PERFORMANCE EVALUATION OF DISTRIBUTED VIDEO CODING SCHEMES

R. Bernardini, R. Rinaldo, P. Zontone

DIEGM, University of Udine,
Via delle Scienze 208
33100 Udine - Italy
{bernardini,rinaldo,pamela.zontone}@uniud.it

A. Vitali

ST Microelectronics
Via C. Olivetti 2
20041 Agrate Brianza, Italy
andrea.vitali@st.com

ABSTRACT

Distributed Video Coding (DVC) has recently been proposed for emerging scenarios, where sources of correlated video do not communicate, as in some video-surveillance applications, or to simplify the coder for video equipment with power consumption constraints. In this paper we propose and compare different DVC schemes. In particular, we propose the use of the wavelet transform and analyze the performance of encoders based on turbo-codes and on a novel modulo-reduction procedure. The proposed schemes do not need feedback from the receiver and use statistical models for the estimation of the required bit-rate. Experimental results show that the proposed schemes have good performance when compared with similar asymmetric video compression schemes, and that DVC can be an interesting option in appropriate scenarios.

***Index Terms*—** Distributed video coding, Source coding,

1. INTRODUCTION

Distributed Source Coding (DSC) refers to the compression of multiple correlated sources that do not communicate with each other. These sources send their compressed outputs to a common decoder that performs joint decoding. In this situation the challenging problem is to achieve the same efficiency of a joint coder without requiring that the source encoders communicate with each other. Specifically, let $\{(X_i, Y_i)\}$ be a sequence of independent and identically distributed drawings of a pair of correlated discrete random variables X and Y. In [1], Slepian and Wolf showed that $R = H(X, Y)$, i.e., the minimum rate required when X and Y are coded *jointly*, is also sufficient when X and Y are coded in a distributed scenario, where the coders for X and Y operate *separately.* This remarkable result of Information Theory has been extended to lossy source coding by Wyner and Ziv in [2]. Recently, several practical schemes based on the distributed source coding principle have been proposed for video coding. As originally proposed by Wyner, these schemes are based on the use of channel codes [6, 7, 8]. The reference scheme of [9], with two separate coders and a joint decoder, assumes that the video sequence is divided into *key* frames (i.e., the odd frames of the sequence), and Wyner-Ziv (WZ) frames (the even frames). One coder codes the key frames without knowledge of the WZ frames and send them to the decoder. The decoder computes, by interpolating the key frames, a prediction of the WZ frames that will be used as *side-information* in the distributed coding paradigm. In [9], each bit plane of the pixels of the side information frames is modeled as a corrupted noisy version of the corresponding bit plane of the WZ frames. Thus, the WZ frames can be compressed, without access to the the key frames but assuming knowledge of the joint distribution, by sending only the parity bits of a systematic turbo coder, that the joint decoder will use to *correct* the bits in which the WZ and the side information bit plane differ (see Fig. 1). By modelling X^K as a length-K vector of i.i.d. binary random variables, and Y^K as the output of a memoryless binary symmetric channel with cross-over probability p and input X^K, optimal coding/decoding can be performed by using an ideal channel coder achieving capacity. In this case, the redundancy rate to recover X^K from Y^K, with vanishing error probability as K increases, would be equal to $H(p) = -p \log_2 p - (1-p) \log_2 (1-p) = H(X|Y)$ bit/symbol. The described distributed scheme would require $H(X, Y) = H(Y) + H(X|Y)$ bit/symbol, as predicted by the Slepian-Wolf result. The scheme described in [9] is extended

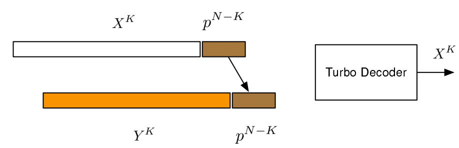

Fig. 1. DVC: 1. The WZ coder transmits the parity bits for the bit-plane X^K; 2. The decoder computes the side information, i.e., a noisy version Y^K of X^K, and recovers X^K using a systematic turbo decoder with input Y^K and the parity bits.

to the transform domain in [10], where the bit-planes of the Discrete Cosine Transform (DCT) of the WZ frames are compressed as described before. In [3], Pradhan and Ramchandran presented a syndrome based coding/decoding procedure, successively extended to a video coding system in [4].

