FEEDBACK CHANNEL IN PIXEL DOMAIN WYNER-ZIV VIDEO CODING: MYTHS AND REALITIES *

[1]Catarina Brites, [2]João Ascenso, and [3]Fernando Pereira

[1, 3] Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001, Lisbon, Portugal
phone: + (351) 21 8418462, fax: + (351) 21 8418472
[2] Instituto Superior de Engenharia de Lisboa – Instituto de Telecomunicações
R. Conselheiro Emídio Navarro, 1, 1959-007 Lisbon - Portugal
[1]catarina.brites@.lx.it.pt, [2]joao.ascenso@lx.it.pt, [3]fp@lx.it.pt

ABSTRACT

Wyner-Ziv (WZ) video coding – a particular case of distributed video coding (DVC) – is a new video coding paradigm based on two major Information Theory results: the Slepian-Wolf and Wyner-Ziv theorems. Recently, practical WZ video coding solutions were proposed with promising results. Many of the solutions available in the literature make use of a feedback channel (FC) to perform rate control at the decoder. In this context, this paper intends to analyse the impact of this feedback channel, notably through a number of metrics such as the frequency the feedback channel is used as well as its associated rate. It is also presented a study on the evolution of the decoded frames quality as more parity bits are requested via feedback channel. Those measures are important since they allow characterizing the usage of the feedback channel, and have never been presented in the literature.

1. INTRODUCTION

Today's digital video coding paradigm, represented by the standardization efforts of ITU-T VCEG and ISO/IEC MPEG, lies on hybrid DCT and interframe predictive coding with motion compensation. In this coding framework, the encoder is typically 5 to 10 times more complex than the decoder [1], mainly due to the motion estimation/compensation task; after all, it is the encoder that has to take all coding decisions, and is responsible to achieve the best performance, while the decoder remains a pure executer of the encoder "orders". This kind of architecture is well-suited for applications where the video is encoded once and decoded many times, i.e. one-to-many topologies, such as broadcasting or video-on-demand, and the cost of the decoder is more critical than the cost of the encoder.

In recent years, with emerging applications such as wireless low-power surveillance, multimedia sensor networks, wireless PC cameras and mobile camera phones, the traditional video coding architecture is being challenged. These applications have very different requirements than those of the broadcast video delivery systems. For some applications, a low power consumption both at the encoder and decoder sides is essential, e.g. in mobile camera phones. In other types of applications, notably when there is a high number of encoders and only one decoder, e.g. surveillance, low cost encoder devices are necessary. And for other applications, at least a flexible allocation of the codec complexity between the encoder and decoder is very much welcome.

Distributed video coding fits well these scenarios, since it enables to exploit the video statistics, partially or totally, at the decoder only. This paradigm targets a flexible allocation of complexity between encoder and decoder which may have as a rather important case the low encoding complexity. From the Information Theory, the Slepian-Wolf theorem [2] states that it is possible to compress two statistically dependent signals, X and Y, in a distributed way (separate encoding, jointly decoding) using a rate similar to that used in a system where the signals are encoded and decoded together, i.e. like in traditional video coding schemes. The complement of Slepian-Wolf coding for lossy compression is the Wyner-Ziv (WZ) coding [3]. WZ coding deals with the lossy source coding of X considering that Y, known as side information (a guess of X), is only available at the decoder. The side information is usually interpreted as an attempt made by the decoder to obtain an estimate of the original frame. In the WZ coding scenario, error correcting codes are used to improve the quality of the side information until a target quality for the final decoded frame is achieved.

One of the most interesting DVC approaches is the turbo-based pixel domain Wyner-Ziv coding scheme presented in [4], where the decoder is responsible to explore all the source statistics, and therefore to achieve compression following the Wyner-Ziv coding paradigm. In the low encoding complexity DVC approaches available in the literature [4] [6], it is often used a feedback channel in order to request more parity bits and thus successfully correct the errors in the side information. The feedback channel is a unidirectional channel used by the decoder to inform the encoder that the current biplane was successfully decoded or more parity bits are necessary for its successful decoding.

