MULTIPLE DESCRIPTION VIDEO CODING AND ITERATIVE DECODING OF LDPCA CODES WITH SIDE INFORMATION

Olivier Crave [1,2], *Christine Guillemot* [1], *and Béatrice Pesquet-Popescu* [2]

[1] IRISA / INRIA - Campus Universitaire de Beaulieu, 35042 Rennes Cedex, FRANCE

[2] TELECOM ParisTech, Signal and Image Proc. Dept. - 46, rue Barrault, 75634 Paris Cedex 13, FRANCE

ABSTRACT

In this paper, we propose the use of multiple description coding to increase the robustness of distributed video coding while keeping good rate-distortion performance. The video sequence is structured into key frames and Wyner-Ziv frames. For each type of frame, two descriptions are generated by a multiple description scalar quantizer and sent on a loss-prone channel. When both Wyner-Ziv descriptions are received, they are jointly decoded along with the side information. We investigate the influence of the amount of redundancy and of the iterative decoding of the descriptions on the performance.

Index Terms— Video codecs, Robustness, Quantization, Iterative methods, Codes.

1. INTRODUCTION

In the most widely used form of distributed video coding, the input sequence is structured into groups of picture (GOP), each one contains a certain number of key frames (KF) and Wyner-Ziv (WZ) frames (WZF). WZ coding [1] refers to lossy compression with side information (SI) at the decoder. The WZF are intraframe-encoded but interframe-decoded. In practice, a WZ encoder is obtained by adding a quantizer before a Slepian-Wolf (SW) encoder [2]. SW encoders individually compress the WZF but decode them with the help of a SI. The most commonly used practical SW encoders are based on channel coding principles. For each WZF, the decoder generates the SI by interpolation or extrapolation of previously decoded KF. The SW decoder combines the SI and the received parity bits to recover the WZ symbol stream. WZ coding enables low-complexity video encoding, while the complexity is shifted to the decoder.

Our goal is to improve the robustness of this type of coding structure while at the same time keeping good rate-distortion (RD) performance, a low-complexity at the encoder and preventing the introduction of quality fluctuations in the decoded video sequences along time. A few works investigated the use of distributed source techniques to increase the robustness of video transmission. In [3]-[4], WZ coding is used as a forward error correction mechanism. In [5] a WZ codec is added to a multiple description coding (MDC) scheme to combat packet loss. In our own work [6], a WZ codec is embedded into a temporal MDC scheme. The solution we propose here consists in generating two descriptions for each type of frame (KF and WZF) by a multiple description scalar quantization (MDSQ) [7]. This provides two balanced descriptions that contain partial information about WZF and KF. We only consider on-off channels (a description is either well received or fully lost). When the two descriptions are well decoded at the receiver, a maximum quality is achieved. When only one is received, the quality remains acceptable without introducing too much *flickering* if the redundancy is properly controlled. To exploit the redundancy between the descriptions, the latter scheme is further improved by jointly iteratively decoding the descriptions at the central decoder.

The remainder of this paper is organized as follows. In Section 2, we describe the video codec. In Sections 3 and 4, the iterative decoding of LDPCA codes and the optimal inverse quantization are detailed. In Section 5, we present experimental results, before drawing the concluding remarks in Section 6.

2. MULTIPLE DESCRIPTION VIDEO CODING WITH SIDE INFORMATION

A practical MDC scheme with common decoder-only SI was presented in [8] and tested on Gaussian sources. The proposed video codec is largely based on the same approach but, this time, the decoders have access to different SI. Diggavi et al. [9] have defined the RD region for this problem in the general case.

