PERFORMANCE ANALYSIS OF A DISTRIBUTED VIDEO CODING SYSTEM – APPLICATION TO BROADCASTING OVER AN ERROR-PRONE CHANNEL

J. Farah[1], C. Yaacoub[1,3], F. Marx[2], and B. Pesquet-Popescu[3]

(1) Faculty of Sciences & Computer Engineering, Holy-Spirit University of Kaslik, Jounieh, Lebanon
(2) France Telecom, Paris, France
(3) Ecole Nationale Supérieure des Télécommunications, Paris, France

ABSTRACT

In this paper, we present a novel approach for determining the theoretical compression bound for Wyner-Ziv coded video sequences transmitted over error-prone channels. This estimation takes into account the amount of motion inside each particular frame, as well as the transmission channel conditions. We compare the analytical results with those of a practical Wyner-Ziv coding system where a turbo-decoder acts as a joint decoder for recovering the compressed information and correcting errors due to transmission channel impairments. Furthermore, we show how the entropy limits can be used for estimating the optimal amount of information to be transmitted per frame, in case of a broadcasting application where a feedback channel is not available. Simulation results show a small degradation in the system rate-distortion performance, compared to the case of source-channel coding based on feedback information from the receiver.

1. INTRODUCTION

During the last few years, distributed source coding [1-4] has made tremendous progress in the world of video communication. Its aim is to counteract the high complexity of conventional compression techniques such as the H26x or MPEG standards based on predictive interframe coding. The concept emerged from Slepian's and Wolf's theory [5]: Given two statistically dependent sources X and Y, with Y being separately compressed to its entropy limit H(Y), X can be transmitted at a rate which is very close to the conditional entropy H(X/Y), known as the Slepian-Wolf limit. The application of this concept to lossy source coding is known as the Wyner-Ziv coding [6], which permits a simple encoder structure and a much more efficient compression for source X, provided that Y is perfectly recovered as side information at the receiver. This scenario was applied in wireless sensor networks [3] where several information sources are collected and processed by a central base. It was also proposed for low-complexity video encoders in [7] where intra-frame source compression is realized by adequate puncturing of the parity bits generated at the output of a turbo-encoder. The systematic information is not transmitted but rather replaced by side information available at the receiver, leading to inter-frame decoding. One of the major drawbacks of this scheme is that it employs a feedback channel between the transmitter and the receiver. Therefore, its application in a real-time environment or a broadcasting application is rather impossible. On the other hand, the theoretical lower bound on the achievable compression performance of such a distributed coding system still needs to be determined, especially for the realistic case of an error-prone transmission system. In [8], a source-channel codec was proposed for a broadcasting system: a Wyner-Ziv bitstream is generated at a fixed rate and employs an MPEG decoded bitstream as side information. In this work, we first consider a source-coding system employing feedback information. Then, we generalize our study to the case of a broadcasting system: still employing independent frame encoding, the transmitter relies on a low-complexity estimation of the theoretical bounds to vary the necessary amount of transmitted data for each particular frame, depending on the transmission channel conditions and on the amount of motion in the frame.

The remainder of the paper is organized as follows: in section 2, we start by describing the joint source-channel codec based on rate-compatible punctured turbo-coding [9]. Then, in section 3, we present our approach for the estimation of the system's compression limit in the presence of noise. Simulation results, as well as theoretical curves, are presented and discussed in section 4.

2. DESCRIPTION OF THE JOINT SOURCE-CHANNEL CODING SYSTEM

The distributed joint source-channel coding system considered in this study can be represented by the block-diagram in Figure 1. The intra-frame encoding system compresses only the even frames in the video sequence. Odd frames are considered to be perfectly recovered at the receiver as side information used as systematic data for source-channel decoding and in the reconstruction of even frames. Side information of a particular even frame is generated by an average interpolation of the two adjacent (preceding and succeeding) odd frames [7].

It can be verified that the statistics of the residual signal d resulting from subtracting an even frame from the corresponding interpolated frame follow a Laplacian distribution: $P(d) = \frac{\alpha}{2} e^{-\alpha|d|}$. (1)

The parameter α can be approximately estimated on the receiver side using the available odd frames. It can also be estimated by the encoder and transmitted as side information to the receiver.

