ROBUST VIDEO TRANSMISSION BASED ON DISTRIBUTED MULTIPLE DESCRIPTION CODING

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

[1] IRISA / INRIA
Campus Universitaire de Beaulieu, 35042 Rennes Cedex, FRANCE
[2] GET / ENST, Signal and Image Proc. Dept.
46, rue Barrault, 75634 Paris Cedex 13, FRANCE

ABSTRACT

This paper proposes systematic lossy description coding for robust video transmission over error-prone channels. The problem of error propagation is addressed here by first structuring the data to be encoded into two descriptions. In a first approach, the two descriptions are constructed by splitting odd from even frames. Each description is separated into two sub-sequences, one being conventionally-encoded, the other one being coded with a Wyner-Ziv encoder. This amounts to having a systematic lossy Wyner-Ziv coding of every other frame of each description. This error control system can be used as an alternative to Automatic Repeat reQuest (ARQ) or Forward Error Correction (FEC), i.e. the additional bitstream can be systematically sent to the decoder or can be requested, similarly to an ARQ request. The amount of redundancy is mostly controlled by the quantization of the Wyner-Ziv data. This first approach leads to satisfactory lateral rate-distortion performance, however suffers from high redundancy which penalizes the central description. To cope with this problem, the approach is then extended to the use of motion-compensated temporal filtering (MCTF) for the Wyner-Ziv frames, in which case only the low-frequency subbands are WZ-coded and sent in the descriptions.

1. INTRODUCTION

Due to the real-time nature of envisioned data streams, multimedia delivery usually makes use of transport protocols, i.e. User Datagram Protocol (UDP) and/or Real-time Transport Protocol (RTP) which do not include control mechanisms which would guarantee a level of Quality of Service (QoS). The data transmitted may hence suffer from losses due to network failure or congestion. Traditional approaches to fight against losses mostly rely on the use of Automatic Repeat reQuest (ARQ) techniques and/or Forward Error Correction (FEC). ARQ offers to the application level a guaranteed data transport service. However, the delay induced by the retransmission of lost packets may not be appropriate for multimedia applications with delay constraints. FEC consists in sending redundant information along with the original information. The advantage of FEC is that there is no need for a feedback channel. However, if the channel degrades rapidly owing to fading or shadowing, or if the estimated probability of transmission errors is lower than the actual value, then the FEC parity information is not sufficient for error correction. Hence, the video quality may degrade rapidly, leading to the undesirable *cliff* effect.

Multiple description coding (MDC) has been recently considered for robust video transmission over lossy channels. Several correlated coded representations of the signal are created and transmitted on multiple channels. The problem addressed is how to achieve the best average rate-distortion (RD) performance when all the channels work, subject to constraints on the average distortion when only a subset of channels is correctly received.

Practical systems for generating descriptions that would best approach these theoretical bounds have also been designed considering the different components of compression systems : decorrelating transform and quantization.

Wyner-Ziv (WZ) coding can also be used as a forward error correction (FEC) mechanism. This idea has been initially suggested in [1] for analog transmission enhanced with WZ-encoded digital information. The analog version serves as side information (SI) to decode the output of the digital channel. This principle has been applied in [2], [3] to the problem of robust digital video transmission. The video sequence is first conventionally encoded, e.g., using an MPEG coder. The resulting bitstream constitutes the systematic part of the transmitted information which could be protected with classical FEC. Errors in parts of the bitstream, e.g. the temporal prediction residue in conventional predictive coding, may still lead to predictive mismatch and error propagation. The video sequence is in parallel WZ-encoded, and the corresponding data are transmitted to facilitate recovery from this predictive mismatch. The Wyner-Ziv data can be seen as extra coarser descriptions of the video sequence, which are redundant if there is no transmission error. The conventionally encoded stream is decoded and the corrupted data is reconstructed using error concealment techniques. The reconstructed signal is then used to generate the SI to decode the WZ-encoded data. However, error propagation in the MPEG-encoded stream may negatively impact the quality of the SI and degrades the RD performance of the system.

