A Hybrid Encoder/Decoder Rate Control for Wyner-Ziv Video Coding with a Feedback Channel

Denis Kubasov, Khaled Lajnef, and Christine Guillemot
IRISA / INRIA Rennes, Campus Universitaire de Beaulieu,
Avenue du Général Leclerc, 35042 RENNES Cedex - France
Email: {denis.kubasov,khaled.lajnef,christine.guillemot}@irisa.fr

Abstract— **This paper describes a hybrid coder/decoder rate control for a Wyner-Ziv video coding scheme with a feedback channel. The approach is first based on a method to estimate at the encoder the minimum rate required for the Slepian-Wolf (SW) encoded data. This estimation makes use only of the Lapacian correlation model. The decoder then estimates the bit error rate (BER) for each decoded bit plane from likelihood ratios computed at the output of the SW decoder. The robustness of the BER estimation is further improved by the use of an error detection mechanism based on a cyclic redundancy checksum. Experimental results show that rate-distortion performances are comparable to those which would be obtained by computing the Hamming distance between the original and the decoded data. In addition the hybrid coder/decoder rate control reduces the decoding complexity as well as the usage of the return channel.**

I. INTRODUCTION

Distributed source coding (DSC) has emerged as an enabling technology for sensor networks. It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. DSC finds its foundation in the seminal theorems of Slepian and Wolf [1] for lossless encoding and of Wyner and Ziv [2] for the lossy case. In practice, a Wyner-Ziv (WZ) codec is obtained by concatenating a Slepian-Wolf codec with a quantiser. The most widely used practical Slepian-Wolf encoders are based on the channel coding principles [3], [4]. Given two binary correlated signals X and Y, the statistical dependence between them is modelled as a virtual correlation channel, and Y is considered as a noisy version of X. The encoder computes the parity bits for X (a systematic channel code is used) and sends them to the decoder (the systematic bits are discarded). The decoder removes the 'virtual' channel noise, and thus estimates X given the received parity bits and the side information (SI) Y regarded as a noisy version of the codeword systematic bits.

Recently video compression has been recast into a distributed source coding framework leading to distributed video coding (DVC) systems targeting mainly low coding complexity and error resilience functionalities. Correlated samples (pixels or transform coefficients) from different frames are regarded as outputs of different sensors. A comprehensive survey of first DVC solutions can be found in [6]. The DVC architecture considered here is based on the solution proposed in [5] and further developed in [6]. In this architecture, the rate of the SW coder (here a turbo coder) is controlled via a feedback channel: when the residual bit error rate (BER) at the output of the turbo decoder exceeds a given threshold, more bits are requested from the encoder via the feedback channel. This amounts to controlling the rate of the code by selecting different puncturing patterns at the output of the turbo code. In many implementations so far, the residual BER was assumed to be the true error rate given by the Hamming distance between the (original) sequence (per bit plane) to be encoded and the reconstructed sequence, which is not realistic.

This paper describes a hybrid coder/decoder rate control strategy based on three novel techniques. The first technique is used by the encoder to estimate the minimum rate required. This allows reducing the number of parity bits requests from the decoder, hence reducing the decoding complexity. The second technique is a method to estimate the residual BER at the output of the turbo decoder. The approach makes use of the log-likelihood ratios (LLRs) available at the output of the decoder. The rate-distortion (RD) performances obtained using this estimate of the BER are comparable to those obtained with the true error rate. However, estimation errors sometimes may lead to annoying local visual artifacts. The robustness of the estimation is thus further improved by coupling the error rate estimation to an error detection mechanism based on a cyclic redundancy checksum (CRC).

II. WYNER-ZIV CODING FRAMEWORK

The codec architecture considered is depicted in Fig. 1. The input sequence is structured into groups of frames (GOF), in which key frames are Intra-coded (using H.264/AVC Intra) and intermediate frames are WZ coded (called WZ frames). Each WZ frame is encoded independently of the other frames. The WZ data are transformed (using a 4x4 DCT) and organised in 16 subbands $WZ^j, j = \overline{1,16}$. Each DCT band is uniformly quantised using a number of levels, specified in the selected quantisation matrix. For a given rate point, one can associate a pair of parameters consisting of the QP parameter for the key-frames and the number of quantisation levels per band of the WZ frames. Each bit plane is fed to the turbo encoder. The systematic bits are discarded and only the parity bits of the turbo coder are stored in a buffer. Initially the encoder sends only a subset of the parity bits.

