Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec *

José Quintas Pedro[1], Luís Ducla Soares[2], Catarina Brites[1], João Ascenso[3], Fernando Pereira[1]
Carlos Bandeirinha[4], Shuiming Ye[4], Frédéric Dufaux[4], Touradj Ebrahimi[4]

[1] Instituto Superior Técnico – Instituto de Telecomunicações
[2] Instituto Superior de Ciências do Trabalho e da Empresa – Instituto de Telecomunicações
[3] Instituto Superior de Engenharia de Lisboa – Instituto de Telecomunicações
[4] Signal Processing Institute, Ecole Polytechnique Fédérale de Lausanne

ABSTRACT

Wyner-Ziv (WZ) video coding is an emerging video coding paradigm based on two major Information Theory results: the Slepian-Wolf and Wyner-Ziv theorems. One of the most interesting and used WZ video coding architectures makes use of a feedback channel (FC) to perform rate control at the decoder; in this context, the Slepian-Wolf coding module is typically based on turbo coding with puncturing. Because WZ coding is not based on the prediction loop used in conventional video coding but rather on a statistical approach where a decoder estimation of the frame to be coded is 'corrected' by the encoder, it provides intrinsic error resilience capabilities. This paper intends to study the error resilience performance of a feedback channel based transform domain WZ codec using appropriate scenarios and conditions, notably in comparison with the best performing H.264/AVC standard

1. INTRODUCTION

All the available ITU-T and ISO/IEC MPEG video coding standards rely on the powerful hybrid block-based motion compensation/DCT transform (MC/DCT) architecture which mostly targets applications where the video content is encoded once and decoded multiple times. In such scenarios, the video codec architecture is primarily driven by the one-to-many model of a single complex encoder and multiple light (cheap) decoders. The complexity burden of the encoder is mainly associated with the motion estimation and compensation tasks, which account for a major share of the coding gain in rate-distortion (RD) performance.

Distributed video coding, a new video coding paradigm, fits well in these scenarios, since it enables to explore the video statistics, partially or totally, at the decoder only, relying on a lower encoding complexity. One of the most interesting and used Wyner-Ziv video coding architectures [1] makes use of a feedback channel to perform rate control at the decoder. In this context, the Slepian-Wolf coding module is typically based on turbo coding with puncturing; the turbo coded bits (WZ bitstream) have the task to correct the estimation errors in the so-called side information (SI) which is an estimation made by the decoder of the original frame to be coded. Because this coding architecture always performs Intra coding in the sense that each frame is coded independently of its predecessors and successors, there is no prediction loop as it exists in all conventional video codecs. This characteristic brings to this coding approach intrinsic error resilience capabilities. The objective of this paper is to study in detail the error resilience performance of a FC based transform domain WZ codec for packet based networks.

This paper is organized as follows: Section 2 presents a brief review of the WZ video codec used in this paper; while Section 3 describes the error resilience evaluation scenarios to be used (standalone and comparative), Section 4 presents and discusses the error resilience performance evaluated under appropriate conditions. Finally, Section 5 summarizes the paper and proposes future work.

* The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org).

2. THE TRANSFORM DOMAIN WYNER-ZIV (TDWZ) VIDEO CODEC

The TDWZ video codec adopted in this paper [2], which architecture is presented in Figure 1 adopts an architecture inspired on the WZ coding architecture proposed in [1]; however, the algorithms used for the various codec modules are different [2][3].

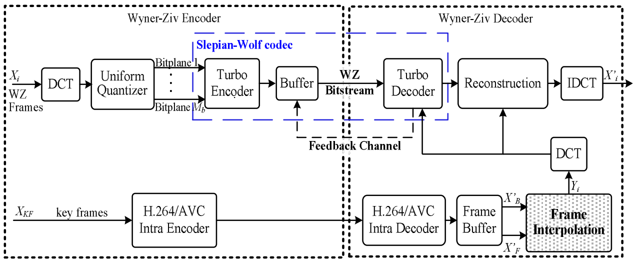


Figure 1 - TDWZ video codec architecture.