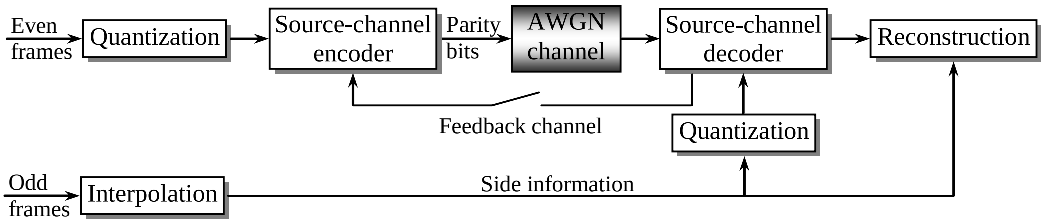

Figure 1. Block-diagram of the pixel-domain joint source-channel coding system.

Compression of the even frames starts by a uniform scalar quantization to obtain M-bit representations of the eight-bit pixels. In this work, we consider M = 1, 2 or 4 in order to obtain different source coding rates. The source-channel encoder consists of a parallel concatenation of two 16-state quadri-binary convolutional encoders separated by an internal interleaver and resulting in a minimum global coding rate of 2/3. The generator polynomials in octal notation are $(23, 35, 31, 37, 27)_8$ from [10]. As for the interleaver length, it is set to the size of the quantized video frame. At the encoder output, systematic information is discarded, while parity information is punctured and transmitted to the decoder. Therefore, the system maximum compression rate is 0.5, which corresponds to the case where all the generated parity bits are transmitted. At the receiver side, turbo-decoding is realized by iterative Soft Input Soft Output (SISO) decoders based on the Max-Log-MAP (Maximum A Posteriori) algorithm [11]. However, we modified the metric calculations in order to take into account the 16 possible transitions between any couple of trellis states. Moreover, the conditional probabilities in the turbo-decoding process must rely on the residual signal statistics between the even frames and the side information, on one hand, and the channel conditions (the Additive White Gaussian Noise channel over which the parity information is transmitted) on the other (see equation (3) developed later).

In the case of a transmission system with a feedback channel, the decoder requests parity bits from the transmitter until a symbol error rate of 10^{-3} is reached. However, in the case of a broadcasting application, the system must rely on a pre-determined amount of parity information that can be estimated using the theoretical compression limits (H_f) calculated by the transmitter for the even frames: for a fixed transmission rate (for example, the average bitrate D_t obtained by the system with feedback), the theoretical limits are used by the encoder to determine a quasi-optimal partitioning scheme of the total amount of parity bits between the different frames. The transmission rate for a particular frame f will be:

$D_f = D_t \cdot H_f / H_t$, where H_t is the sum of the compression limits over the sequence frames.

After source-channel decoding, the reconstruction block is used to recover an eight-bit version of the even frame using the available side information [7]: each decoded symbol is compared to the corresponding side information. If the latter lies within the same quantization interval of the decoded symbol, the reconstructed pixel will take the value of the side information. Otherwise, it will be clipped to the interval boundary closest to the side information.

3. ANALYTICAL ESTIMATION OF THE SYSTEM LOWER COMPRESSION BOUND

To derive the theoretical compression bound of the practical pixel-domain Wyner-Ziv coding system in figure 1, we consider (figure 2) a discrete source Z being transmitted over an AWGN channel to yield a received discrete sequence X. In addition, side information Y is available at the decoder in such a way that the residual difference between Z and Y exhibits a Laplacian distribution. Since variables X, Y and Z are quantized representations of the video frame pixels, we further assume the presence, in the proposed model, of a hard decision device at the channel output. Throughout the paper, $\{i_1, i_2, ..., i_M\}$ will denote the binary representation of a pixel i, i_1 being the most significant bit. Note that, in the absence of AWGN, the system compression limit would be the conditional entropy H(Z|Y). Due to the presence of AWGN, the uncertainty about the transmitted source increases and so the new compression limit is H(X|Y).