The work presented was developed within DISCOVER, a European Project (http://www.discoverdvc.org), funded under the European Commission IST FP6 program.

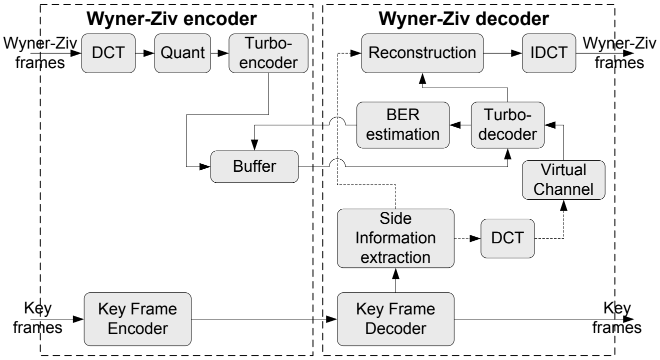


Fig. 1. The architecture of the Wyner-Ziv codec

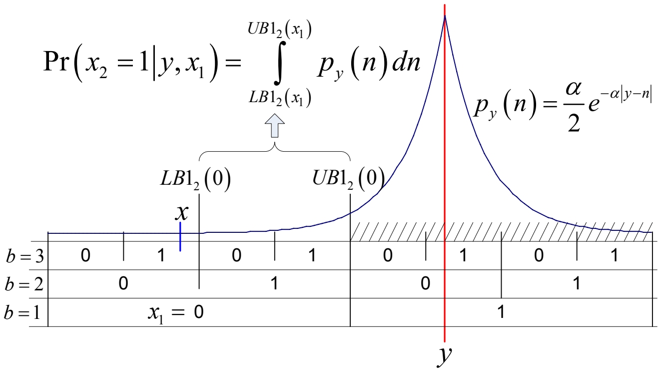


Fig. 2. Conditional probability of the bit x_2 given the SI y and the previous bit $x_1 = 0$

The decoder constructs the SI using motion compensated interpolation of the key frames as described in [7]. A Laplacian model is adopted for the virtual correlation channel, its variance being estimated from the residue obtained by motion compensating the key frames [8]. For each bit plane per band, upon reception of the first parity bits, the turbo decoder runs a log-MAP decoding algorithm. The decoding of the bit plane follows a hard decision based on LLRs; and the residual BER is estimated as the Hamming distance between the decoded and the original bit planes. If BER turns out to be large, e.g. $> 10^{-3}$, the decoder requests for more parity bits through the return channel. After turbo decoding, estimates of the quantised values are computed, given the decoded quantisation index and the SI.

III. MINIMUM RATE ESTIMATION AT THE ENCODER

In order to reduce the number of parity bit requests to be made by the decoder (which has a strong impact on decoding complexity), the encoder determines the minimum number of parity bits to be sent (R_{min}) per bit plane and per band. This minimum rate R_{min} is given by the conditional entropy $H(X|Y)$ between the data to be encoded and the SI. This conditional entropy depends on the crossover probability $p_{co} \equiv \Pr(X \neq Y)$ as follows:

$$H(X|Y) = -p_{co}\log_2 p_{co} - (1-p_{co})\log_2(1-p_{co}). \quad (1)$$

The correlation model, assumed to be Laplacian between the subband samples to be WZ-encoded and the corresponding samples in the SI, is supposed to be known to the encoder. The Laplacian parameter α estimated at the decoder can for example be transmitted periodically via the return channel. The quantity p_{co} is then estimated from the Laplacian correlation model as follows. The crossover happens in the b-th bit plane for a particular DCT coefficient x when the b-th bit x_b differs from the bit $\hat{x}_b$ which will be estimated at the decoder using the SI DCT coefficient y and the previous bits of x $(x_{b-1}, ..., x_2, x_1)$, i.e., when

$$x_b \neq \hat{x}_b \equiv \arg\max_{i=0,1} \Pr(i|y, x_{b-1}, ..., x_2, x_1), \quad (2)$$

where $\Pr(i|y, x_{b-1}, ..., x_2, x_1)$ is the *a posteriori* probability of the event $x_b = i$. The calculation of $\Pr(x_2 = 1|y, x_1)$ is illustrated in Fig. 2.

The crossover probability for the bit plane b is then obtained by taking the expectation of the probability of (2) (given the Laplacian distribution of Y), which is averaged over the whole band WZ^j as:

$$p_{co} = \frac{1}{N}\sum_{x\in WZ^j}\left[\int_{-\infty}^{+\infty}\Pr(x_b \neq \hat{x}_b)\frac{\alpha}{2}e^{-\alpha|y-x|}dy\right], \quad (3)$$

where N is the size of the band.