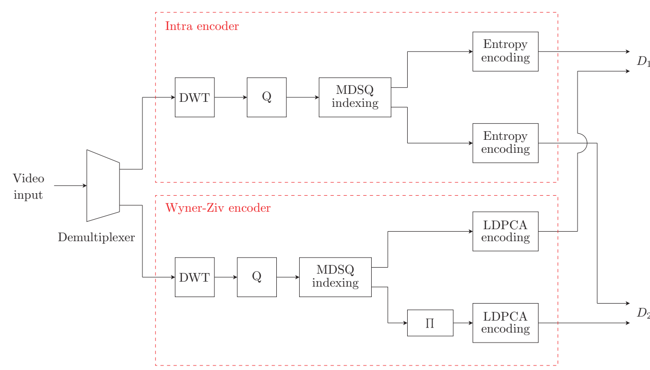

Fig. 1: Video encoder.

The implementation of the proposed video encoder is described in Fig. 1. The input video is first decomposed into two sequences, one that will be encoded as KF, and one that will be encoded as WZF. The KF are encoded using a conventional intra encoder (in our case, JPEG2000), which consists in a spatial discrete wavelet transform (DWT) followed by a MDSQ. After quantizing the source on a given alphabet, two indexes are assigned to the resulting discrete source symbols. This index assignment introduces redundancy or correlation in the transmitted streams in order to increase the robustness of the scheme. Finally, the two sequences of indexes are entropy coded using variable-length codes. The WZF go through a DWT, a uniform quantization and the resulting quantized coefficients are also mapped to a pair of indexes by the same index assignment as the one used for the KF. The two sequences of indexes are encoded bitplane per bitplane using a SW encoder based on low-density parity-check (LDPC) accumulate (LDPCA) codes, and only the produced accumulated syndromes are buffered into two descriptions. LDPCA codes were described in [10] as an efficient way of using LDPC codes in a rate-adaptive distributed source coding scheme. After the encoding is completed, one description of the KF is combined with one description of the WZF to obtain two descriptions that contain data for the entire sequence.

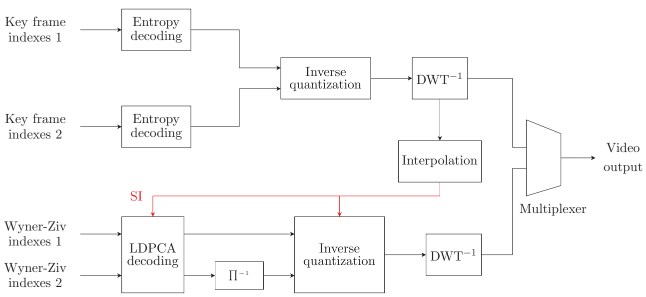

Fig. 2: Video decoder.

At the central decoder (see Fig. 2), the descriptions for the KF are entropy decoded separately and the two indexes are combined using a matrix lookup. The transform coefficients are reconstructed and inverse transformed to retrieve the KF. A SI is interpolated from the KF to decode the WZF with the accumulated syndrome bits transmitted in small amounts upon the decoder request via the feedback channel. In a first approach, the two WZ descriptions are decoded separately with the same SI. In a second approach, we propose the use of a cross-decoding algorithm described in Section 3 to perform a joint decoding of the two descriptions. After decoding the indexes, a simple matrix lookup is performed to get the quantized coefficients and the reconstruction is done with the help of the SI as described in Section 4. The coefficients are finally inverse transformed to retrieve the WZF which are combined with the previously decoded KF. The procedure is essentially the same at the side decoders: the received description for the KF is decoded and interpolated to decode the description for the WZF.

3. ITERATIVE DECODING OF LDPCA CODES

At the central decoder, in order to reduce the bitrate, the two descriptions can be jointly decoded to exploit the redundancy between them. This idea was first suggested in [11] in the case of turbo cross-decoding of two descriptions. We propose to generalize this approach to the case where an extra SI is available at the decoder and using LDPCA codes. The correlation between the descriptions comes from the index assignment matrix used by the MDSQ. Knowing this dependency, iterative decoding of two descriptions can be performed on LDPCA codes by exchanging the extrinsic information between the two decoders (see Fig. 3).