In this context, several key questions may arise: Is the usage of this feedback channel a limitation by itself ? How much decoder requests are made to successfully decoded a given bitplane ? How much is the feedback channel rate ? How is the evolution of the decoded frame quality with the (increasing) number of parity bits received ? Although many authors use WZ coding architectures including the feedback channel, the answers to these questions have never been presented in the literature. These questions deserve attention in order to characterize the usage of the feedback channel as any other tool in the architecture, and following that improve its performance, for example by reducing the number of requests.

This paper proposes to study the myths and realities around the WZ feedback channel, notably by measuring, in the context of a pixel domain WZ video codec: 1) the number of decoder requests for each bitplane; 2) the feedback channel rate; 3) the evolution of the decoded frame quality with the number of requests. The paper is organized as follows: Section 2 presents a brief summary of the IST-PDWZ codec. This summary is necessary to study, in Section 3, the feedback channel, which constitutes the novel part of this paper regarding what exists in the literature. Experimental results are presented and analysed in Section 4. Conclusions and some future work topics are presented in Section 5.

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

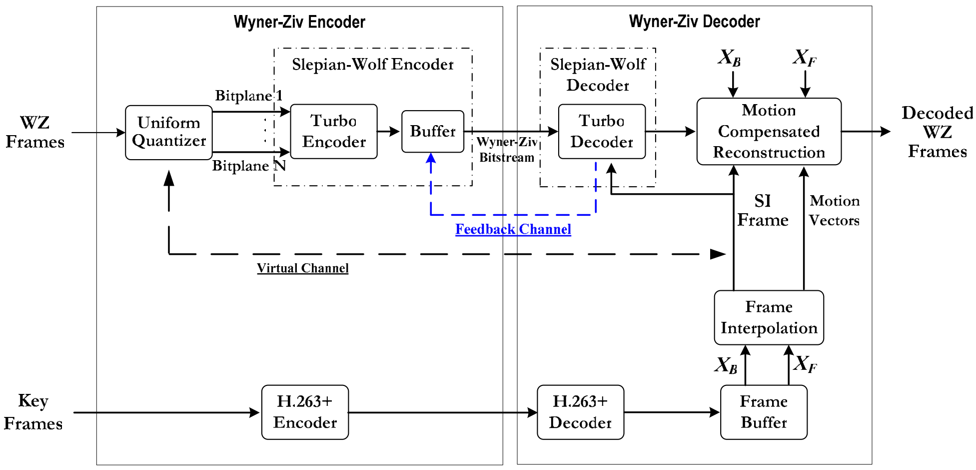


Figure 1 – The IST-PDWZ video codec architecture.

2. THE IST-PIXEL DOMAIN WYNER-ZIV VIDEO CODEC

Figure 1 illustrates the architecture of the IST-PDWZ video codec proposed in [5]; this codec is an improved PDWZ video coding solution which follows the same architecture as the one proposed by Aaron et al. in [4]. There are however some major differences regarding the coding solution proposed in [5], notably a more efficient side information generation scheme at the decoder by using motion compensated frame interpolation with spatial motion smoothing and a reconstruction function based on motion compensation with improved performance (for more details the reader should consult [6]).

In a nutshell, the coding process is as follows: the video frames are organized into key frames and Wyner-Ziv frames. The key frames are traditionally intraframe coded. The Wyner-Ziv frame pixel values are quantized using a 2^M-level uniform scalar quantizer; in this case, $2^M \in \{2, 4, 8, 16\}$. Over the resulting quantized symbol stream, constituted by all the quantized symbols of the WZ frame using M levels, bitplane extraction is performed. Each bitplane is then independently turbo encoded, starting with the most significant one. The parity bits produced by the turbo encoder are stored in the buffer and transmitted in small amounts upon decoder request via the feedback channel; the systematic bits are discarded.

At the decoder, the frame interpolation module is used to generate the side information frame, an estimate of the WZ frame, based on previously decoded frames, X_B and X_F. For a Group Of Pictures (GOP) length of 2, X_B and X_F are the previous and the next temporally adjacent key frames but any GOPs lengths are allowed. For the decoder to make good use of the side information it needs to have reliable knowledge of the model that characterizes the statistical relation between the (original) WZ frame and the corresponding side information. A Laplacian distribution is used to model the residual distribution between the WZ frame to be encoded and the side information since it provides a good fit for the residual.