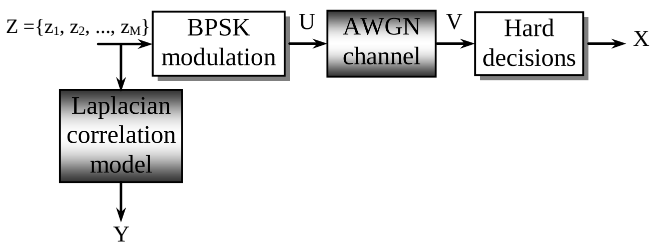

Figure 2. Theoretical model for the derivation of the system compression limit in the presence of noise.

Let N be the additive noise vector: $N = \{n_1, n_2, ..., n_M\}$, where n_k has a normal distribution with zero mean and variance σ^2. The hard decision function at the output of the AWGN channel is given by: $f_d(v) = 1$ if $v \geq 0$, 0 otherwise.

The theoretical achievable compression rate for noisy transmission is estimated using:

$$H(X \mid Y) = -\sum_{i=0}^{2^M-1} \sum_{j=0}^{2^M-1} P(Y = j) \cdot g\left[P(X = i \mid Y = j)\right], \quad (2)$$

where $g(x) = x \cdot \log_2(x)$. The conditional probability can be calculated by:

$$P(X = i \mid Y = j) = \sum_{k=0}^{2^M-1} P\left(Z = k, f_d(U + N) = i \mid Y = j\right). \quad (3)$$

Since the additive noise is independent from the source bits:

$$P\left(x_1 = i_1, x_2 = i_2, ..., x_M = i_M \mid Y = j\right) = \sum_{k=0}^{2^M-1} P\left(Z = k \mid Y = j\right) P\left(f_d\left(U+N\right) = i \mid Z = k\right). \quad (4)$$

The noise samples being independent, we can write:

$$P\left(f_d\left(U+N\right) = i \mid Z = k\right) = \prod_{r=1}^{M} P\left(f_d\left(u_r + n_r\right) = i_r \mid z_r = k_r\right). \quad (5)$$

Let $P_N = P\left(n_r > 1\right) = P\left(n_r < -1\right) = (1/2)\mathrm{erfc}\left(1/\sigma\sqrt{2}\right)$, (6)

where erfc(·) stands for the complementary error function.
It can be verified that:

$$P\left(f_d\left(u_r + n_r\right) = i_r \mid z_r = k_r\right) = \begin{cases} 1 - P_N, & \text{if } i_r = k_r \\ P_N, & \text{if } i_r \neq k_r \end{cases}. \quad (7)$$

On the other hand, the conditional probability $P\left(Z = k \mid Y = j\right)$ is calculated using the Laplacian distribution of the residual signal between the even frames and the corresponding side information:

$$P\left(Z - Y = d\right) = c\frac{\alpha}{2} e^{-\alpha\left|d_{Z-Y}\right|}, \quad (8)$$

where $d_{Z-Y} = d \cdot 2^{8-M}$ and c is a scaling factor used to take into account the discrete and bounded values of the quantized pixels Y and Z.

$c = (2/\alpha) \Big/ \sum_{i=-(2^M-1)}^{2^M-1} e^{-\alpha\left|d_i\right|}$, with $d_i = i \cdot 2^{8-M}$. Therefore:

$$c = (2/\alpha) \Bigg/ \left[1 + 2e^{-\alpha 2^{8-M}} \frac{1 - e^{-\alpha 2^{8-M}(2^M - 1)}}{1 - e^{-\alpha 2^{8-M}}}\right]. \quad (9)$$

On the other hand:

$$P\left(Z - Y = d\right) = \sum_j P\left(Z = k = d + j \mid Y = j\right) P\left(Y = j\right) \quad (10)$$

The number of non-zero terms in (10) is equal to the number $L_{d_{k-j}}$ of couples (k,j) that yield the residual difference k – j = d. In the case of an equiprobable source, these couples can be considered to be equally likely: For example, considering a 2-bit quantization, the possible values of d are 0, ±64, ±128 and ±192, with a decreasing order of probability of occurrence. For a particular value of the difference, for example d = 64, couples (64,0), (128,64) or (192, 128) have the same probability to occur. Therefore:

$$P\left(Z = k = d + j \mid Y = j\right) = \frac{2^M}{L_{d_{k-j}}} P\left(Z - Y = d\right). \quad (11)$$