This problem is addressed here by first structuring the data to be encoded into two descriptions. In a first approach, the two descriptions are constructed by splitting odd from even frames. Each description is separated into two sub-sequences, one being conventionally-encoded, the other one being coded with a Wyner-Ziv encoder. This amounts to having a systematic lossy Wyner-Ziv coding of every other frame of each description. This error control system can be used as an alternative to ARQ or FEC, i.e. the additional bitstream can be systematically sent to the decoder or can be requested, similarly to an ARQ request, depending upon the existence of a return channel and/or the tolerance of the application to latency. The amount of redundancy is mostly controlled by the quantization of the Wyner-Ziv data. This first approach leads to satisfactory lateral RD performance, however suffers from high redundancy which penalizes the central description. To cope with this problem, the approach is then extended to the use of motion-compensated temporal filtering (MCTF) for the Wyner-Ziv frames, in which case only the low-frequency subbands are WZ-coded and sent in the descriptions.

The paper is organized as follows. Sections 2 and 3 describe the two schemes. Section 4 reports the simulation results of the proposed codecs. Conclusions are drawn in Section 5.

2. SYSTEMATIC LOSSY DESCRIPTION CODING IN PIXEL DOMAIN

We first consider the MDC coding architecture depicted in Fig. 1 (encoder) and 2 (decoder). At the encoder, the source is first divided into two sequences leading to two *non redundant* descriptions of the input sequence. The two descriptions are constructed by splitting odd from even frames as shown in Fig. 3. Every alternate frame (considered as *key frames*) of each description is conventionally encoded while the other frames are WZ-encoded. The sub-sequence of key frames is first temporally transformed using a 3-band motion-compensated temporal filter [4]. Each frequency band resulting from the temporal filtering is then encoded with an EZBC coder [5]. The remaining frames (Wyner-Ziv frames) are transformed (with an integer 4×4 block-based discrete cosine transform (DCT)) and quantized with a uniform scalar quantizer. The transformed coefficients are structured into spatial subbands and each bit-plane of the quantized subbands is then separately turbo-encoded. The resulting parity bits are punctured and transmitted. At the side decoders, the key frames are decompressed and the SI is generated by interpolating the intermediate frames from the key frames. The turbo decoder then corrects this SI using the parity bits. The parity sequences are stored in the buffer and transmitted in small amounts upon decoder request via the feedback channel [6].

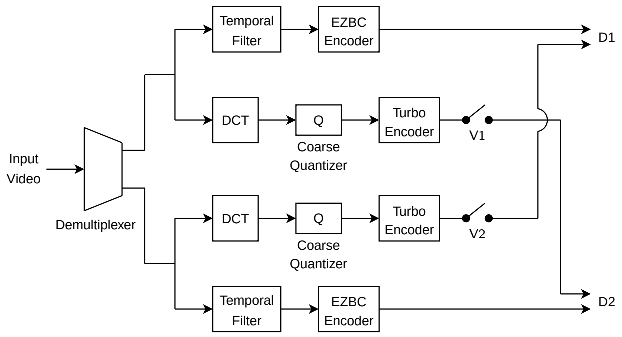

Figure 1: Implementation of the systematic lossy description encoder in pixel domain.

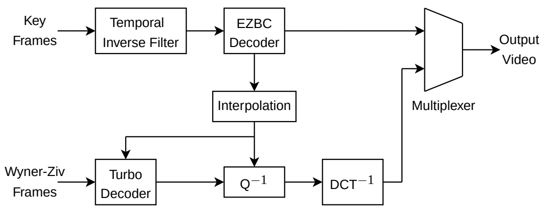

Figure 2: Implementation of the systematic lossy description side decoder in pixel domain.

Original sequence	0	1	2	3	4	5	6	7	8	9
Description 1	HF	WZ	BF	WZ	HF	WZ	HF	WZ	BF	WZ
Description 2	WZ	HF	WZ	HF	WZ	BF	WZ	HF	WZ	HF

Figure 3: The sequence is split into its even and odd frames. WZ symbols refer to WZ-encoded frames, HF and BF symbols refer to high- and low-frequency subbands.