IV. CONFIDENCE MEASURE

This section describes a method to estimate the error rate at the output of the turbo decoder. The technique is based on a measure of confidence on the estimated bits derived from the likelihood ratio Λ_i:

$$\Lambda_i = \log_{10}\frac{Pr(X_i = 1|Y)}{Pr(X_i = 0|Y)} \quad (4)$$

where X_i is the bit decoded at the position i and Y is the SI value. The quantity $|\Lambda_i|$ is used as a confidence measure for a particular bit i: the bit i is considered uncertain if $|\Lambda_i|$ is below a certain threshold $\log_{10} T$. For example, putting $T = 99$ means that only bits with $\max(Pr(X_i = 1|Y), Pr(X_i = 0|Y)) \geq 0.99$ will be considered successfully decoded. To obtain the total confidence score for the bit plane, the number of bits with $|\Lambda_i| < \log_{10} T$ is computed, divided by the bit plane size:

$$ConfScore = \frac{count\{|\Lambda_i| < \log_{10} T\}}{N} \quad (5)$$

where N is the size of the bit plane. If the confidence score is more than 10^{-3}, the turbo decoder requests for more parity bits from the buffer as shown in Fig. 3.

V. CYCLIC REDUNDANCY CHECK

The stopping criterion based on the confidence score (5) performs well in terms of rate-distortion performance, showing almost no deterioration compared to the "ideal" error detection, involving the original sequence. However, localised visual artifacts (blocks having incorrect intensity values) can happen in the reconstructed video sequence. This is due to the fact that

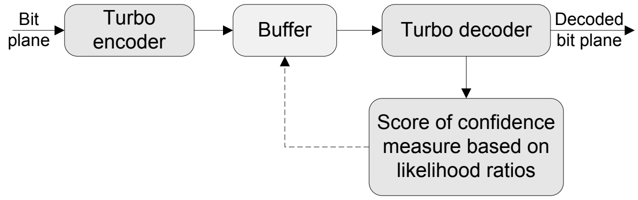


Fig. 3. Turbo encoder/decoder using confidence measure

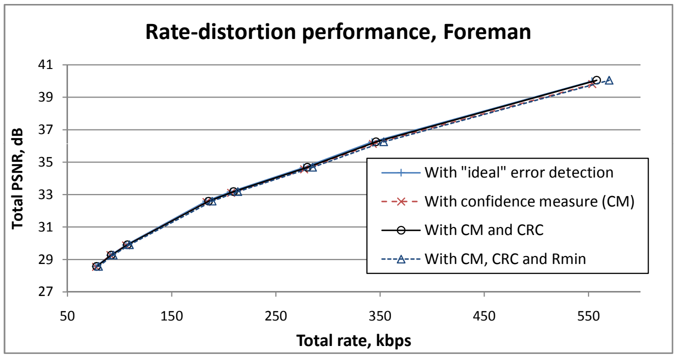
for some bit planes the actual BER happens to be higher than 10^{-3}. Since the likelihood ratio $|\Lambda_i|$ is computed using the probabilities depending on the estimated Laplacian distribution parameter α, it might not be always correct, thus some bits can have confidence value $|\Lambda_i|$ greater than the threshold $\log_{10} T$, while being incorrect. These incorrect bits lead to the incorrect reconstruction of quantisation bins for certain DCT coefficients, and finally they appear in the decoded frame as blocks having incorrect intensity values. The PSNR of the decoded sequence does not suffer much from these artifacts, because the under-estimation of the BER only happens on a limited number of frames and bit planes.

Setting T to a higher value, to try to avoid these visual artefacts, is not desirable because it will lead to unnecessarily increase in the number of parity bits requests. We propose to employ a cyclic redundancy checksum (CRC) to help the decoder to detect remaining errors in a bit plane. Transmitting a CRC of sufficient strength is a well-known error detection mechanism. It is computed as the reminder of the division of the polynomial, corresponding to the sequence of bits, by another polynomial chosen beforehand. A description of an algorithm computing CRC can be found in [9]. In the experiments reported in Section VI, an 8-bit CRC is computed and transmitted per bit plane. The decoder also computes a CRC on the decoded sequence. If BER $> 10^{-3}$, or the two CRCs do not match, it then requests more parity bits.