For an M-bit quantization, the residual difference takes the following values: $d_i = i \cdot 2^{8-M},\ i = -(2^M - 1), ..., 2^M - 1$. (12)
We can express L_{d_i} in terms of i by the following relationship: $L_{d_i} = 2^M - \left|i\right|$. (13)

Finally:

$$H\left(X \mid Y\right) = -\sum_{i=0}^{2^M-1}\sum_{j=0}^{2^M-1} \frac{1}{2^M} g\left[\sum_{k=0}^{2^M-1}\left(c\frac{\alpha}{2}\frac{2^M e^{-\alpha\left|d_{k-j}\right|}}{L_{d_{k-j}}} \prod_{r=1}^{M}\left\{\left(1 - P_N\right)\delta_{i_r - k_r} + P_N\left(1 - \delta_{i_r - k_r}\right)\right\}\right)\right], \quad (14)$$

where δ is the Kronecker delta function.

Note that, in the absence of noise, $P_N = 0$. Therefore,

$$P\left(f_d\left(U+N\right) = i \mid Z = k\right) = \begin{cases} 0, & \text{if } i \neq k \\ 1, & \text{if } i = k \end{cases}. \quad (15)$$

In this case, H(X|Y) reduces to:

$$H\left(Z \mid Y\right) = -\sum_{i=0}^{2^M-1}\sum_{j=0}^{2^M-1} \frac{1}{2^M} g\left[c\frac{\alpha}{2}\frac{2^M e^{-\alpha\left|d_{i-j}\right|}}{L_{d_{i-j}}}\right]. \quad (16)$$

In the following section, we will compare the analytical results obtained by equation (14) with an approximate calculation of H(X|Y). The latter can be derived by mapping the AWGN channel to a Binary Symmetric Channel (BSC) by an equivalence of the stability functions of the two channels [12]. This mapping was adopted in [13] for the design of source-channel codes. The relationship between the crossover probability p of the BSC and the symbol energy per noise density ratio (E_S/N_0) of the AWGN channel can be written as:

$$p = \frac{1}{2}\left(1 - \sqrt{1 - \exp\left(-2\frac{E_s}{N_0}\right)}\right). \quad (17)$$

On the other hand, the theoretical compression limit of a binary source transmitted over a BSC channel is related to its entropy limit H(Z|Y) by [1]:

$$H\left(X \mid Y\right) = \frac{H\left(Z \mid Y\right)}{C\left(p\right)}, \quad (18)$$

where C(p) is the capacity of the BSC channel.

4. PRACTICAL RESULTS

In Figures 3 to 5, we represent the achievable compression rate (R: xdB) obtained by simulating the source-coding system, as well as the theoretical lower bound (H: xdB) obtained for different values of the E_S/N_0 ratio and different numbers of quantization bits per pixel (M). 50 even frames were considered for the Carphone video sequence and 180 for the Foreman sequence. The curves labeled 'R' and 'H' are obtained in the absence of noise, whereas the label 'H: xdB (M)' designates the theoretical compression bound estimated by the channel mapping method.

The results obtained for the Carphone sequence with M = 4 show a gap in the achieved compression towards the theoretical limit between 0.06 and 0.14 for the case of noiseless transmission. In the case where E_S/N_0 = 1 dB, the gap range increases to [0.11 ; 0.18]. The high values of the gap, of the achievable compressions and of the theoretical limits correspond to the low values of the parameter α measured for M = 4 (figure 6). In fact, a low value of α indicates a high level of motion in the frame, making the corresponding side information less reliable for the turbo-decoding process. In this case, a greater amount of parity information is required from the encoder. Besides, we noticed that, starting from E_S/N_0 = 3 dB, the theoretical limits for both noisy and noiseless transmission become almost similar, whereas the practical system performances become almost identical to the noiseless case for $E_S/N_0 \geq 5$ dB.