In this paper we consider different approaches based on the distributed source coding paradigm. In particular, we propose the use of the wavelet transform, instead of the DCT transform of [10]. In [10], the Authors suppose that the decoder can request, via a feedback channel, additional bits until turbo-decoding gives correct reconstruction, within a small probability of error. In this paper, we consider a more realistic scenario where key-frames are actually coded, and where a CRC is added to the WZ bit stream, in order to recognize when the turbo-decoder reconstruction is successful. We also propose a procedure to estimate the required WZ bit rate in case no feedback channel can be used. Finally, we propose a novel DVC scheme which makes use of a modulo-reduction transformation before actual coding of the WZ frames. This scheme does not require any feedback channel and does not make use of turbo-codes. In the experimental section, these schemes are compared with intra-coding of the WZ frames and intra-coding with maximum likelihood estimation at the decoder.

2. WAVELET DOMAIN DVC

In the following, we will describe the procedures to code the WZ frames, since we suppose that the decoded key frames are available at the receiver. We consider two scenarios. In the first, the WZ frames are encoded independently of the key frames, and the key frames are encoded and decoded using a conventional intraframe codec. This is the original framework considered for Wyner-Ziv coding, e.g., in [9, 10]. In the second scenario, all frames (key frames and WZ frames) are available at the encoder. This scenario is interesting for the design of a low-complexity video coder, with no motion compensation, and where half of the frames (the WZ frames) are coded using distributed source coding techniques. This framework is considered, for example, in [5].

Wyner-Ziv Wavelet Domain. We propose a modification of the scheme presented in [10]. We operate on the Wavelet Transform of the WZ frames, with the advantage to possibly allow for multiresolution and bitstream scalability. At the encoder, the wavelet transform coefficients are grouped together to form coefficient subbands. Each subband is then quantized using a uniform quantizer and, for each subband, the corresponding bit-planes are independently coded using a *Rate Compatible Punctured Turbo* (RCPT) turbo coder. We transmit the parity bits relative to all the wavelet coefficients of each subband. At the decoder, the side information is generated from the key frames using a temporal interpolation based on Motion Compensated (MC) interpolation with symmetric motion vectors [12]. As explained before, the parity bits are used to recover the bit planes of the wavelet transform of the WZ frames from those of the side-information.

Puncturing allows to transmit a subset of the parity bits generated at the encoder. As mentioned before, in [10] a feedback channel is used to request additional parity bits until correct decoding is possible. In our implementation, instead, we consider the transmission of a 16 bit CRC code for each bit-plane. If the transmitted CRC does not match with the decoded bit-plane, the decoder requests additional parity bits from the encoder buffer until the reconstructed bit-plane matches the CRC and the decoding is declared to be successful.

As in [10], the iterative turbo decoder uses information about already decoded bit planes to improve a-priori knowledge while decoding the next bit plane. Moreover, since the WZ frames are typically quantized more coarsely than the key frames, the decoder implements a Maximum Likelihood reconstruction strategy, where the WZ wavelet coefficient is reconstructed as the value in the quantization interval which is closest to the value of the side-information.

Wyner-Ziv Wavelet Domain with rate estimation. This scheme does not use a feedback channel, which might not be available, e.g., in video streaming for delay sensitive applications. Since the decoder cannot make requests to the WZ encoder, it is necessary that the latter estimate the required parity bits for each wavelet coefficient bit plane. To this respect, at the encoder we need an estimate of the bit error probability between the WZ and the side information wavelet coefficient bit planes. Note that, in both scenarios we are considering[1], the side information is not available at the encoder.

Let us denote with $X_j(t)$ and $Y_j(t)$ the wavelet coefficient of subband j and frame t in the WZ and side information frames, respectively. We assume a Laplacian distribution for the difference $e_j(t) = X_j(t) - Y_j(t)$, namely,

$$f_{e_j(t)}(a) = \frac{\alpha_j(t)}{2} e^{-\alpha_j(t)|a|}, \tag{1}$$

and that $X_j(t)$ and $e_j(t)$ are independent. The parameter $\alpha_j(t)$ or, equivalently, the variance $\sigma^2_{j,t} = \mathrm{E}[(X_j(t) - Y_j(t))^2]$, is estimated as the mean squared error between $X_j(t)$ and an approximation $\tilde{Y}_j(t)$ of the side information. Specifically, in the first scenario, $\tilde{Y}_j(t)$ is the wavelet coefficient of the average of the WZ frames closest to the current frame and, in the second scenario, of the average of its closest key frames.