The TDWZ video codec works as follows: a video sequence is divided into WZ frames and key frames. The key frames are inserted periodically with a certain Group of Pictures (GOP) size; most results available in the literature use a GOP of 2 which mean that odd and even frames are key frames and WZ frames, respectively. While the key frames are traditionally intraframe coded, the WZ frames are DCT transformed, quantized with the quantization matrices defined in [2], representing each one a rate distortion point (RD). Turbo encoding is then performed and the parity bits are stored in the buffer and transmitted in small amounts upon decoder request via the FC. At the decoder, the frame interpolation module is used to generate the SI frame, an estimate of the WZ frame, based on previously decoded frames, X'_B and X'_F. For a GOP length of 2, X_B and X_F are the previous and the next temporally adjacent decoded key frames. The SI is then used by an iterative turbo decoder to obtain the decoded quantized symbol stream. The decoder requests for more parity bits from the encoder via the FC whenever the adopted request stopping criteria has not been fulfilled; otherwise, the bitplane turbo decoding task is considered successful. The SI is also used in the reconstruction module, together with the decoded quantized symbol stream, to help in the WZ frame reconstruction task. After all DCT coefficients bands are reconstructed, a block-based 4×4 inverse discrete cosine transform (IDCT) is performed and the reconstructed WZ frame is obtained. To finally get the decoded video sequence, decoded key frames and WZ frames are mixed conveniently.

3. ERROR RESILIENCE EVALUATION SCENARIOS

In order to evaluate the error resilience of the TDWZ video codec, its RD performance in the presence of channel errors will be first studied for different error conditions and after compared to the H.264/AVC RD performance, which represents the state-of-the-art in hybrid video coding. The performance evaluation will be carried out considering two different scenarios: i) Standalone TDWZ: errors in the two main components of the bitstream, key frames and WZ frames are considered; and ii) TDWZ is compared against the H.264/AVC codec. In all the scenarios, the FC used by the TDWZ decoder to ask for more parity information will be considered error free.

3.1 Scenario 1: TDWZ Standalone Evaluation

The objective of this first scenario is to evaluate the error resilience of the various parts of the TDWZ codec bitstream in the presence of channel errors, with a certain packet loss ratio (PLR). This includes evaluating the error resilience of the WZ part of the bitstream, as well as its dependence on reliable SI. For this evaluation, three separate experiments will be performed for a GOP of 2.

3.1.1 Only Key Frames Corrupted

In the first experiment in this scenario, errors will only be applied to the H.264/AVC part of the bitstream (i.e. the key frames), while the WZ part remains error free. This experiment allows evaluating the importance of having good quality SI at the decoder (which in this case is derived from corrupted key frames). To create a more error resilient stream, the H.264/AVC encoder exploits the available error resilient coding tools, notably flexible macroblock ordering (FMO), in this case using a checkerboard pattern. Since the H.264/AVC frames received at the decoder will be corrupted, error concealment in the key frames will be necessary in order to improve the SI quality used at the decoder. In this experiment, the error concealment included in the JM (version 11) software is used [4]. Because only Intra coding is performed, the H.264/AVC software Intra error concealment is used; this error concealment is based on spatial interpolation of the missing samples based on the adjacent blocks. In this case, each packet corresponds to a slice of 64 bytes [5].