The frames encoded as key frames in the first description are encoded as Wyner-Ziv frames in the second description and vice versa. Therefore, if both descriptions are received, the decoder so far only uses the key frames to reconstruct the sequence. On the other hand, if only one description is received, the decoder uses the Wyner-Ziv information in the received description to reconstruct the missing frames. The amount of redundancy is defined by the quantization of the Wyner-Ziv frames: The coarser the quantization, the higher the Wyner-Ziv bitrate. So far, when we use the scheme in a FEC scenario, the Wyner-Ziv streams are systematically sent and discarded at the central decoder. Further work will be dedicated to a possible use of the Wyner-Ziv bits even when both descriptions are received in order to improve the quality of the central decoder. In the ARQ scenario, the Wyner-Ziv streams are only sent if requested by the decoder. In the results reported later on, only the FEC scenario is considered.

It is important to notice that the Wyner-Ziv bitrate not only depends on the degree of quantization of the Wyner-Ziv frames but also on the quality of the SI, and therefore on the degree of quantization of the key frames.

3. SYSTEMATIC LOSSY DESCRIPTION CODING IN MCTF DOMAIN

To reduce the Wyner-Ziv bitrate and improve the RD performance of the central decoder, we propose a second architecture where the Wyner-Ziv frames are transformed by the same 3-band temporal filter than the one used for the key frames. Furthermore, before entering the Wyner-Ziv encoder, the subbands are lowpass filtered such that only the low-frequency subbands are WZ-encoded. The codec architecture is depicted in Fig. 4 (encoder) and 5 (decoder). For this codec, we consider the approach of separating the frames according to the Group Of Pictures (GOP) size of the temporal filter to obtain the two sub-sequences as shown in Fig. 6. At the side decoders, the SI is obtained by transforming the interpolated frames with a MCTF and only the resulting low frequencies are used as SI to decode the Wyner-Ziv subbands. The WZ-decoded low-frequency subbands are combined with the high-frequency subbands of the interpolated frames to get a sequence of subbands that is finally inverse filtered.

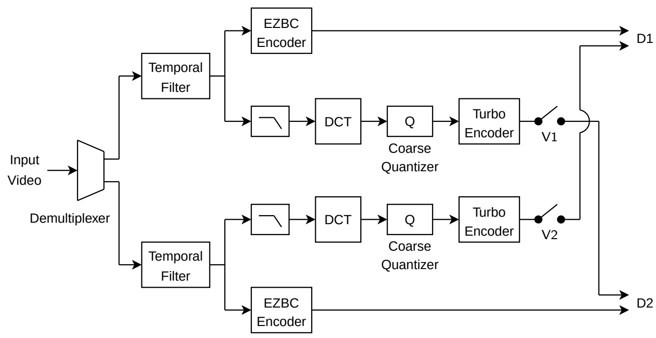

Figure 4: Implementation of the systematic lossy description encoder in MCTF domain.

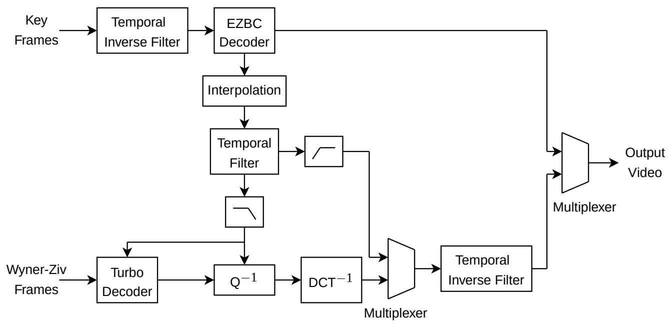

Figure 5: Implementation of the systematic lossy description side decoder in MCTF domain.

We will see in section 4 that since only the low frequencies are WZ-encoded, the RD performances at the central decoder are better than that of the previous scheme.

4. SIMULATION RESULTS

The results of the two schemes are compared with the 3-band MDC scheme described in [7]. In our simulations, we have chosen one level of temporal subband decomposition and the MCTF is performed using Hierarchical Variable Size Block Matching (HVSBM)

Original sequence	0	1	2	3	4	5	6	7	8	9
Non-redundant	HF	BF	HF	WZ	WZ	WZ	HF	BF	HF	WZ
Description 1	HF	BF	HF		WZ		HF	BF	HF	
Description 2		WZ		HF	BF	HF		WZ		HF

Figure 6: The sequence is split according to the GOP size of the 3-band temporal filter. WZ symbols refer to WZ-encoded frames, HF and BF symbols refer to high- and low-frequency subbands.

algorithm with block sizes varying from 64×64 to 4×4 and an $1/8$th pel accuracy.