Using this model, we can calculate for a given quantizer the cross-over probability $p^b_j(t)$ for each bit-plane. As explained above, $R^b_j(t) = H(p^b_j(t))$ is a good estimate of the required WZ bit rate. The actual number of bits used in the experiments we present later, is computed on the basis of operational curves describing the performance of the turbo encoder/decoder for binary sources and a memoryless binary channel (see Fig. 2.)

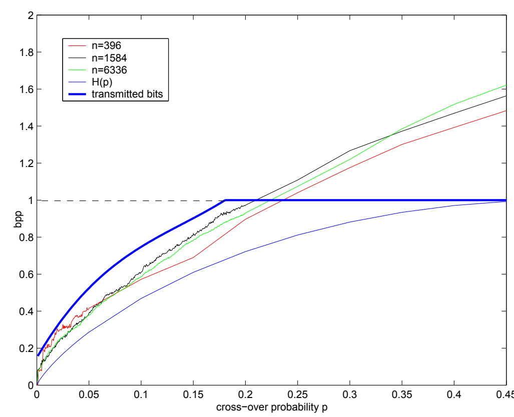

Fig. 2. Average bit rate (100 simulations) required by the turbo decoder, for a given cross-over probability p, and different codeword lengths. The lengths are chosen to match the actual subband dimensions considered in the experiments.

3. MODULO REDUCTION DVC

The scheme, shown in Fig. 3, is an improvement of the one presented in [13], in that the parameters are estimated from statistical models instead of depending on experimental constants. The scheme does not use turbo-codes and does not require feedback from the receiver. Let X and Y be the coefficients of the wavelet transform of the t-th WZ frame and key frame, respectively. The wavelet coefficients are processed by means of a modulo-reduction function ϕ_M shown in Fig. 4. After the reduction block, the coefficients are quantized with quantization step Δ and compressed by means of an efficient wavelet coder [11]. At the decoder side, the reduced and quantized $\overline{X}$ is reconstructed. As in Section 2, the side information is generated using temporal interpolation. From the knowledge of the output of the decoding process, the receiver deduces that the folded wavelet coefficient $\overline{X}$ belongs to $I = [\tilde{X}' - \Delta/2, \tilde{X}' + \Delta/2]$, for some X', and that X belongs to the inverse image $\phi_M^{-1}(I)$. X is reconstructed in a maximum likelihood (ML) sense by taking the value $\tilde{X} \in \phi_M^{-1}(I)$ closest to Y. Note that, disregarding quantization, exact recovery of X from $\overline{X}$ and Y is possible as soon as $|X - Y| < M/2$.

[1]Even if the WZ encoder can access the key frames, we do not want to locally calculate the side information, since this is a complex task, involving motion estimation, that can be done at the decoder, but not at the encoder.

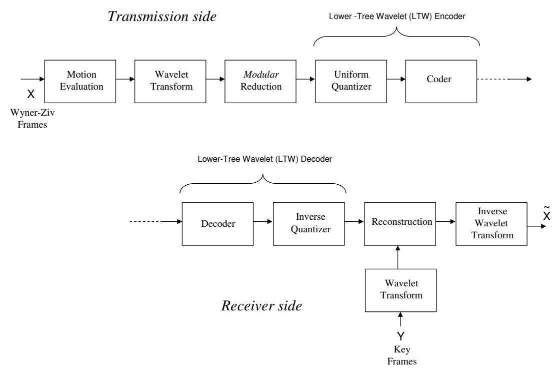

Fig. 3. Modulo Reduction DVC: The proposed scheme.

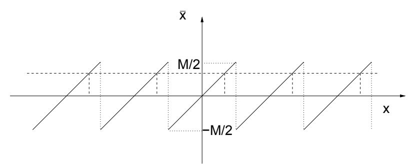

Fig. 4. Graph of the folding function associated with reduction modulo M.

The overall distortion at the decoder comprises two contributions, one caused by the quantization error, and the other caused by the large error one makes when $|X - Y| > M/2$. As before, we consider the model $Y = X + e$, with e independent from X and Laplacian distributed with parameter α. Again, this parameter can be estimated from sample averages as explained in Section 2. In the common hypothesis of small Δ, it is possible to show that the overall distortion between X and its reconstruction $\hat{X}$ can be approximated as

$$D \quad \simeq \quad \frac{\Delta^2}{12} + M^2 \frac{-2\,\beta \exp(-\alpha\, M)\,(\exp(-\alpha\, M) + 1)}{(\exp(-\alpha\, M) - 1)^3}. \quad (2)$$

The first term in Eq. 2 represents the contribution of the quantization error. The second term, instead, can be seen as the contribution of the *overload error* introduced by the modulo reduction procedure. In the experiments we present in the next section, the Motion Estimation module chooses M so that these two terms contribute equally to the overall distortion.