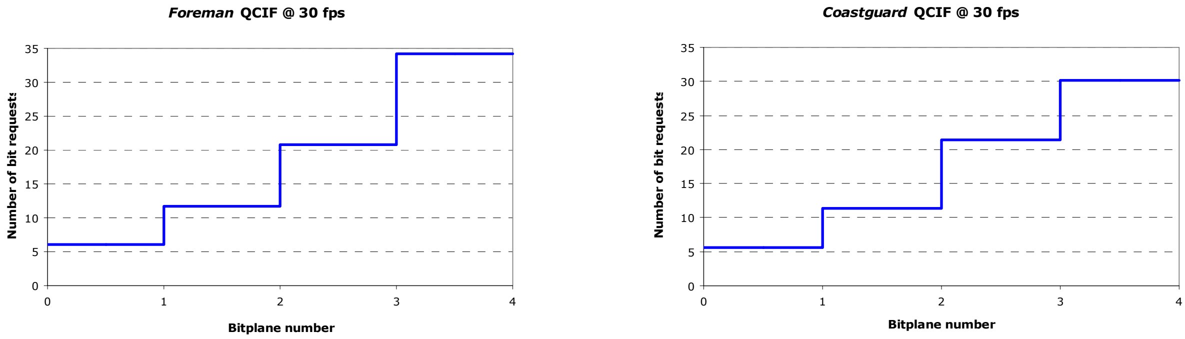
The side information is used by an iterative turbo decoder to obtain the decoded quantized symbol stream. The turbo decoder is constituted by two soft-input soft-output (SISO) decoders; each SISO decoder is implemented using the Maximum *A Posteriori* (MAP) algorithm. It is assumed that the decoder has ideal error detection capabilities regarding the current bitplane error probability P_e. For example, if $P_e > 10^{-3}$, the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the current bitplane turbo decoding task is considered successful and another bitplane starts being turbo decoded. The side information is also used in the motion compensated reconstruction module, together with the decoded quantized symbol stream, to help in the WZ frame reconstruction task [5].

A typical myth about this architecture states that the usage of the feedback channel is a weakness by itself. Although it is evident that the usage of a feedback channel is not realistic for applications that are intrinsically unidirectional, and therefore it is more than a weakness in that context, its usage for bidirectional applications is clearly possible and, in fact, wise if it allows to provide better overall codecs, e.g. more efficient in RD performance. This reasoning is clearly more relevant in the context of WZ coding since the decoder plays there a more central role, e.g. in terms of motion estimation, and thus it is only natural that the same happens also for rate control whenever possible. Not using the feedback channel when it is available and can bring benefits does not look wise, very much like not using the characteristics of the human visual system when it is known that the decoded video frames are to be seen by a human with certain well known characteristics. It is however true that the feedback channel precludes offline encoding (the video is encoded but decoded much later) since the decoder will only ask the encoder in real-time the bits needed while decoding.

A different issue may be the question of round trip delay since the feedback channel implies a certain delay in the coding of each frame. In fact, the delay is directly related to the frequency of feedback channel usage which depends on how fast the iterative turbo decoder can correct the side information errors since the turbo decoder requests more parity bits or not depending on the amount of errors still present in the side information. The round trip delay myth associated with the feedback channel presence mostly falls when considering the type of applications for which WZ coding seems to be relevant, e.g. sensor network, and wireless surveillance. In this context, big round trip delays are not expected and thus should not be a major problem. However the answer may be different if a high number of feedback requests are made; this shows another reason why the study proposed in this paper for the first time is important.