3.1.2 Only WZ Frames Corrupted

In the second experiment, errors will be only applied to the WZ part of the bitstream. This leaves the H.264/AVC part of the transmitted bitstream error free, thus making it possible to see how much the decoded video quality drops, when the WZ frames are corrupted but the SI at the decoder is still intact. With this target, 3 cases are studied:

Case A) *Decoder rate control with protocol retransmission* - For this case, and since a FC is available, it is assumed that, whatever the packet size, the network protocol may ask for the retransmission of lost packets until they are correctly received. This implies the WZ frames quality is always the same for each RD point with an increase in the rate associated to the packet loss. More precisely, the rate for each RD point is computed as:

$$R_{PLR} = R_{Intra} + R_{WZ} \times \frac{1}{(1-PLR)} \quad (1)$$

where R_{Intra} and R_{WZ} is the error free rate for Intra and WZ frames, respectively and 1/(1-PLR) is the average number of transmissions for a PLR packet loss in the 0-1 range.

Case B) *Decoder rate control without protocol retransmission* - In this case, there is no network protocol to control the retransmission of the lost packets. If one packet is lost, the WZ decoder is fed with a full 0s packet and asks for another packet of WZ bits; this means that nothing happens in terms of turbo error correction and thus improvement of the decoded data. In this case, one packet corresponds to a chunk of WZ bits, i.e. to the bits sent for each WZ decoder request.

Case C) *Ideal encoder rate control* - In this case, it is supposed that the encoder performs ideal rate control, i.e. the bit rate needed for successfully decoding each bitplane for the error free case is determined *a priori* and used for decoding the corresponding corrupted bitstream. Furthermore, if the bit error probability of the decoded bitplane is higher than 10^{-3}, because some packets are lost, the decoder uses the corresponding bitplane of the side information,since no additional parity bits are requested in this case. The header of the WZ bitstream, which contains critical information such as image size, quantization parameters, and Intra period, is assumed to be received correctly.

3.1.3 Both Key frames and WZ Frames Corrupted

Finally, a third experiment will be performed, where errors are applied to both the H.264/AVC and WZ frames. In this case, the same error concealment as described in 3.1.1 will be applied for the key frames. Also the same packet sizes will be used.

3.2 Scenario 2: TDWZ Comparative Evaluation

In this scenario, the error resilience performance of the TDWZ and H.264/AVC codecs will be compared. The three cases for WZ frames error corruption described in 3.1.2 will be considered. This RD performance comparison would need to be complemented by a complexity comparison which is not the topic of this paper.

4. ERROR RESILIENCE PERFORMANCE EVALUATION

This section evaluates the error resilience performance of the adopted TDWZ codec, notably with video sequences which represent different types of video content.

4.1 Test Conditions

The tests carried out used two sequences, notably Foreman (with Siemens logo) and Hall Monitor. All frames were tested for both sequences, which mean 149 frames for Foreman and 165 frames for Hall Monitor; both sequences were tested at a temporal resolution of 15 Hz and a QCIF spatial resolution. The TDWZ video codec was configured to use GOP length of 2. Four quantization matrices have been used which means there are four RD points for each evaluation case; key frames were encoded using the H.264/AVC Extended Profile. Only luminance data has been coded.

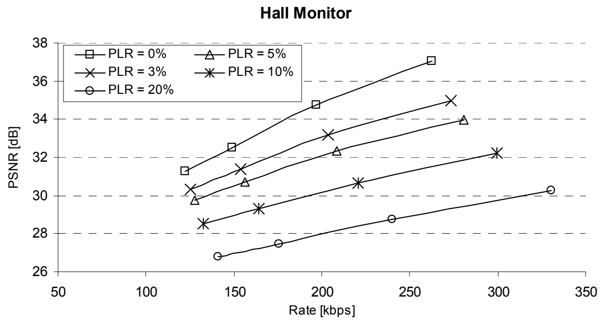
For this error resilience study, a communication channel that is characterized by the error patterns provided in [6] with different PLR is considered. These patterns are those also used in JVT. The testing sequences were corrupted with packet loss rates of 3%, 5%, 10%, or 20%, and each corrupted sequence was compared with its corresponding error free sequence (PLR=0%).

For the various PLR and RD points, the average (in time) PSNR will be measured and also averaged using 10 error pattern runs (with same PLR but different patterns).