4. EXPERIMENTAL RESULTS

In this section we compare the performance of all the distributed coding schemes introduced in the previous sections. The key frames are encoded as intra-frames using a standard H.264/AVC coder. The turbo code is a Rate Compatible Punctured Turbo (RCPT) code with a puncturing period equal to 33. For the schemes where we use a Wavelet transform, this transform is computed using the well known 9/7 biorthogonal Daubechies filters, using a three level pyramid. In summary, we compare the performance (in scenarios 1 and 2) of the following schemes:

- Wyner-Ziv Wavelet Domain;
- Wyner-Ziv Wavelet Domain with rate estimation;
- Modulo-reduction;
- Intra coding of the WZ frames with ML decoding using the side information (joint decoding);
- Intra coding, no joint decoding. This is the simplest scheme, which does not make use of the distributed coding paradigm.

In scenario 2, we also consider the following scheme.

- Intra coding of the difference $X - X_{AV}$, where X_{AV} is the average of the key frames closest to the current frame.

We consider 299 frames of the QCif Foreman and 73 frames of the Teeny sequence, coded at 30 frames/s. Only the luminance component of the WZ frames is considered. In the first set of simulations, we set a quantization parameter QP, for the H.264 encoder, in order to have an average PSNR, for the key frames, close to 33 dB. Fig. 5 shows the rate - PSNR performance for the Foreman sequence (scenario 2) with all the considered schemes. As we can see, for scenario 2, Intra coding of the difference $X - X_{AV}$ with joint decoding performs better than the other schemes. Note that this scheme can not be used in scenario 1. The modulo reduction scheme and the Intra coding with joint decoding have similar results; in particular at high bitrates, they outperform the Wavelet Domain Wyner-Ziv scheme. At low bitrates, the Wavelet Domain Wyner-Ziv scheme with feedback has a slightly better performance than the other two schemes. However, note that it requires feedback from the receiver to send the right amount of WZ bits. Finally, the Intra coding scheme performs well when the rate is high. Similar results are obtained with the Teeny sequence (see Fig. 6), a very high motion sequence. Differently from the previous figure, in the case of the Teeny sequence, the Intra coding with joint decoding, outperforms the Modulo-reduction scheme and the Wyner-Ziv Wavelet domain scheme for all the bitrates. Moreover, after a given rate, the Modulo-reduction scheme clearly outperforms the Wyner-Ziv scheme. From the previous plots, we can also notice that the Wyner-Ziv scheme with rate estimation has low performance compared to the others schemes. This is due to the fact that we can not always transmit a sufficient number of parity bits to the turbo decoder and if this happens the quality of the reconstructed frames drops rapidly.

Finally, we consider coding of the Foreman sequence in scenario 1. As it can be seen from Fig. 7, the performance of the Wavelet Domain Wyner-Ziv, modulo reduction, and Intra coding with joint decoding schemes are very similar. The modulo reduction scheme has a slightly better performance than the Intra coding with joint decoding.

5. CONCLUSIONS

In this paper, we proposed and compared several schemes based on the recently proposed paradigm of Distributed Video Coding. DVC is an interesting option for emerging applications where geographically separated sources capture correlated video. Moreover, since the computational burden is transferred to the decoder, DVC can be applied in scenarios where the coders can be as simple as possible. Another practical advantage is that the paradigm could be more robust against transmission errors, since the decoder exploits statistical rather than deterministic correlation. We proposed a Wavelet Domain scheme with a rate estimation procedure to avoid feedback from the receiver. Moreover, we described a novel scheme, based on modulo reduction, which does not require any feedback and does

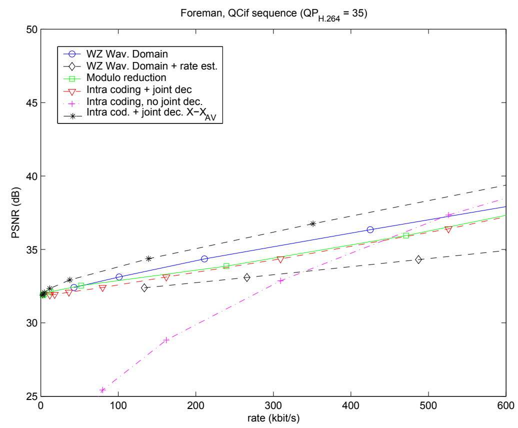

Fig. 5. Rate-PSNR performance for the *Foreman* sequence (scenario 2), the key frames are compressed using a QP = 35.