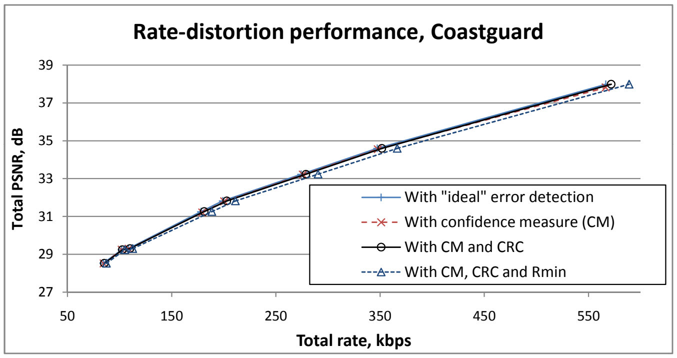
It is worth pointing out that, when the two CRCs match, the resulting BER is zero and the condition BER $> 10^{-3}$ seems to be unnecessary. But from the rate saving point of view, we assume that the CRC does not have sufficient length to be used alone without the confidence measure. That is why we keep both conditions, and CRC is actually used only in cases when the confidence measure described in the previous subsection under-estimates the error rate.

VI. EXPERIMENTAL RESULTS

The rate-distortion performance of the proposed method has been assessed on the entire Foreman and Coastguard sequences at QCIF/15 Hz (149 frames), with a GOF size of 2. The confidence level has been set to a reasonably high value 0.99 ($T = 99$). Our experiments with confidence measure without CRC showed very similar results for a wide range of T value. This can be explained by a so-called cliff-effect of turbo codes, resulting in a very rapid growth of the likelihood ratio Λ_i for the majority of bits' positions i as soon as the decoder has received enough parity bits for the bit plane. When the confidence measure is coupled with a CRC, the choice of T is not very important, as explained in Section V. So we kept the same value $T = 99$ for all evaluated methods. An 8-bit CRC has been computed with the standard polynomial $x^8 + x^2 + x + 1$ for each bit plane. For each rate point, a value of the quantisation parameter (QP) has been chosen for H.264 Intra-coding of the key frames. To each value of the QP parameter, we use a corresponding quantisation matrix for the different subbands resulting from the DCT transform of the Wyner-Ziv frames. These quantisation matrices are given in [5], and have been defined so that key-frames and Wyner-Ziv frames have comparable average quality. Fig. 4 shows the average PSNR of the Luminance component for both key frames and Wyner-Ziv frames versus the total bit-rate. It can be observed that, when using the proposed hybrid encoder/decoder rate control solution, there is almost no rate-distortion deterioration compared to the "ideal" error detection. A small rate augmentation in the case when minimum rate estimation is used, results from rare R_{min} over-estimations. These over-estimations can be explained by the fact that the exploited model for the correlation noise does not fit perfectly the real data. At the same time the decoder complexity is significantly reduced, which is confirmed by Fig. 5, showing the number of turbo-decoder requests. Fig. 6 shows the frame-by-frame PSNR of the decoded sequence at one RD point. The PSNR of certain frames is improved significantly (by up to 2dB) when using the CRC. The visual quality is also significantly better, the annoying local artefacts are removed.

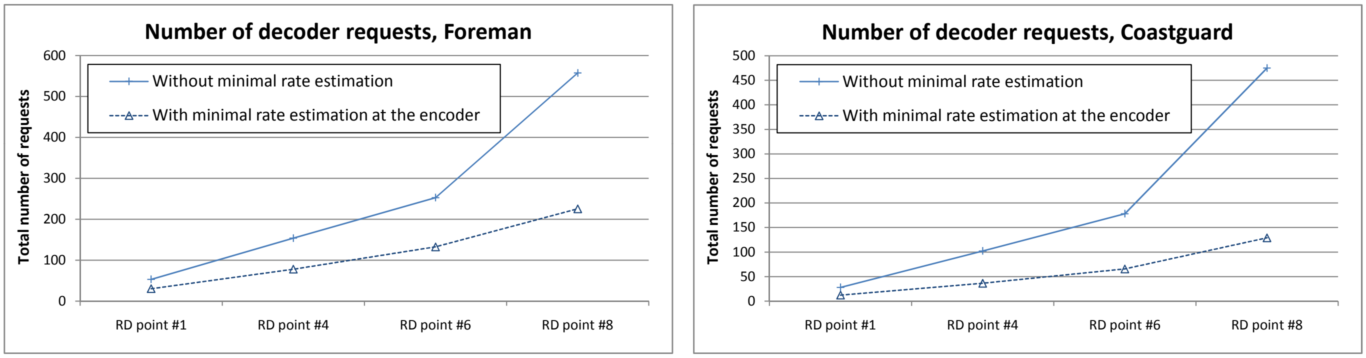


(a) Foreman sequence, QCIF/15 Hz, 149 frames



(b) Coastguard sequence, QCIF/15 Hz, 149 frames

Fig. 4. Rate-distortion performance with BER estimation versus true error rate.