Let $\{X_n, n = 1, 2, ..., N\}$ denote the samples of a memoryless i.i.d source. This source is encoded at an average rate of r bits per sample (bps) per channel using a multiple description encoder, producing two correlated bitstreams, $\boldsymbol{u}^{(s)} = \{u_1^{(s)}, \ldots, u_{rN}^{(s)}\}, s = 1, 2$. Each bitstream is separately encoded using an LDPCA encoder. At the receivers, a bitstream of information bits is obtained from the SI, $\boldsymbol{y} = \{y_1, \ldots, y_N\}$.

The receiver separately decodes each bitplane, starting with the

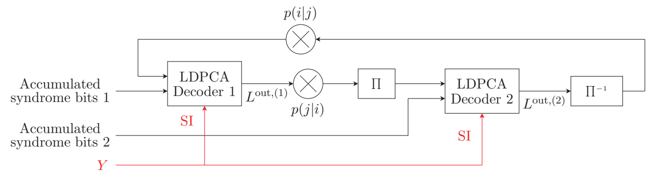

Fig. 3: LDPCA cross-decoding of two descriptions with side information.

most significant one. We only describe the transfer of information from the first decoder to the second decoder, the other being symmetric. When the estimates of the source bits of the first decoder converge, the log-likelihood ratio (LLR) $L^{\text{out},(1)}_{(k-1)r+t}$ is sent from the check nodes to the variable node $v^{(1)}_{(k-1)r+t}$ where $k = 1, \ldots, N$, $t = 1, \ldots, r$. This information can be seen as an extrinsic information that can be sent to the other description as an *a priori* information. The probability distribution for the bits that constitute the second description can be calculated from the bit probabilities,

$$\begin{aligned} & P(u^{(2)}_{(k-1)r+t} = 1) \\ = \quad & P(u^{(2)}_{(k-1)r+t} = 1 | u^{(1)}_{(k-1)r+t} = 1) \times P(u^{(1)}_{(k-1)r+t} = 1) \\ + \quad & P(u^{(2)}_{(k-1)r+t} = 1 | u^{(1)}_{(k-1)r+t} = 0) \times P(u^{(1)}_{(k-1)r+t} = 0) \end{aligned} \tag{1}$$

The samples being i.i.d, the conditional probabilities do not depend on k. Therefore, we can write, $\forall k \in \{1, \ldots, N\}$,

$$P(u^{(2)}_{(k-1)r+t} = 1 | u^{(1)}_{(k-1)r+t} = 1) = \sum_{\substack{l: b_t(l)=1 \\ m: b_t(m)=1}} P(j = m | i = l) \tag{2}$$

$$P(u^{(2)}_{(k-1)r+t} = 1 | u^{(1)}_{(k-1)r+t} = 0) = \sum_{\substack{l: b_t(l)=0 \\ m: b_t(m)=1}} P(j = m | i = l) \tag{3}$$

where $l \in \{1, \ldots, M\}$, $m \in \{1, \ldots, M\}$, M is the size of the alphabet of the indexes, and $\{b_t(l), t = 1, \ldots, r\}$ is the binary representation for the quantizer index l. Here, i and j are the row and column indexes of the index assignment matrix. The conditional probabilities are obtained from the index assignment matrix by using the distribution model of the quantized wavelet coefficients in the subbands. We suppose the distribution model to be Gaussian for the approximation subband and Laplacian for the detail subbands. The estimation parameters are determined for all subbands of the original image using the subbands of the SI. The LLRs for the second description are obtained from the bit probabilities

$$\begin{aligned} P(u^{(2)}_{(k-1)r+t} = 1) = & \\ & \sum_{\substack{l: b_t(l)=1 \\ m: b_t(m)=1}} P(j = m | i = l) \times P(u^{(1)}_{(k-1)r+t} = 1) \\ + & \sum_{\substack{l: b_t(l)=0 \\ m: b_t(m)=1}} P(j = m | i = l) \times P(u^{(1)}_{(k-1)r+t} = 0), \end{aligned} \tag{4}$$

$$P(u^{(2)}_{(k-1)r+t} = 0) = 1 - P(u^{(2)}_{(k-1)r+t} = 1), \tag{5}$$

$$L^{\text{in},(2)}_{(k-1)r+t} = \log \frac{P(u^{(2)}_{(k-1)r+t} = 1)}{P(u^{(2)}_{(k-1)r+t} = 0)} \tag{6}$$