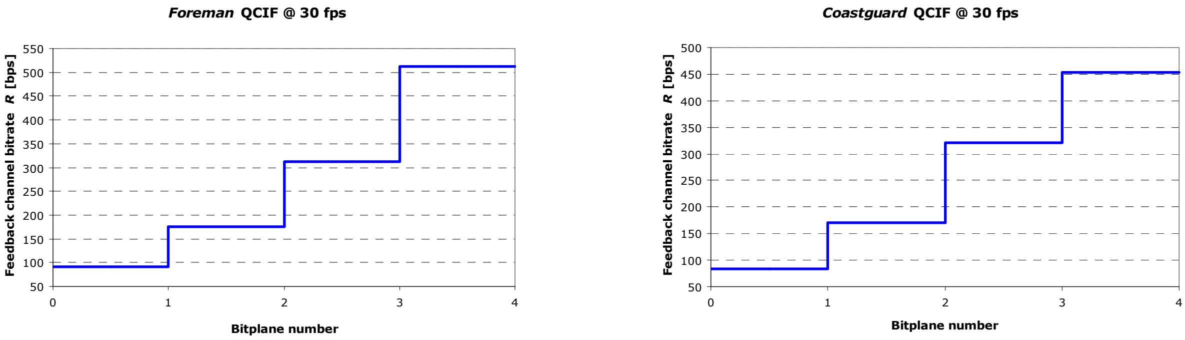
3. FEEDBACK CHANNEL BEHAVIOR AND IMPACT

After analysing some of the myths related to the WZ feedback channel, it is important to analyse the realities which will be here presented using some adequate metrics. In the adopted WZ coding architecture, the feedback channel has the role to adapt the bitrate to the changing statistics between the side information and the frame to be encoded, i.e. to the quality (or accuracy) of the frame interpolation (Section 2). Therefore, contrary to conventional codecs, it is the decoder responsibility to perform rate control and in this way to guarantee that only a minimum of parity bits are sent in order to correct the mismatches/errors present in each bitplane, and thus to achieve the minimum rate for a target quality.



a) b)

Figure 2 – Average bit request number for 1, 2, 3 and 4 bitplanes of the: a) *Foreman* and b) *Coastguard* QCIF sequences.



a) b)

Figure 3 – Average feedback channel rate for 1, 2, 3 and 4 bitplanes of the: a) *Foreman* and b) *Coastguard* QCIF sequences.

Since this decoder rate control operation based on the feedback channel is central in this architecture, it is important to be aware of its behaviour and impact in order to design more efficient algorithms to perform rate control, e.g. with a lower number of requests. In the following subsections, some feedback channel statistics are presented.

3.1 Measuring the Number of Bit Requests

During the decoding of a given bitplane, the decoder may request one or more times to the encoder for more parity bits. The number of requests depends mainly on the side information quality, on the number of bitplanes and on the accuracy of the correlation noise model used to characterize the residual between the WZ frame and the side information. In this paper, the Laplacian distribution parameter is estimated online at the decoder and at the block level [8] since it allows to achieve better RD performance.

To have an insight on how the number of requests varies with the temporal correlation of the video sequence, it is proposed here to measure, at the bitplane level, and for each frame, the number of parity bits requests. Thus, it is measured, for each WZ frame of a video sequence, the number of requests needed towards a successfully decoding of a certain number of bitplanes. The average number of decoder requests D, at the bitplane level, was computed from:

$$D = \frac{\sum_{b=1}^{M}\sum_{i=1}^{N} r_{bi}}{N} \qquad (1)$$

where M is the number of bitplanes, i.e. the number of quantizer levels (see Section 2), N is the total number of WZ frames and r_{bi} is the number of requests at bitplane b and WZ frame i.

3.2 Measuring the Feedback Channel Rate

After the average number of requests per bitplane is known, it is possible to measure the feedback channel rate for each bitplane. In order to measure the feedback channel rate, it is assumed that only one bit is required by the decoder to inform the encoder if more parity bits are needed or not to successfully decode the current bitplane. If more parity bits are needed, the decoder sends the bit '1' via the feedback channel; otherwise, the bit '0' is transmitted and the encoder, receiving such bit, sends parity bits for the next bitplane to be decoded. Note that the turbo coding operation is performed by sending the parity bits by decreasing order of relevance, this means from the most significant to the least significant bitplane. Since only one bit is transmitted via the FC for each decoder request, the feedback channel rate R, at the bitplane level, can be obtained from:

$$R = \frac{\sum_{b=1}^{M}\sum_{i=1}^{N} n_{bi}}{N} \times f \qquad (2)$$

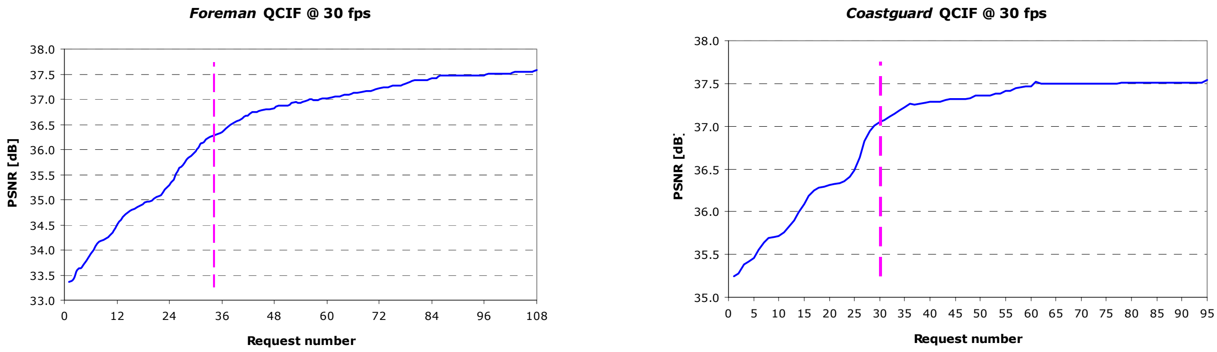
where f is the WZ frame rate and n_{bi} is the number of bits sent via the FC for bitplane b and WZ frame i ; as in (1), M is the number of bitplanes and N is the total number of WZ frames.

3.3 Measuring the Evolution of WZ Decoded Frames Quality Versus the Number of Bit Requests

Since the parity bits are successively requested to improve the decoded quality, it is important to know what is the evolution of this quality with the number of bit requests in order to design more adequate request strategies. The pseudo-algorithm to obtain the quality of WZ decoded frames as the number of requests increases is presented in the following:

1) For a given chunk (small amount) of parity bits received, the turbo decoder decodes the current bitplane;
2) After the turbo decoding operation, the WZ frame is reconstructed and the PSNR associated with the decoded frame is computed;
3) Then, the current bitplane error probability P_e is computed;

4.a) If $P_e > 10^{-3}$, the decoder requests for more parity bits from the encoder, transmitting the bit '1' via the feedback channel and



a) b)

Figure 4 – Wyner-Ziv decoded frames quality (considering 4 bitplanes) with the number of decoder requests for: a) *Foreman* and b) *Coastguard* QCIF sequences.

returning to step 1);

4.b) If $P_e \leq 10^{-3}$, the current bitplane turbo decoding task is considered successful and the turbo decoding of another bitplane starts. In this case, the decoder sends, via the feedback channel, the bit '0' to inform the encoder that parity bits for another bitplane should be transmitted (thus returning back to step 1); if there are no more bitplanes, the decoding of another WZ frame starts.

For both situations 4.a) and 4.b), the PSNR of the decoded frame is recorded thus providing not only the final but also the intermediate decoded quality and thus the quality evolution with the number of decoder bit requests.

4. EXPERIMENTAL RESULTS

This section will present and analyse the results regarding the metrics proposed in the previous section. These results will allow to have for the first time a better knowledge of the reality of the feedback channel in the context of WZ video coding.

4.1 Evaluation Conditions

The results presented in this section consider all frames of the *Foreman* and *Coastguard* QCIF video sequences, this means 400 and 300 frames, at 30 frames per second (fps). The test conditions for the frame interpolation and motion compensated reconstruction modules are [5]:

- **Frame interpolation:** 8×8 block size, ±8 pixels for the search range of the full block motion estimation, and ±2 pixels for the search range of the bi-directional motion estimation.
- **Motion compensated reconstruction:** Block size remains unchanged; the threshold to determinate if the block is motion compensated or not is an average difference of 0.15 per pixel and the search range is ±4 pixels.

The key frames are encoded with H.263+ intra with a quantization parameter (QP) equal to 13, 10, 8, 5, respectively, depending on the number of decoded Wyner-Ziv bitplanes; using these QP values for the key frames allows to have almost constant decoded video quality for the full set of frames (key frames and WZ frames). The Wyner- Ziv frame rate is 15 frames per second (i.e. GOP length of 2); as usual for WZ coding, only the luminance component is taken into account.

4.2 Analysis of Results

Analysing the Number of Bit Requests

Figure 2 a) and Figure 2 b) show the average number of decoder requests *D* (see Section 3.1) when the WZ frames are encoded with 1, 2, 3 and 4 bitplanes for the *Foreman* and *Coastguard* QCIF sequences, respectively. As expected, the number of bits requests increases with the number of bitplanes to be decoded. The increase is not constant across the bitplanes; it depends mainly on the correlation between the side information and the WZ frame for each bitplane; this correlation is usually high for the 1st bitplane (which represents the coarsest bin) and decreases for higher bitplane numbers (finer grained bins).