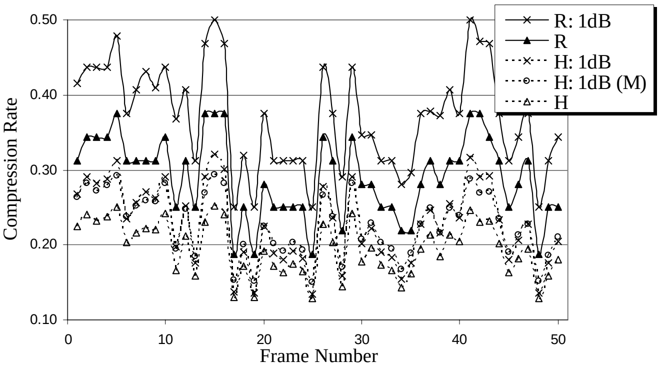

Figure 3. Achievable compression rate for M = 4 (Carphone).

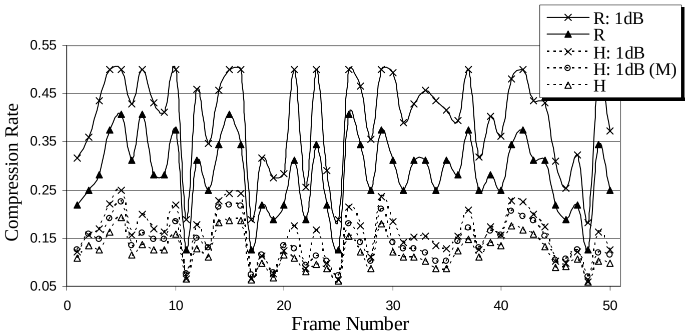

Figure 4. Achievable compression rate for M = 1 (Carphone).

Similar effects were observed for M = 2 and M = 1. However, at E_S/N_0 = 1 dB, the gap range increases to [0.14 ; 0.25] for M = 2 and [0.12 ; 0.33] for M = 1 (figure 4). The high values observed in the practical compression levels, for both cases of noisy and noiseless transmissions, are due to the fact that when the number of quantization levels is low, the correlation between the quantized pixels of the even frames and those in the interpolated frames decreases dramatically, which can be observed from the very low values of the parameter α. This leads to important fluctuations in the achievable compression levels from a frame to another and explains the wider ranges in the performance gaps compared to those obtained for M = 4.

In the case of the Foreman sequence, the level of motion throughout the sequence is much more important than for the Carphone case (note the very low values of the alpha parameter for a certain number of frames in figure 6). In fact, as seen in figure 5, the maximum gap in the system compression performance is very important and reaches almost 0.2 for M = 4. Furthermore, the entropy limit is sometimes higher than the system maximum compression limit (0.5), as in frames 136 to 166. In this case, the decoding bit error rate will saturate at higher values than the 10^{-3} target, leading to a more important distortion than for the rest of the sequence.

On the other hand, we clearly notice that, in general, the theoretical entropy limits obtained by our method are very close to the approximate calculation method. In fact, the latter is less accurate when the compression limits are high, especially for low E_S/N_0 ratios, i.e. for high values of the mapped BSC cross-over probability. However, the closeness between the results of the two estimation methods proves the validity of our theoretical model.

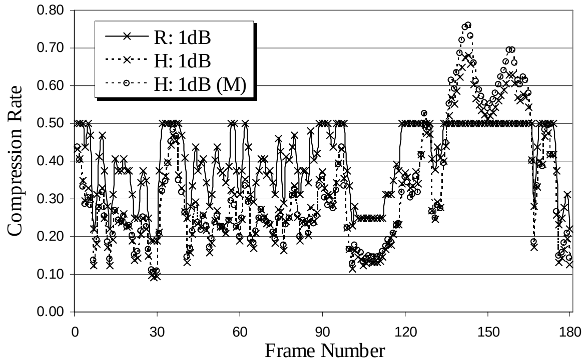

Figure 5. Achievable compression rate for M = 4 (Foreman).

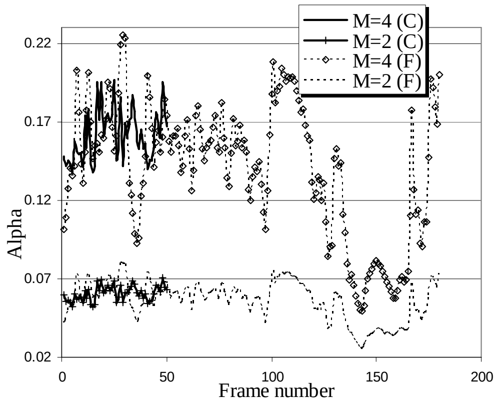

Figure 6. The alpha parameter of the Laplacian distribution for the Carphone (C) and Foreman (F) sequences.