These LLRs are used by the second decoder as *a priori* information. After convergence, the second decoder generates extrinsic log-likelihoods for the first decoder. An interleaver before the encoding of one of the descriptions is necessary to make sure that the information contained in one description is not redundant with the information contained in the other description for a given bitrate. The transfer of information back to the first decoder is carried out in a similar fashion. For a given bitrate for the accumulated syndrome bits, this cross-decoding is performed until the two descriptions are decoded or until the number of iterations reaches a certain threshold (the results shown in Section 5 were obtained for a threshold set to 18), in which case more accumulated syndrome bits are requested by the decoder.

4. OPTIMAL INVERSE QUANTIZATION

We consider the residue between the subbands of wavelet coefficients for the original image X and those that were obtained from the SI Y to be governed by a Laplacian distribution. The original model implies that $X = Y + U$, where the noise U has a Laplacian distribution with zero mean and variance $2/\alpha^2$. Since we are minimizing the mean-square error, the optimal estimate $\hat{x}_{opt}$ of the source x is given by

$$\hat{x}_{opt} = E[x | x \in \bigcup_{k=1}^{K} [z_i^k, z_{i+1}^k), y] \tag{7}$$

where $z_0 < z_1 < \ldots < z_N$ represent the quantization levels, N being the number of quantization levels. The number K of intervals of quantization for a given x depends on the number of received descriptions and the number of diagonals in the index assignment matrix. At the central decoder, $K = 1$. At the side decoders, K is the number of quantized coefficients in the row or in the column of the index assignment matrix that corresponds to the received index. Knowing the expression of the *pdf* for the correlation noise between x and y, we finally obtain:

$$\hat{x}_{opt} = \left(\sum_{k=1}^{K} g(k) \right) / \left(\sum_{k=1}^{K} h(k) \right) \tag{8}$$

where

$$g(k) = \begin{cases} \begin{array}{l} \left(\frac{1}{\alpha} + z_i^k\right) e^{\alpha(y - z_i^k)} \\ \quad - \left(\frac{1}{\alpha} + z_{i+1}^k\right) e^{\alpha(y - z_{i+1}^k)} \end{array} & \text{if } y < z_i^k, \\ \begin{array}{l} \left(\frac{1}{\alpha} - z_i^k\right) e^{-\alpha(y - z_i^k)} \\ \quad - \left(\frac{1}{\alpha} + z_{i+1}^k\right) e^{-\alpha(z_{i+1}^k - y)} \\ \quad +2y \end{array} & \text{if } y \in [z_i^k, z_{i+1}^k), \\ \begin{array}{l} \left(\frac{1}{\alpha} - z_i^k\right) e^{\alpha(z_i^k - y)} \\ \quad - \left(\frac{1}{\alpha} - z_{i+1}^k\right) e^{\alpha(z_{i+1}^k - y)} \end{array} & \text{if } y \geq z_{i+1}^k \end{cases}$$

and

$$h(k) = \begin{cases} e^{\alpha(y - z_i^k)} - e^{\alpha(y - z_{i+1}^k)} & \text{if } y < z_i^k, \\ \begin{array}{l} 2 - e^{-\alpha(y - z_i^k)} \\ \quad -e^{-\alpha(z_{i+1}^k - y)} \end{array} & \text{if } y \in [z_i^k, z_{i+1}^k), \\ e^{\alpha(z_{i+1}^k - y)} - e^{\alpha(z_i^k - y)} & \text{if } y \geq z_{i+1}^k. \end{cases}$$

5. EXPERIMENTAL RESULTS

For the experiments, we illustrate with the sequence Foreman (QCIF format at 15 Hz). The GOP size is set to two (even frames are encoded as KF and odd frames as WZF). The bitrates used for the KF are 114, 266, 418 and 532 kbs. The PSNR at the side decoders is calculated by averaging the PSNR obtained by both side decoders. The number of bitplanes per subband for the WZ indexes were set manually to ensure that the difference of quality between the decoded KF and the decoded WZF remains low at all decoders (the difference is always less than 1 dB at the central decoder and less than 2 dB at the side decoders). The KF were encoded using JPEG2000 (VM 5.2) in an embedded multiple description coding system as described in [12].