The tests have been made for four rate-distortion points for the Wyner-Ziv bitrate corresponding to different amounts of redundancy. In the following, the various quantization indexes will be referred to as Q_i with $i = 1, \dots 4$. The higher is i, the higher are the bitrate and the quality.

The schemes of section 2 and 3 will be referred to as Scheme 1 and Scheme 2 respectively. For the experiments, we have considered the sequence Hall Monitor (QCIF format and 15 Hz) and Foreman (CIF format and 30 Hz). The bitrates used for the key frames are 80, 100, 150 and 200 kBit/s for Hall Monitor and 500, 1000, 2000 and 3000 kBit/s for Foreman.

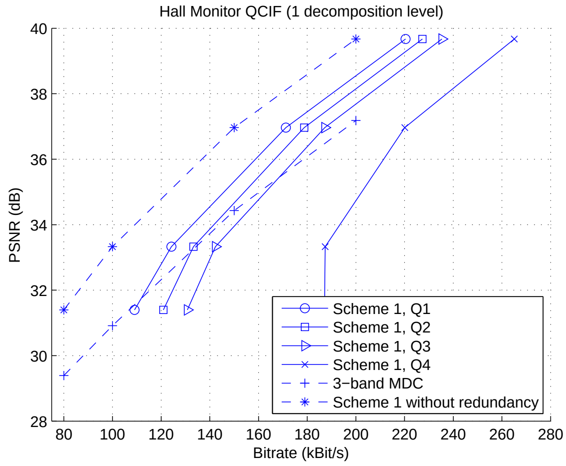

Figure 7: Central distortions of Scheme 1 compared with the 3-band MDC codec (Hall Monitor QCIF sequence, 15 Hz).

Fig. 7 shows the performances of Scheme 1 at the central decoder for Hall Monitor. The bitrate corresponds to the global rate (both descriptions). Scheme 1 without redundancy systematically outperforms the 3-band MDC scheme with at least 2 dB of improvement for 80 kBit/s. As expected, when we add some redundancy, the PSNR values decrease. But for quantification indexes lower than $Q4$, Scheme 1 still outperforms the 3-band MDC scheme for the highest bitrates. Fig. 8 shows the performances of Scheme 1 at the side decoder. The results look similar to the ones at the central decoder, except that the amplitude of the PSNR values has decreased. The two figures show that it is possible to use Scheme 1 with an additional Wyner-Ziv stream with a quality lower than $Q4$ and outperforms the 3-band MDC scheme for this type of fixed camera sequence.

Fig. 9 shows the performances of Scheme 1 at the central decoder for Foreman. One can see that the PSNR values for Scheme 1 without redundancy are comparable with the values of the 3-band MDC scheme. It follows that Scheme 1 with an added WZ-encoded stream always performs worst, and the higher is Q_i, the lower are the RD performances at the central decoder. Fig. 10 shows the performances of Scheme 1 at the side decoder. The 3-band MDC scheme outperforms Scheme 1 for all choices of Q_i. This time, Scheme 1

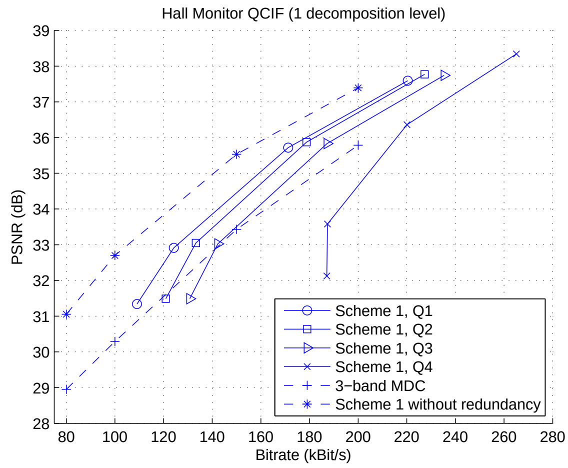

Figure 8: Lateral distortions of Scheme 1 compared with the 3-band MDC codec (Hall Monitor QCIF sequence, 15 Hz).