4.2 TDWZ Standalone Evaluation

4.2.1 Only Key Frames Corrupted

For the Hall Monitor and Foreman sequences, the overall TDWZ RD performance after key frames corruption is presented in Figure 2 for the overall set of frames (WZ frames are error free).



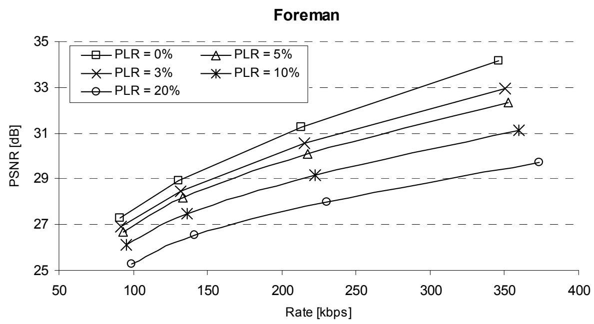


Figure 2– TDWZ RD performance for key frames corruption for Hall Monitor and Foreman.

It is important to notice that, for the various RD points, the parity bits sent do not correspond to all bitplanes of all DCT coefficients of the WZ frame; this means that for the bitplanes which do not have parity bits available at the decoder, the bitplanes from the side information are always used, which may have additional errors introduced by the transmission channel, explaining the different final quality for each RD point (see Figure 2). The bitplanes, for which parity bits were sent, will always have an error probability less than 10^{-3}.

4.2.2 Only WZ Frames Corrupted

Case A) The RD performance for this case is depicted in Figure 3, for the Foreman and Hall Monitor sequences.

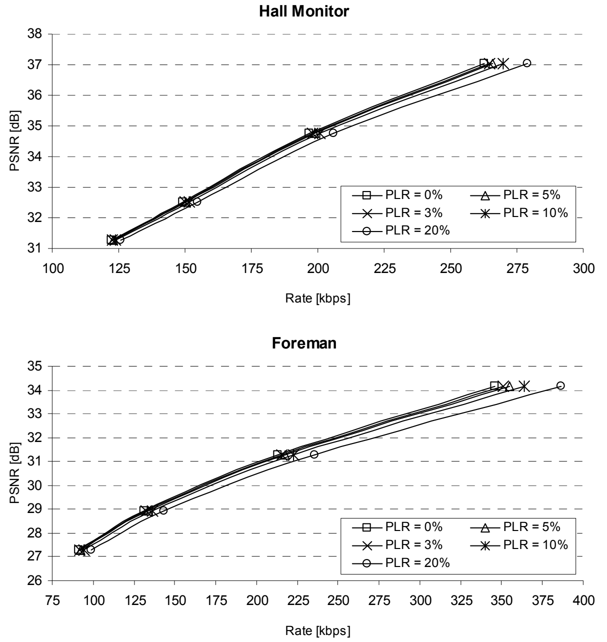


Figure 3 - RD performance for WZ frames corruption (case A) for Hall Monitor and Foreman.

The final quality is the same for each PLR because the lost packets are requested again and retransmitted by the encoder, meaning that all the needed parity bits are always received even if after some retransmissions. In this case, a bitrate loss is observed when the PLR≠0, with an increase expressed by (1).

Case B) In this case, the RD performance (shown in Figure 4) is similar to case A) although the received bit chunks are not the same but just the same amount since there are no retransmissions here. This shows that if a certain bitrate is used it is not relevant which precise bit chunks (packets) are lost or used, since the turbo decoder will always converge after decoding the same amount of chunks/packets. This type of behaviour is not observed in the hybrid video coding schemes where some information is more important than other (e.g. motion vectors); a clear advantage of the proposed TDWZ codec.