Analysing the Feedback Channel Rate

The average feedback channel rate *R* (see Section 3.2) is illustrated in Figure 3 a) and Figure 3 b) for the *Foreman* and *Coastguard* QCIF sequences respectively, when the WZ frames are encoded with 1, 2, 3 and 4 bitplanes. As it can be observed, the number of bits transmitted through the feedback channel increases with the number of bitplanes to be decoded; the correlation between the side information and the WZ frame is the main reason for that behaviour, as explained above.

Table 1 shows the IST-PDWZ total bitrate results for the *Foreman* and *Coastguard* QCIF sequences as well as the corresponding FC rate results (plotted in Figure 3). Comparing the results in Table 1, it is possible to conclude that the FC rate corresponds to almost 0.02% of the total bitrate for the first bitplane of the *Foreman* and *Coastguard* sequences; the fourth bitplane spends a significant amount of the FC rate, with FC rate percentages around 0.05% and 0.04% for the *Foreman* and *Coastguard* sequences respectively. These values suggest that rate control techniques to minimize the number of requests are needed.

Analysing the Evolution of WZ Decoded Frames Quality Versus the Number of Bit Requests

Figure 4 a) and Figure 4 b) illustrate the evolution of the reconstructed WZ frames quality after decoding 4 bitplanes with the number of decoder requests. As it can be observed, the maximum number of decoder requests is 108 for the *Foreman* sequence and 95 for the *Coastguard* sequence. For video sequences characterized by constant and well defined motion, like the *Coastguard* sequence, the frame interpolation algorithm employed at the decoder generates high quality side information, i.e. a good estimate of the WZ frame. The side information is therefore well interpolated, so few errors exist between the side information and the original WZ frame. Since there are few errors to be corrected by the turbo decoder, only few decoder requests are needed and thus a lower maximum number of decoder request is achieved. This concept is confirmed in Figure 5: the higher the side information PSNR (Figure 5 b), the lower is the number of decoder requests (Figure 5 d) since there are fewer errors to be corrected between the side information and the original WZ frame. The inconstant motion that characterizes the *Foreman* sequence justifies the fluctuations in the side information quality (Figure 5 a) and the corresponding fluctuations on the number of decoder requests per frame (Figure 5 b).

In Figure 4 a) and Figure 4 b), the pink dashed lined represents the average number of decoder requests; for the *Foreman* sequence,