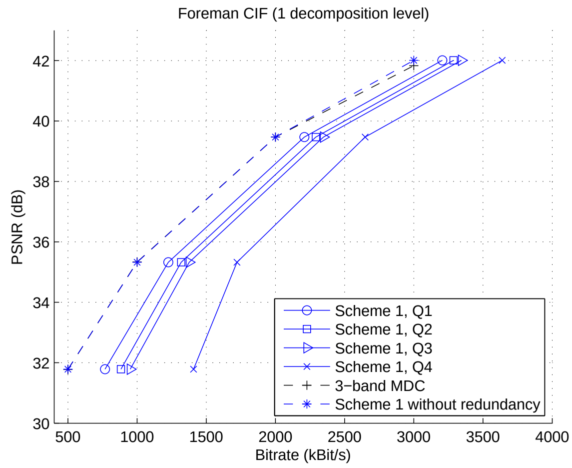

Figure 9: Central distortions of Scheme 1 compared with the 3-band MDC codec (Foreman CIF sequence, 30 Hz).

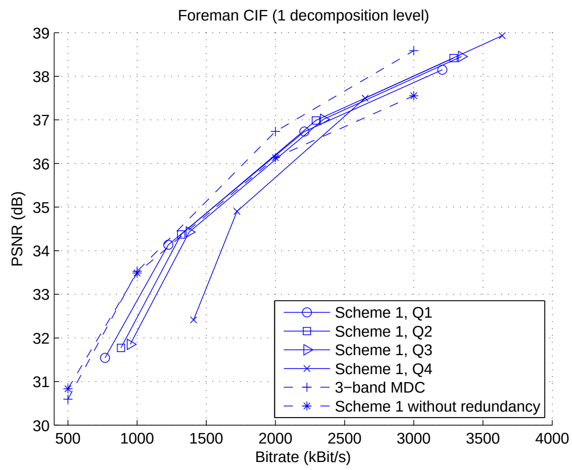

Figure 10: Lateral distortions of Scheme 1 compared with the 3-band MDC codec (Foreman CIF sequence, 30 Hz).

does not perform efficiently because of the high motion nature of the sequence.

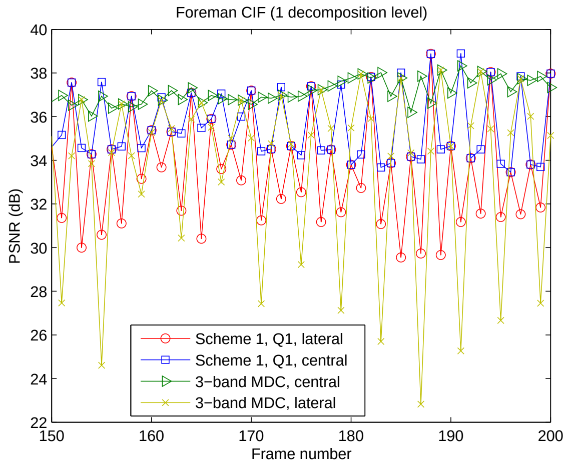

Figure 11: Central and lateral PSNR variation from the 150th to the 200th frame of the Foreman sequence (CIF, 30 Hz) at 1225 kBit/s.

We want to point out that it is not really fair to compare the schemes only with the mean PSNR (the average PSNR between the frames being received and the frames being lost and interpolated with or without extra information) because the fluctuations of PSNR between the frames is not taken into account. Fig. 11 shows the PSNR variation from the 150th to the 200th frame of the Foreman sequence at 1225 kBit/s for Scheme 1 using the quantization matrix Q_1 and the 3-band MDC scheme at the central and side decoder. At the side decoder, this figure shows that the PSNR values of the 3-band MDC scheme drop sharply (as low as 22.8 dB) when the missing frames are simply interpolated whereas it is more stable for Scheme 1 (the lowest value being 29.5 dB), even though the mean PSNR value is higher (+0.3 dB) for the 3-band MDC scheme than for Scheme 1. However, at the central decoder, the 3-band MDC scheme performs better and is more stable than Scheme 1, simply because the redundant data which is discarded in scheme 1 is used in the 3-band MDC scheme to reduce the quantization noise. In order to stabilize the quality at the central decoder, we could implement the new 3-band MCTF scheme described in [8].