Case C) The RD plots in Figure 5 show the RD performance at various PLRs for the sequences Hall Monitor and Foreman, respectively. These results show that the TDWZ codec can compensate quite well channel errors in WZ frames. In particular, the performance loss at the 3% loss rate is very low for all the considered RD points, for the two test sequences. At higher loss rates (10% and 20%), the quality loss is still lower than 1dB at low bitrates, and it only increases slightly at higher bitrates. To evaluate how the packet size affects the RD performance of the codec, different packet sizes (64, 128, 512, and 1024 bytes) with different RD point have been tested. The results show that, for different packet sizes, the RD performance is basically the same.

4.1.3 Both Key Frames and WZ Frames Corrupted

Case A) The results are shown in Figure 6; as observed, the final decoded quality is the same as for case A when only key frames are corrupted, but there is a rate increase caused by the corruption of the WZ bitstream, which will always reach the decoder regardless the erroneous chunks in WZ bitstream. The quality degradation across PLR is due to the degradation of the quality of the key frames, as more errors are introduced, which in turn produces SI with less quality, meaning that the bitplanes for which parity bits were not sent would be with errors and it is not possible to correct them.

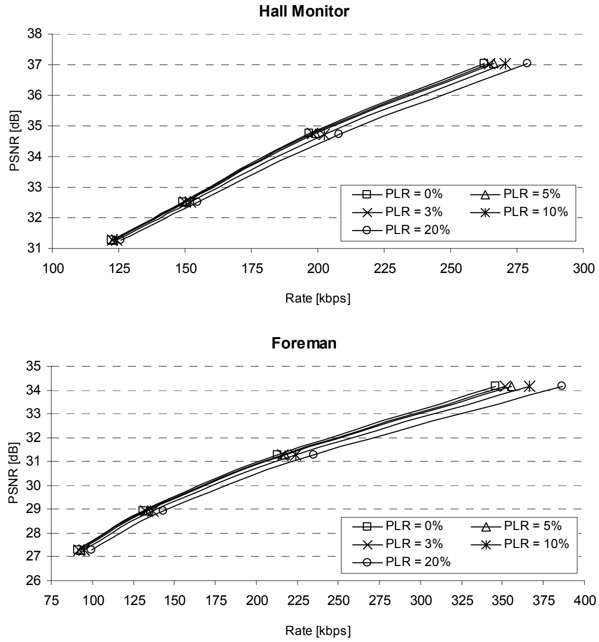


Figure 4 - RD performance for WZ frames corruption (case B) for Hall Monitor and Foreman.

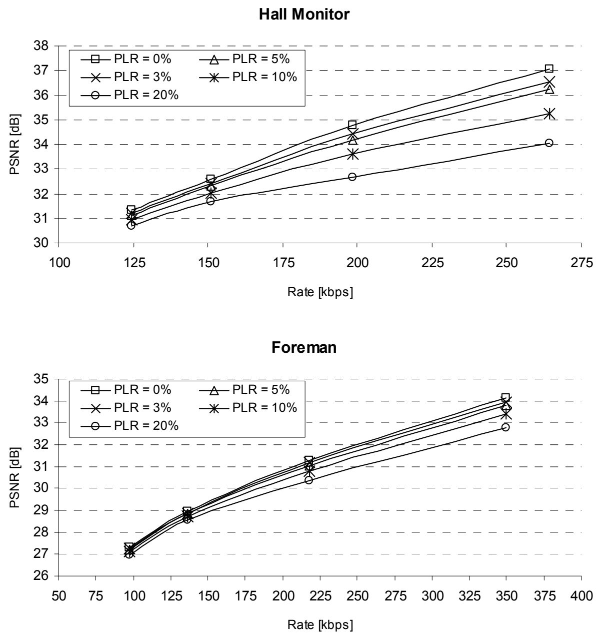


Figure 5 - RD performance for WZ frames corruption (case C) for Hall Monitor and Foreman.

Case B) The results for case B are presented in Figure 7; they are similar to the results for case A when both frame types are corrupted (as expected). The PSNR degradation due to the corruption of the key frames is the same as for case A which means that if a certain parity rate is used it is rather irrelevant which precise chunks are received.