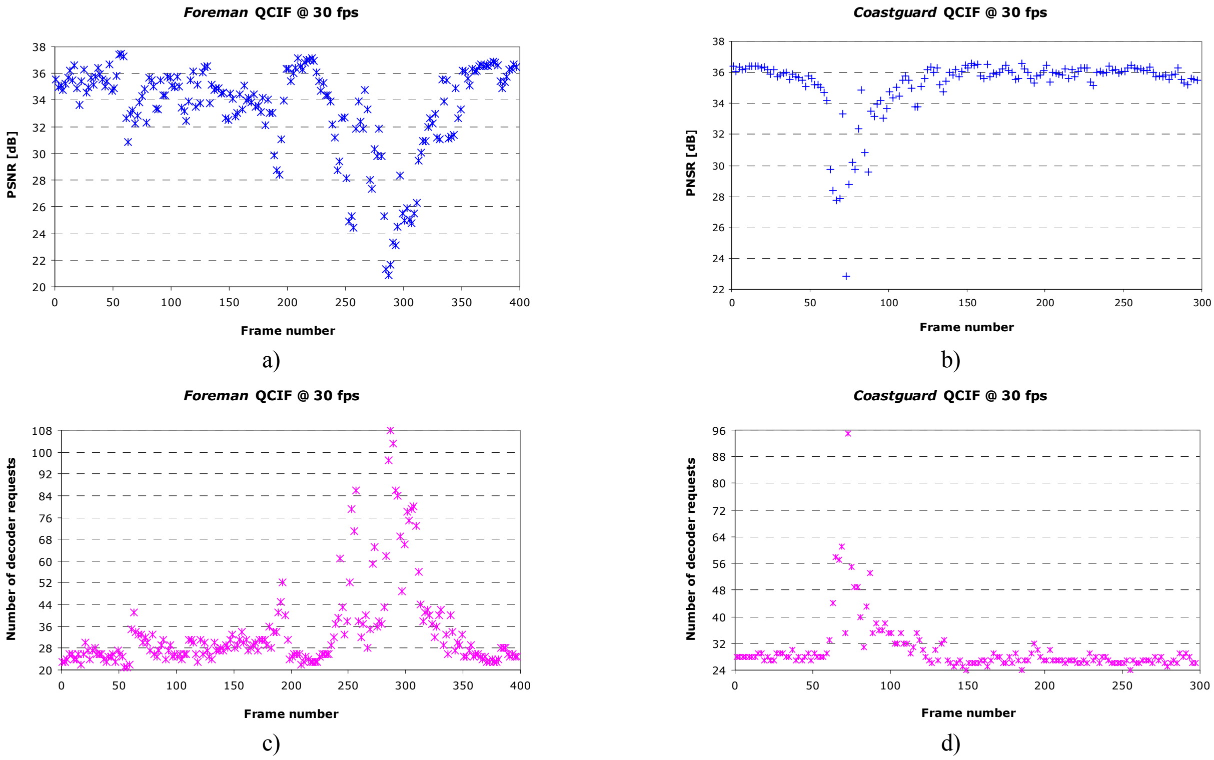


Figure 5 – Side information PSNR (considering 4 bitplanes) and the corresponding number of decoder requests at the frame level for the *Foreman* – a) and c) – and *Coastguard* – b) and d) – QCIF video sequences.

Table 1 – IST-PDWZ total bitrate and FC rate results for 1, 2, 3 and 4 bitplanes of the *Foreman* and *Coastguard* QCIF sequences.

Bitplane Nr.	*Foreman*		*Coastguard*	
	Total bitrate [kbps]	FC rate [bps]	Total bitrate [kbps]	FC rate [bps]
1	364.12	90.98	385.85	84.16
2	518.46	174.95	550.69	171.14
3	720.62	312.36	770.45	321.64
4	1125.64	512.34	1137.29	453.12

that value rounds 34.2 while for the *Coastguard* sequence 30.2 is the value obtained.

As already mentioned, the performance of the motion interpolation framework is the main reason behind this difference. As Figure 4 illustrates, the reconstructed frame quality increases quickly with the number of requests until the average number of decoder requests is reached (pink dashed line). After that, increasing the number of requests typically corresponds to a small increase in the average PSNR since for high number of requests only a few reconstructed frames still have errors to correct, i.e. can still improving its quality. In this region, the frames which have no more errors to be corrected contribute with the quality (PSNR) reached in the last request. Thus, the curve asymptotically converges to the average PSNR of the decoded WZ frames.

As is also shown in Figure 4, the number of decoder requests may be rather high depending on the targeted rate/quality. This may be a disadvantage for certain applications since it implies high latency in the decoding procedure bringing back the myth of round trip delay.

5. FINAL REMARKS

This paper intends to study the role of the feedback channel in pixel domain Wyner-Ziv video coding by presenting and analysing for the first time some relevant statistics. Based on the results obtained, it is planned to design in future work rate control algorithms which minimize the number of requests, notably targeting low latency requirements. The current study gives insights on how to develop improved rate control solutions, especially the information regarding the evolution of WZ decoded frames quality versus the number of bit requests.

REFERENCES

[1] ISO/IEC International Standard 14496-10:2003, "Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding".

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

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

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv Coding for Motion Video", *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, November 2002.

[5] J. Ascenso, C. Brites, and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding", *IEEE International Conference on Advanced Video and Signal Based Surveillance*, Como, Italy, September 2005.

[6] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *5th EURASIP*, Slovak Republic, July 2005.

[7] A. Aaron, E. Setton, and B. Girod, "Towards Practical Wyner-Ziv Coding of Video", *IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003.

[8] C. Brites, J. Ascenso, and F. Pereira, "Studying Temporal Correlation Noise Modelling for Pixel Based Wyner-Ziv Video Coding", *IEEE International Conference on Image Processing*, Atlanta, GA, USA, October 2006.