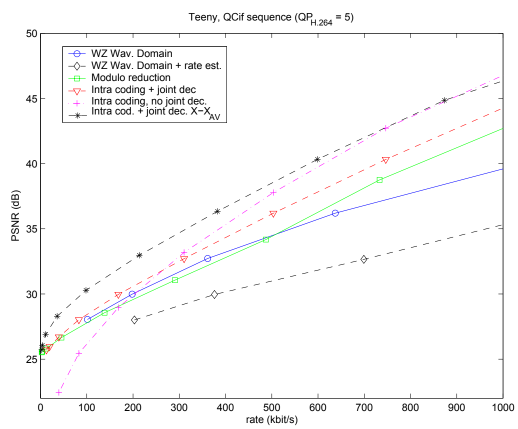

Fig. 6. Rate-PSNR performance for the *Teeny* sequence (scenario 2), the key frames are compressed using a QP = 5.

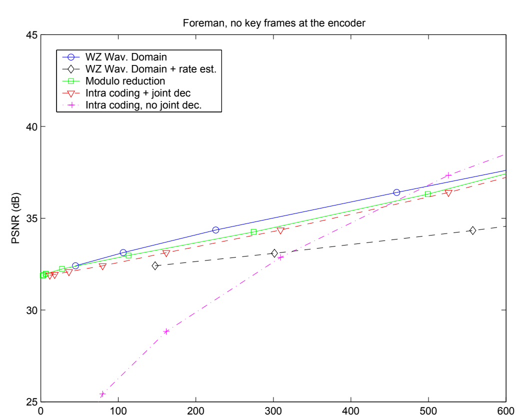

Fig. 7. Rate-PSNR performance for the *Foreman* sequence (scenario 1), the key frames are not accessible at the encoder and are compressed using a QP = 35.

not make use of turbo codes. Besides presenting several advantages, the proposed schemes have comparable or better performance than state-of-the-art DVC coders. Experiments show that DVC can currently provide some gain over more conventional techniques, at least for some sequences and bit rates.

6. REFERENCES

[1] D. Slepian and J.K. Wolf, “Noiseless coding of correlated information sources”, *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471-480, July 1973.

[2] A. 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, no. 1, pp. 1-10, Jan. 1976.

[3] S.S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction”, in *Proc. DCC ’99*, Snowbird, UT, Mar. 1999, pp. 158-167.

[4] R. Pury and K. Ramchandran, “PRISM: A new robust video coding architecture based on distributed compression principles”, in *Proc. Allerton Conf. on Communication, Control, and Computing*, Urbana-Champaign, IL, Oct. 2002.

[5] M. Tagliasacchi, A. Trapanese, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, “Intra Mode Decision Based on Spatio-temporal Cues in Pixel Domain Wyner-Ziv Video Coding”, *IEEE International Conference on Acoustic, Speech and Signal Processing*, Toulouse, May 2006.

[6] J. Garcia-Frias, “Compression of correlated binary sources using turbo codes”, *IEEE Commun. Lett.*, vol.5, no. 10, pp.417-419, Oct. 2001.

[7] J. Bajcsy and P. Mitran, “Coding for the Wyner-Ziv problem with turbo-like codes”, in *Proc. ISIT ’02*, Lausanne, Switzerland, July 2002, p.91.

[8] A. Aaron and B. Girod, “Compression with side information using turbo codes”, in *Proc. DCC ’02*, Snowbird, UT, Apr. 2002, pp.252-261.

[9] A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video”, in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, California, Nov. 2002.

[10] A. Aaron, S. Rane, E. Setton, and B. Girod, “Transform-domain Wyner-Ziv codec for video”, *Visual Communications and Image Processing*, VCIP-2004, San Jose, CA, Jan. 2004.

[11] J. Oliver, M.P. Malumbres, “Low-Complexity multiresolution image compression using wavelet lower trees”, *Proc. DCC ’03*, Snowbird, UT, pp. 133–142.

[12] D. Alfonso, D. Bagni, D. Moglia, “Bi-directionally motion-compensated frame-rate up-conversion for H.264/AVC decoder”, ELMAR Symposium, June 2005, Zadar, Croatia.

[13] R. Bernardini, R. Rinaldo, P. Zontone, A. Vitali, “Wavelet Domain Distributed Coding for Video”, *IEEE International Conference on Image Processing*, Atlanta, GA, October 2006.