Case C) In this case, both the key frames and WZ frames were corrupted with packet loss rates of 3%, 5%, 10%, or 20%; Figure 8 shows the RD performance in these conditions for Hall Monitor and Foreman sequences. The distortions in Figure 8 are much larger than those caused by packet loss only in WZ frames (see Figure 5). The reason is very likely related to the fact that the packet losses in key frames also affects the SI generated based on these key frames, which would then decrease the quality of the decoded WZ frames.

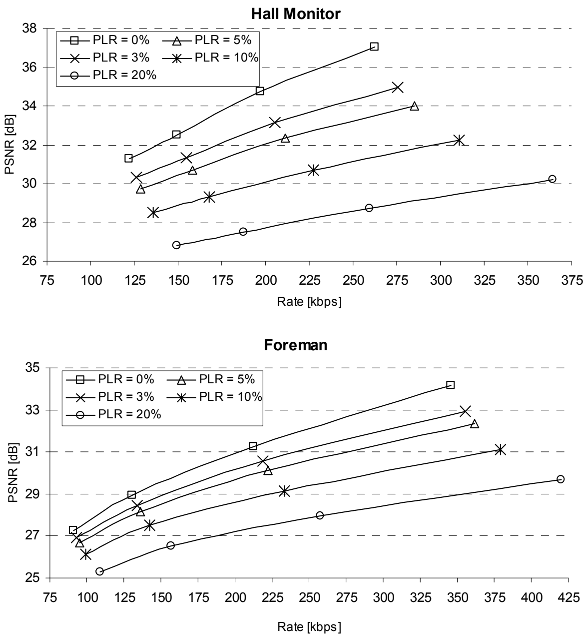


Figure 6 - RD performance for key frames and WZ frames corruption (case A) for Hall Monitor and Foreman.

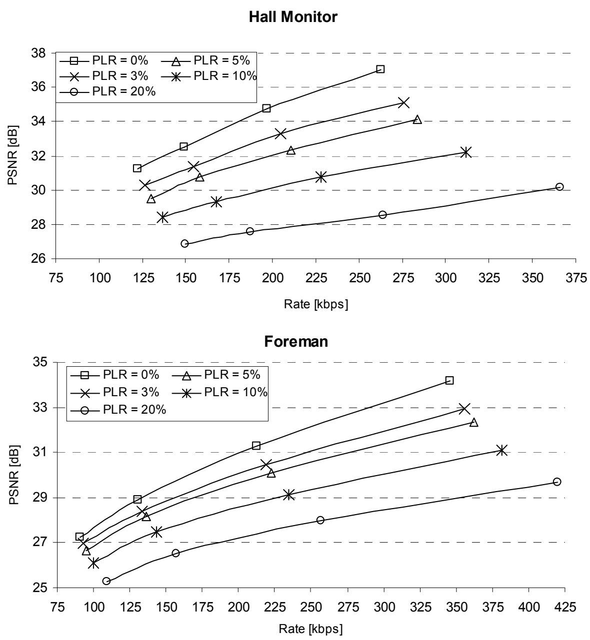


Figure 7- RD performance for key frames and WZ frames corruption (case B) for Hall Monitor and Foreman.

4.3 TDWZ Comparative Evaluation

The comparison between the TDWZ and H.264/AVC codecs is presented in Figure 9, considering the most realistic case B), when both key frames and WZ frames are corrupted. The H.264/AVC codec uses a GOP size of 2 with a IBIB… coding structure. It can be seen that for the error free case the H.264/AVC RD performance is clearly better than the TDWZ codec RD performance. However, in the presence of errors, TDWZ behaves better being more robust than H.264/AVC, confirming its intrinsic error resilience capabilities.