Finally, in figures 7 and 8, we show the rate-distortion curves, i.e. the average PSNR as a function of the average transmission bit rate, obtained for different values of the E_S/N_0 ratio. The bit rates correspond to 15 even frames per second. The theoretical limits (labeled 'TL') are estimated by considering only the influence of quantization, i.e. we suppose perfect recovery of the M-bit frame at the input of the reconstruction block. For the Carphone sequence quantized at M = 4, we find a gap between the theoretical and achieved bit rate from 145 kbps (for a noiseless transmission) to 218 kbps (for E_S/N_0 = 1 dB). For M = 2, the gap is almost 150 kbps at E_S/N_0 = 1 dB and it decreases to 91 kbps for M = 1. The loss in bit rate in the practical compression system due to noise, measured for M = 4, is 13 kbps for E_S/N_0 = 5 dB, 46 kbps for E_S/N_0 = 3 dB, and reaches 124 kbps for E_S/N_0 = 1 dB. In the case of a broadcasting application (curves labeled 'B'), for the same data rate, a loss in the PSNR between 0.85 and 1.1 dB is noticed for M = 2 and 4, towards the system with a feedback channel. By observing the broadcasting system performance at low data rates, we conclude that, in the presence of high noise levels, it is preferable for the decoder to rely on the side information, since the transmitted parity information does not permit a noticeable enhancement in the PSNR.

In the case of the Foreman sequence, the gap towards the theoretical performances is between 100 and 140 kbps for M = 4. However, a loss in the PSNR between 2 and 3 dB is noticed due to the high motion in a great part of the sequence. We also estimated the system performances for the case where the frames corresponding to a theoretical compression limit higher than 0.5 are dropped at the transmitter and replaced by their corresponding side information in the receiver (curves labeled 'D'). This technique permits a considerable gain in the RD performance: for M = 4, a gain in the bit rate between 93 and 127 kbps is observed with a slight increase in the system PSNR.

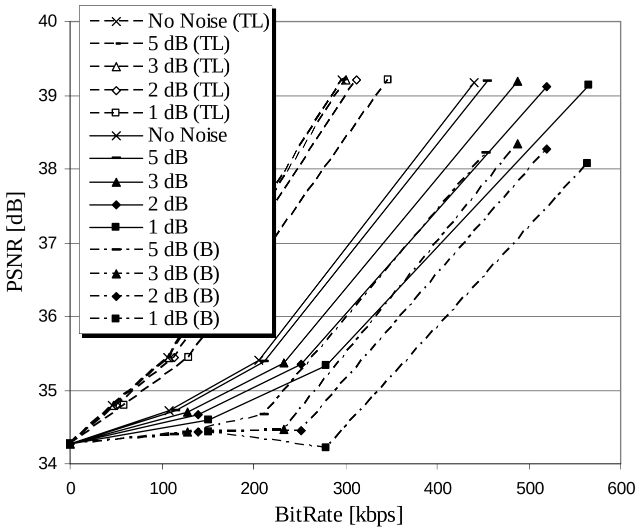

Figure 7. Rate-Distortion curves for the Carphone sequence transmitted in the presence of noise.

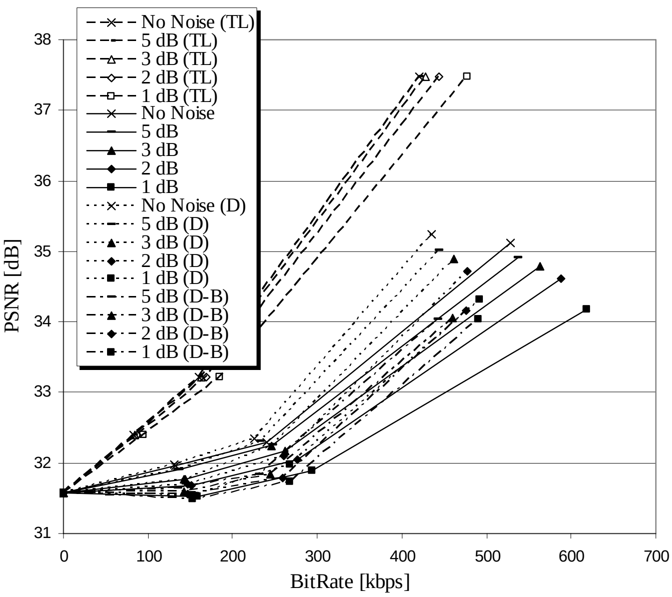

Figure 8. Rate-Distortion curves for the Foreman sequence transmitted in the presence of noise.