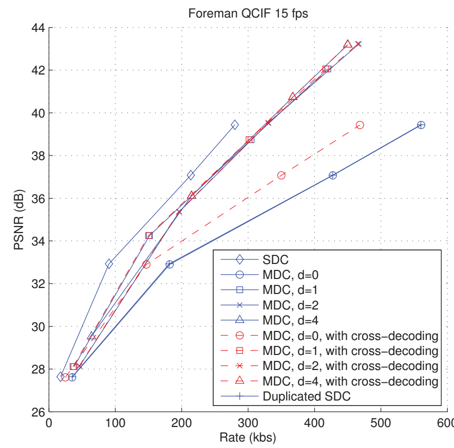

Fig. 4: Rate-distortion comparison of the SDC and MDC schemes for the WZF at the central decoder with and without cross-decoding.

Fig. 4 shows the central RD performance for the WZF with and without performing cross-decoding for comparison. The number of diagonals in the index assignment matrix is $2d + 1$. As expected, the MDC scheme with $d = 0$ performs as well as the single description coding (SDC) scheme with its rate doubled. The MDC schemes with $d = \{1, 2, 4\}$ perform much better and approach the quality achieved by the SDC scheme, but we do not see any clear difference between the three, which perform more or less equivalently. This shows that, when reducing the redundancy, we reach a point ($d = 1$) where less redundancy does not increase the central RD performance because the correlation between the SI and the WZF is too low, and that translates into an increase in the bitrate. The cross-decoding can offer up to 2 dB gain for $d = 0$ but does not improve the performance for higher values of d. One reason is that the cross-decoding information can be thought of as another SI that comes from the other description, and if one SI brings no more information than the other, then it does not contribute to reducing the bitrate and is useless. Another reason is the sub-optimality of the estimation that is needed to evaluate the distribution model of the coefficients in the subbands.

The PSNR results at the side decoders are shown in Fig. 5. As one can see for $d = 1$, with only a slight compromise in the robustness properties of the scheme, the RD performance at the central decoder almost catches up with the one obtained without using MDC and are still quite good at the side decoders compared to the case where the information is duplicated.

Figs. 6 and 7 show the RD performance for all the frames at the central and side decoders respectively. The RD performance progressively increases at the central decoder and decreases at the side decoders when d increases. The DISCOVER curves were obtained using the state of the art DCT-based distributed video coding scheme described in [13]. Its good performance is mainly due to the use of

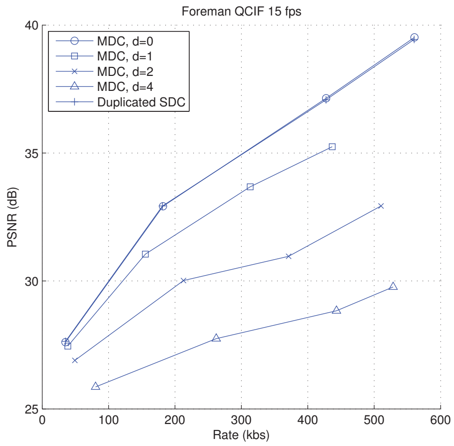

Fig. 5: Rate-distortion comparison of the SDC and MDC schemes for the WZF at the side decoders.

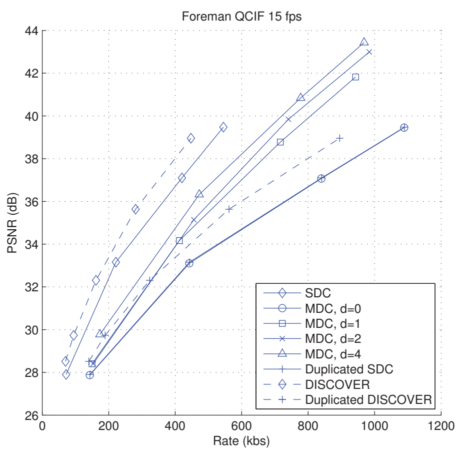

Fig. 6: Rate-distortion comparison of the SDC and MDC schemes for all the frames at the central decoder.