5. FINAL REMARKS

This paper provides a study for the error resilience performance of a FC based transform domain Wyner-Ziv codec. The results confirm the intrinsic error resilience capability of the TDWZ codec, mostly due to the usage of turbo coding. The most interesting result is the fact that, for the adopted test conditions, the TDWZ codec performed better than the H.264/AVC considering an error prone channel.

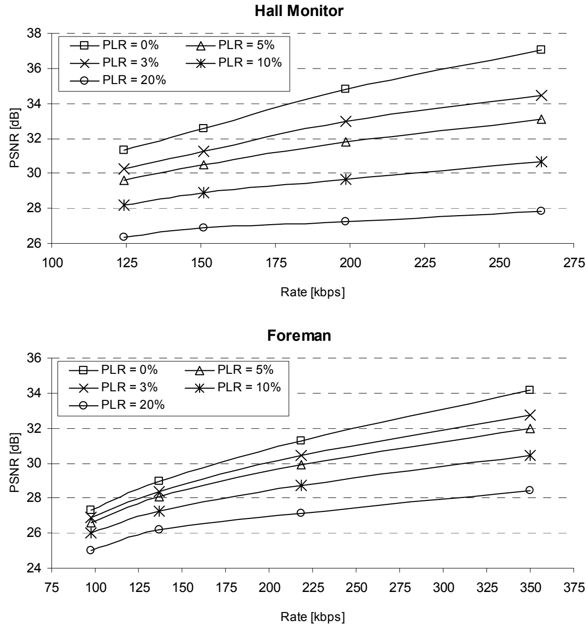


Figure 8 - RD performance for key frames and WZ frames corruption (case C) for Hall Monitor and Foreman.

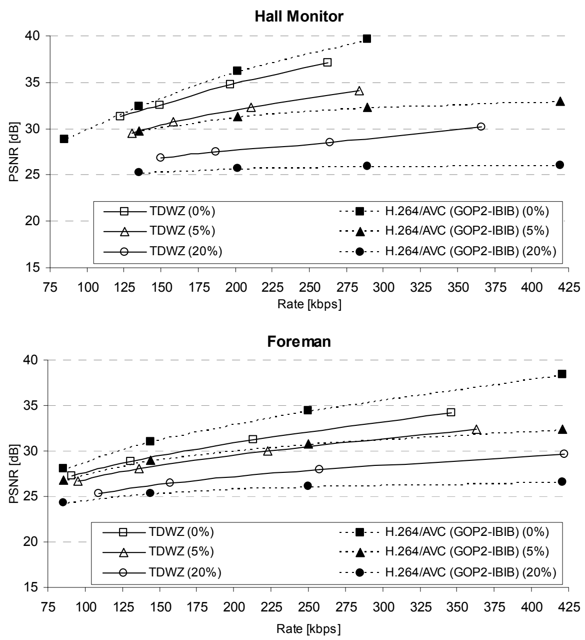


Figure 9 – TDWZ versus H.264/AVC RD performance (errors on key frames and WZ frames in case B) for Hall Monitor and Foreman.

6. REFERENCES

[1] B. Girod, A. Aaron, S. Rane, D. R. Monedero, “Distributed Video Coding”, Proc. of the IEEE, vol. 93, nº 1, pp. 71 - 83, Jan. 2005.

[2] C. Brites, J. Ascenso, F. Pereira, “Improving Transform Domain Wyner-Ziv Video Coding Performance”, ICASSP, Toulouse, France, May 2006.

[3] J. Ascenso, C. Brites, F. Pereira, “Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding”, 5th EURASIP Conf. on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, Jun. 2005.

[4] K.-P. Lim, G. Sullivan, T. Wiegand “Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods”, JVT-K049, Munich, Germany, March 2004.

[5] S. Wenger,”H.264/AVC over IP”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.13, No.7, pp645-656, Jul. 2003.

[6] S. Wenger, “Proposed Error Patterns for Internet Experiments”, Doc. VCEG Q15-I-16r1, VCEG Meeting, Red Bank, NJ, USA, Oct. 1999.