Creating the two descriptions by splitting the sequence into an even and an odd sub-sequences makes the temporal filtering less efficient, the correlation between the frames is weak and it results in poor RD performances at the central decoder. Furthermore, by sending Wyner-Ziv data for all the frames of the sequence we end up with a totally redundant scheme. To solve this problem, we propose Scheme 2 where the frame splitting is done as in Fig. 6 and only the low-frequency subbands are WZ-encoded.

Fig. 12 and Fig. 15 show the performances of Scheme 2 at the central and side decoder for Hall Monitor. Scheme 2 always performs better than the 3-band MDC scheme for all the values of Q_i at both decoders, even with the finest quantization index $Q4$.

Fig. 14 shows the performances of Scheme 2 at the central decoder for Foreman. This time, Scheme 2 performs better than the 3-band MDC scheme for the first three values of Q_i, especially for the larger bitrates ($>$ 1000 kBit/s) where it can be as high as 0.8 dB above the 3-band MDC curve. Unfortunately, the RD performances at the side decoder (Fig. 15) are way below the ones of the 3-band MDC scheme.

Scheme 1 was also tested with the same frame splitting approach than Scheme 2. In that case, it performs better without redundancy than the 3-band MDC scheme at the central decoder. But the 3-band MDC scheme still outperforms Scheme 1 for most of the bitrates when a WZ-encoded stream is added to the descrip-

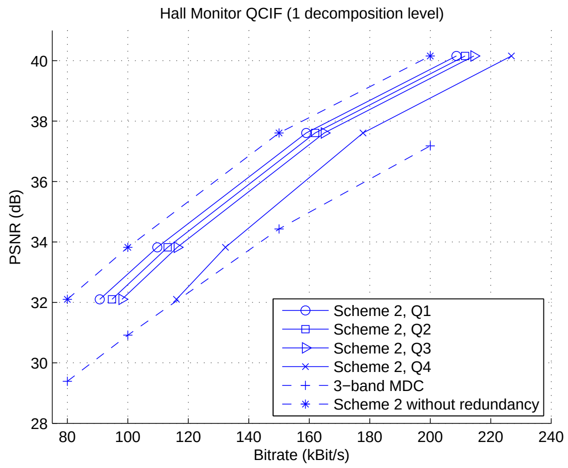

Figure 12: Central distortions of Scheme 2 compared with the 3-band MDC codec (Hall Monitor QCIF sequence, 15 Hz).

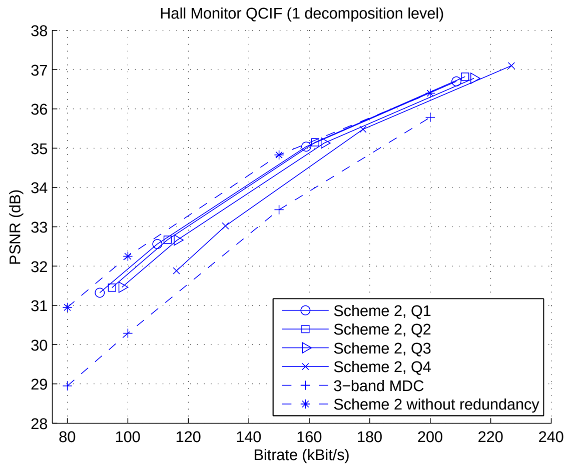

Figure 13: Lateral distortions of Scheme 2 compared with the 3-band MDC codec (Hall Monitor QCIF sequence, 15 Hz).

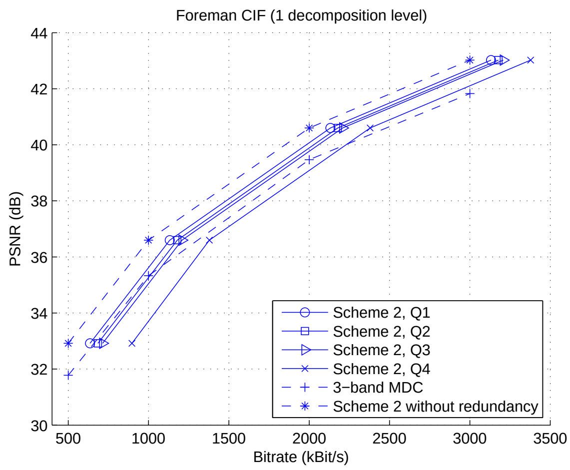

Figure 14: Central distortions of Scheme 2 compared with the 3-band MDC codec (Foreman CIF sequence, 30 Hz).

