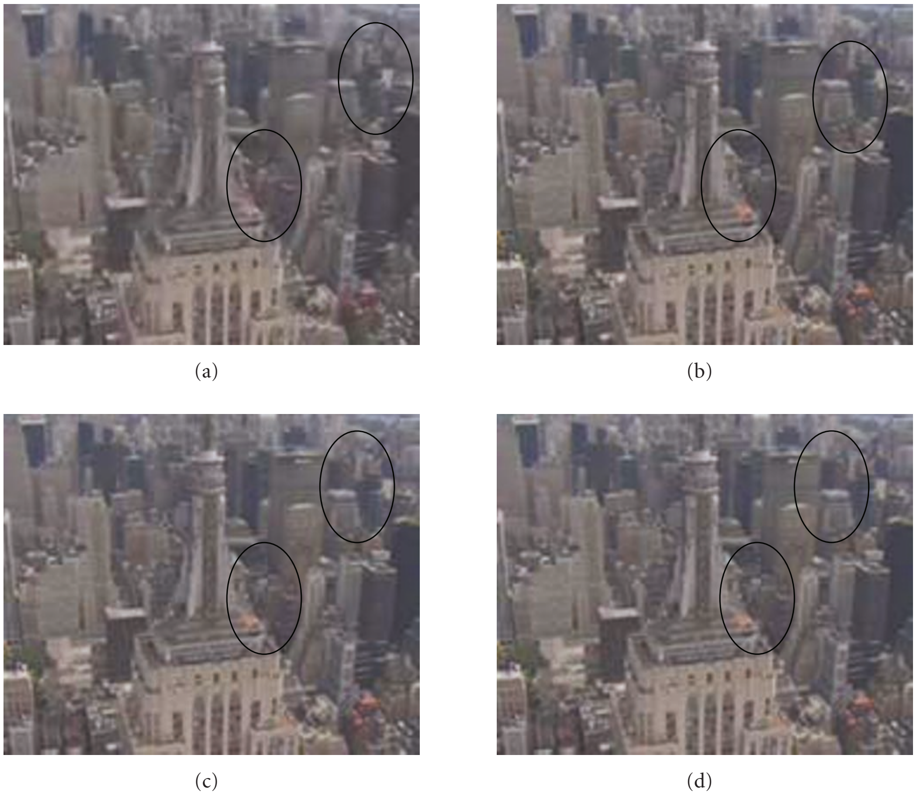

Figure 13: Comparison of the reconstructed 90th frame of City QCIF at 15 fps sequence in the $E_b/N_o = 1$ dB. (a) Original reconstructed frame without FEC PSNR_$Y = 33.42$ dB. (b) Reconstructed by SCTC PSNR_$Y = 37.83$ dB. (c) Reconstructed by RS + TC PSNR_$Y = 39.34$ dB. (d) Reconstructed by ODBTC PSNR_$Y = 40.08$ dB.

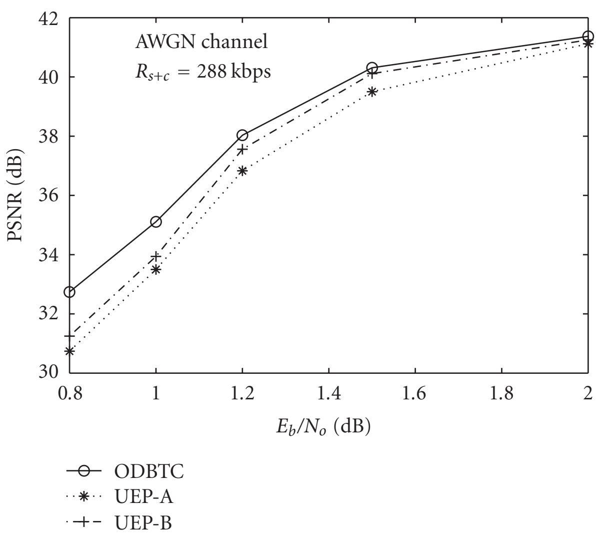

Figure 14: Performance comparison of optimizing different parameters in the proposed technique for City QCIF at 15 fps sequence.