H.264/AV in intra mode which is known to be more efficient than JPEG2000 to encode QCIF frames.

6. CONCLUSION

We presented a robust balanced multiple description video coding scheme with SI where the descriptions are generated by a MDSQ and contain information about all the frames of the input sequence. The results for the WZ frames showed that, at the central decoder, when reducing the redundancy, the RD performance approaches the one obtained by a single description coding scheme until the correlation between the WZ descriptions and the SI becomes too low and does not allow any bitrate reduction. The cross-decoding of WZ descriptions has proven to lead to a rate reduction at the central decoder when the correlation between the descriptions is maximized, but has shown no benefit for lower correlation.

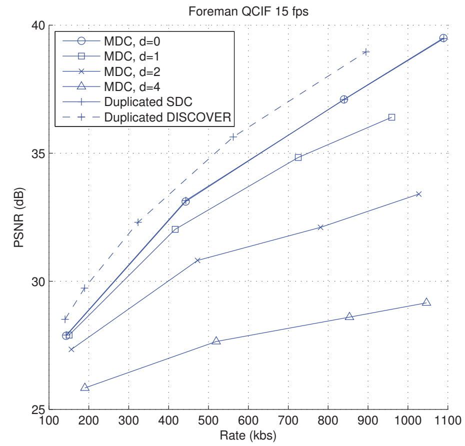

Fig. 7: Rate-distortion comparison of the SDC and MDC schemes for all the frames at the side decoders.

7. REFERENCES

[1] A. D. Wyner and J. Ziv, "The rate distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Th.*, vol. 22, pp. 1–10, January 1976.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Th.*, vol. 19(4), pp. 471–480, July 1973.

[3] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-ziv coding of video: An error-resilient compression framework," *IEEE Trans. Multi.*, vol. 6, no. 2, pp. 249–258, April 2004.

[4] S. Rane, A. Aaron, and B. Girod, "Systematic lossy forward error protection for error resilient digital video broadcasting," in *Proc. VCIP 2004*, 2004.

[5] M. Wu, A. Vetro, and C.C. Wen, "Multiple-description image coding with distributed source coding and side information," *Multimedia Systems and Applications VII*, vol. 5600, pp. 120–127, December 2004.

[6] O. Crave, C. Tillier, C. Guillemot, and B. Pesquet-Popescu, "Distributed temporal multiple description coding for robust video transmission," *EURASIP Journal on Wireless Communications and Networking*, January 2008.

[7] V. A. Vaishampayan, "Design of multiple-description scalar quantizers," *IEEE Trans. Inform. Th.*, vol. IT-39, no. 3, pp. 821–834, May 1993.

[8] O. Crave, C. Guillemot, and B. Pesquet-Popescu, "Multiple description source coding with side information," in *European Signal Processing Conference*, Lausanne, Switzerland, 2008.

[9] S. N. Diggavi and V. A. Vaishampayan, "On multiple description source coding with decoder side information," in *Information Theory Workshop*, San Diego, TX, USA, 2004, IEEE.

[10] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," in *Asilomar Conference on Signals, Systems, and Computers*, 2005.

[11] M. Srinivasan, "Iterative decoding of multiple descriptions," in *Data Compression Conference*, Snowbird, UT, USA, 1999, IEEE, pp. 463–472.

[12] T. Guionnet, C. Guillemot, and S. Pateux, "Embedded multiple description coding for progressive image transmission over unreliable channels," in *International Conference on Image Processing*, Thessaloniki, Greece, 2001, IEEE.

[13] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The discover codec: Architecture, techniques and evaluation," in *Picture Coding Symposium*, Lisboa, Portugal, 2007, IEEE.