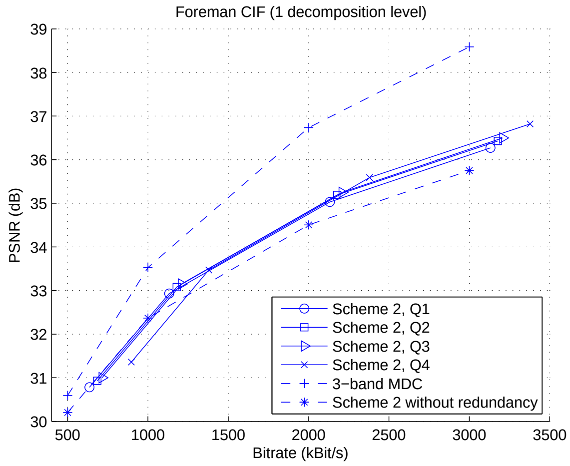

Figure 15: Lateral distortions of Scheme 2 compared with the 3-band MDC codec (Foreman CIF sequence, 30 Hz).

tions. The explanation is that even if the key frames are of a better quality, the distance between the key frames has increased, resulting in a bad quality SI. It follows that the turbo decoding requires more parity bits to reconstruct the Wyner-Ziv frames.

The proposed schemes were also tested with the sequence Flowergarden (CIF format and 30 Hz). For this sequence, Schemes 1 and 2 usually perform better than the 3-band MDC scheme, especially at higher bitrates. Flowergarden is not a fixed camera sequence, but the motion being a translation, the SI performs quite well and the Wyner-Ziv decoding does not require an important amount of parity bits to reconstruct the missing frames.

5. CONCLUSION AND FUTURE WORK

The results show that the RD performances highly depend on the sequence type. If the shooting was made with a fixed camera or if the sequence does not contain an important motion activity, the proposed schemes usually perform better than the 3-band MDC scheme. However, if the sequence contains an important motion activity, the performances of our schemes are usually worst in terms of average PSNR. But, if we take a look at the performances frame-by-frame for this type of sequence at the side decoders, we can see that the variation in quality between the frames is reduced, leading to fewer artifacts.

Our schemes are probably more robust to channel errors than the 3-band MDC scheme at the side decoders but the fact that the Wyner-Ziv information is discarded when both descriptions are received and does not contribute to any improvement in the central decoding quality is an important issue that we plan to solve in future research.

6. ACKNOWLEDGEMENT

The developments have been partly based on the distributed video coding software developed by the European Discover consortium which has been built upon the IST-TDWZ codec [9].

REFERENCES

[1] S. Shamai, S. Verdu, and R. Zamir, “Systematic lossy source/channel coding,” *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

[2] S. Rane, A. Aaron, and B. Girod, “Systematic Lossy Forward Error Protection for Error Resilient Digital Video Broadcasting,” in *Proc. VCIP 2004*, San Jose, CA, USA.

[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, Apr. 2004.

[4] C. Tillier and B. Pesquet-Popescu, “3D, 3-Band, 3-Tap Temporal Lifting For Scalable Video Coding,” in *Proc. ICIP 2003*, Barcelona, Spain, Sept. 14-17, 2003.

[5] 3D MC-EZBC software package, available in the MPEG CVS repository.

[6] A. Aaron and B. Girod, “Compression with side infomation using turbo codes,” in *Proc. DCC 2002*, Snowbird, UT, USA, pp. 252–261, Apr. 2002.

[7] C. Tillier, B. Pesquet-Popescu, and M. Van der Schaar, “Multiple Description Scalable Video Coding,” in *Proc. EUSIPCO 2004*, Vienna, Austria, Sept. 6-10, 2004.

[8] C. Tillier and B. Pesquet-Popescu, “A New 3-band MCTF Scheme For Scalable Video Coding,” in *Proc. PCS 2004*, San Francisco, CA, USA, Dec. 15-17, 2004.

[9] C. Brites, J. Ascenso and F. Pereira, “Improving Transform Domain Wyner-Ziv Video Coding Performance,” in *Proc. ICASSP 2006*, Toulouse, France, May 14-19, 2006.