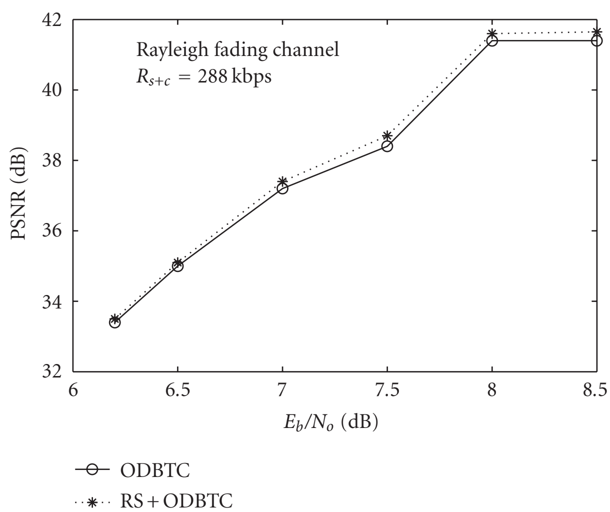

Figure 15: Performance of proposed technique with and without RS code for City QCIF sequence at 15 fps.

DBTC. The results of computer experiments show that the proposed technique provides a more graceful pattern of quality degradation as compared to conventional UEP in literature at different channel errors. The performance using RS code as the outer code is also evaluated.

Important aspects remain open and will be tackled in future extensions of this work. They include better error concealment schemes tailored to the proposed framework; adaptive modulation schemes, and the evaluation of permutation parameters for ARP interleavers.

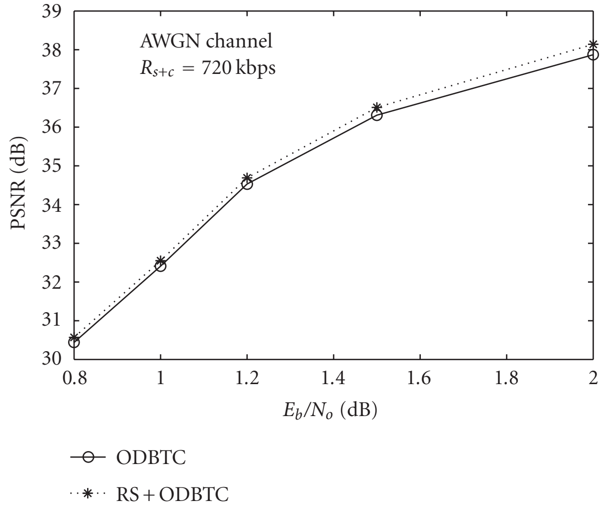

FIGURE 16: Performance of proposed technique with and without RS code for Soccer CIF sequence at 30 fps.

ACKNOWLEDGMENT

We wish to acknowledge support provided by the European Commission under Contract FP6-001765 aceMedia.

REFERENCES

[1] S. Verdü, "Fifty years of Shannon theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2057–2078, 1998.

[2] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Channel-adaptive resource allocation for scalable video transmission over 3G wireless network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 8, pp. 1049–1063, 2004.

[3] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 340–356, 2000.

[4] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 121–131, 2003.

[5] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 1043–1052, 2002.

[6] M. van der Schaar and H. Radha, "Unequal packet loss resilience for fine-granular-scalability video," *IEEE Transactions on Multimedia*, vol. 3, no. 4, pp. 381–394, 2001.

[7] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 819–828, 2000.

[8] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Transactions on Image Processing*, vol. 8, no. 3, pp. 305–320, 1999.

[9] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 511–523, 2002.

[10] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust internet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 357–372, 2001.

[11] N. Sprljan, M. Mrak, and E. Izquierdo, "A fast error protection scheme for transmission of embedded coded images over unreliable channels and fixed packet size," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 3, pp. 741–744, Philadelphia, Pa, USA, March 2005.