Furthermore, when a broadcasting application is considered (curves labeled 'D-B'), the loss in PSNR is only 0.3 dB at E_S/N_0 = 1 dB and reaches 1 dB at E_S/N_0 = 5 dB. For the Foreman sequence, this system permits a transmission at data rates very close to the theoretical bound, but with a loss in the PSNR that can reach 3.4 dB. This loss could be reduced by using complex interpolation techniques for the generation of side information, like motion compensation techniques [7].

5. CONCLUSION

In this paper, we presented an analytical approach for estimating the compression limits of a pixel-domain Wyner-Ziv video coding system with a transmission over error-prone channels. We also implemented a practical source-coding system where compression is achieved using rate-compatible puncturing of turbo-coded sequences. Simulation results show a behavior of the practical compression system performance comparable to the theoretical results: the achievable compression limit for a frame is highly dependent on its content and on the channel conditions. Using the analytical results, the compression level can be predicted for each frame by the encoder and used in a broadcasting system, with a minor loss in the decoding PSNR, compared to the classical feedback-based coding system.

AKNOWLEDGEMENT

This work has been supported by a research grant from the Lebanese National Council for Scientific Research (LNCSR).

REFERENCES

[1] A. Aaron and B. Girod, "Compression with Side Information Using Turbo Codes", *Data Compression Conference*, pp. 252- 261, April 2002.

[2] J. Garcia-Frias and Y. Zhao, "Near-Shannon/Slepian–Wolf Performance for Unknown Correlated Sources over AWGN Channels", IEEE Transactions on Communications, Vol. 53, No. 4, pp. 555-559, April 2005.

[3] S. S. Pradhan and K. Ramchandran, "Distributed Source Coding: Symmetric rates and Applications to Sensor Networks", *Data Compression conference*, pp. 363-372, March 2000.

[4] J. Farah, C. Yaacoub, N. Rachkidy, and F. Marx, "Binary and non-Binary Turbo Codes for the Compression of Correlated Sources Transmitted through Error-Prone Channels", *4th International Symposium on Turbo Codes & Related Topics with the 6th International ITG-Conference on Source and Channel Coding*, Germany, April 2006.

[5] D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Transactions on Information Theory*, Vol. IT-19, pp. 471-480, July 1973.

[6] D.Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Transactions on Information Theory*, Vol. IT-22, pp. 1-10, Jan. 1976.

[7] A. Aaron, R. Zhang, and B.Girod, "Wyner-Ziv Coding of Motion Video", *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, pp. 240-244, Nov. 2002.

[8] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod, "Systematic Lossy Forward Error Protection for Video Waveforms", *IEEE International Conference on Image Processing*, pp. 609-612, Sept 2003.

[9] D.N. Rowitch and L.B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes", *IEEE Transactions on Communications*, pp. 948-959, Vol. 48, June 2000.

[10] D. Divsalar and F. Pollara, "Multiple Turbo Codes", *Military Communication Conference*, pp. 279-285, 1995.

[11] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and Suboptimal Maximum A Posteriori Algorithms Suitable for Turbo Decoding", *European Transactions on Telecommunications*, pp. 119-125, Vol. 8, March-April 1997.

[12] S.-Y. Chung, "On the Construction of Some Capacity-Approaching Coding Schemes", Ph.D. thesis, Massachusetts Institute of Technology, 2000, pp. 137-155.

[13] A.D. Liveris, Z. Xiong, and C. N. Georghiades, "Joint Source-Channel Coding of Binary Sources with Side Information at the Decoder Using IRA codes", *IEEE Workshop on Multimedia Signal Processing*, pp. 53-56, Dec. 2002.