(a) Foreman sequence, QCIF/15 Hz, 149 frames

(b) Coastguard sequence, QCIF/15 Hz, 149 frames

Fig. 5. Number of parity bit requests with and without minimum rate estimation.

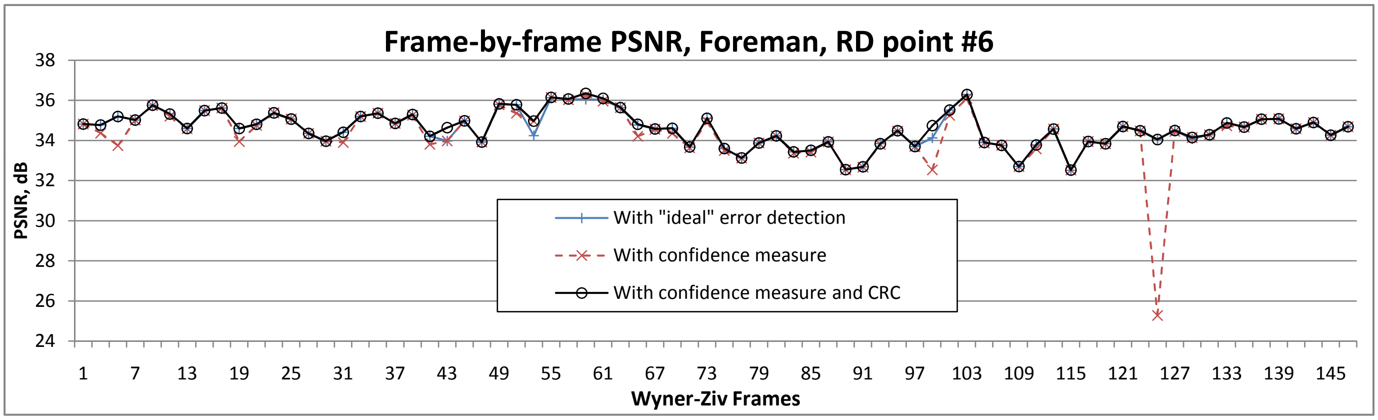


Fig. 6. Frame-by-frame PSNR of the decoded sequence.

VII. CONCLUSION

In this paper, a hybrid encoder/decoder rate control solution for DVC has been presented, including an approach for minimum rate estimation at the encoder side, as well as a robust method to estimate the BER at the output of the turbo decoder based on likelihood ratios. The former allows significant reduction of the decoder complexity, while the latter removes the main limitation of the current Wyner-Ziv video coding architecture, namely the knowledge of the original WZ frame at the decoder. The robustness is achieved by applying an error detection technique based on cyclic redundancy checksums. Experimental results confirm the effectiveness of the proposed method.

ACKNOWLEDGEMENTS

We are thankful to the IST and the Discover software development team for having provided the DVC codec software.

REFERENCES

[1] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. on Information Theory*, Vol. 19, No. 4, July 1973.

[2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Trans. on Information Theory*, Vol. 22, No. 1, January 1976.

[3] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes", *IEEE Comm. Letters*, Vol. 5, pp. 417-419, Oct. 2001.

[4] Da-ke He, En-hui Yang, "On the duality between Slepian-Wolf coding and channel coding", *Proc. of the 2006 IEEE Intl. Symp. Inform. Theory*, Seattle, WA, U.S.A., 2006.

[5] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv Codec for Video", *Proc. SPIE Conference on Visual Communication and Image Processing*, Jan. 2004.

[6] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed Video Coding", *Proc. IEEE, Special issue on advances in video coding and delivery*, vol. 93, No. 1, pp. 71-83, Jan. 2005.

[7] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation With Spatial Motion Smoothing For Pixel Domain Distributed Video Coding", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.

[8] C. Brites, J. Ascenso, F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding", *Int. Conf. on Image Processing (ICIP'2006)*, Atlanta, USA, October 2006.

[9] Ross N. Williams, "A painless guide to CRC error detection algorithms", 1993, http://www.repairfaq.org/filipg/LINK/F_crc_v3.html