[12] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1890–1901, 2005.

[13] J. Thie and D. Taubman, "Optimal erasure protection strategy for scalably compressed data with tree-structured dependencies," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2002–2011, 2005.

[14] R. Hamzaoui, V. Stankovic, and X. Zixiang, "Optimized error protection of scalable image bit streams [advances in joint source-channel coding for images]," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 91–107, 2005.

[15] B. A. Banister, B. Belzer, and T. R. Fischer, "Robust video transmission over binary symmetric channels with packet erasures," in *Proceedings of Data Compression Conference (DCC '02)*, pp. 162–171, Snowbird, Utah, USA, April 2002.

[16] B. Barmada, M. M. Ghandi, E. V. Jones, and M. Ghanbari, "Combined turbo coding and hierarchical QAM for unequal error protection of H.264 coded video," *Signal Processing: Image Communication*, vol. 21, no. 5, pp. 390–395, 2006.

[17] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1710–1720, 2003.

[18] M. Mrak, N. Sprljan, T. Zgaljic, N. Ramzan, S. Wan, and E. Izquierdo, "Performance evidence of software proposal for Wavelet Video Coding Exploration group," in *ISO/IEC JTC1/SC29/WG11/ MPEG2006/M13146, 76th MPEG Meeting*, Montreux, Switzerland, April 2006.

[19] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[20] T. Zgaljic, N. Sprljan, and E. Izquierdo, "Bitstream syntax description based adaptation of scalable video," in *Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT '05)*, pp. 173–178, London, UK, November-December 2005.

[21] M. Mrak, N. Sprljan, and E. Izquierdo, "Motion estimation in temporal subbands for quality scalable motion coding," *Electronics Letters*, vol. 41, no. 19, pp. 1050–1051, 2005.

[22] C. Douillard and C. Berrou, "Turbo codes with rate-m/(m+1) constituent convolutional codes," *IEEE Transactions on Communications*, vol. 53, no. 10, pp. 1630–1638, 2005.

[23] C. Berrou, S. Vaton, M. Jézéquel, and C. Douillard, "Computing the minimum distance of linear codes by the error impulse method," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '02)*, vol. 2, pp. 1017–1020, Taipei, Taiwan, November 2002.

[24] Y. Ould-Cheikh-Mouhamedou and S. Crozier, "Comparison of distance measurement methods for turbo codes," in *Proceedings of Canadian Workshop on Information Theory (CWIT '05)*, pp. 36–39, Montréal, Quebec, Canada, June 2005.

[25] P. Robertson, P. Hoeher, and E. Villeburn, "Optimal and suboptimal maximum a posterioi algorithms suitable for turbo decoding," *European Transactions on Telecommunications*, vol. 8, pp. 119–125, 1997.

[26] C. Berrou, Y. Saouter, C. Douillard, S. Kerouédan, and M. Jézéquel, "Designing good permutations for turbo codes: towards a single model," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 1, pp. 341–345, Paris, France, June 2004.

[27] S. Crozier and P. Guinand, "High-performance low-memory interleaver banks for turbo-codes," in *Proceedings of 54th IEEE Vehicular Technology Conference (VTC '01)*, vol. 4, pp. 2394–2398, Atlantic City, NJ, USA, October 2001.

[28] R. Leonardi, S. Brangoulo, M. Mark, M. Wien, and J. Xu, "Description of testing in wavelet video coding," in *ISO/IEC JTC1/SC29/WG11/ MPEG2006/N7823, 75th MPEG Meeting*, Bangkok, Thailand, January 2006.

[29] G. Zhou, T.-S. Lin, W. Wang, et al., "On the concatenation of turbo codes and Reed-Solomon codes," in *Proceedings of IEEE International Conference on Communications (ICC '93)*, vol. 3, pp. 2134–2138, Anchorage, Alaska, USA, May 2003.