STUDYING THE FEEDBACK CHANNEL IN TRANSFORM DOMAIN WYNER-ZIV VIDEO CODING *

José Quintas Pedro[1], Catarina Brites[1], João Ascenso[2], and Fernando Pereira[1]

[1] 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
jose.pedro@lx.it.pt, catarina.brites@.lx.it.pt, joao.ascenso@lx.it.pt, 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 the feedback channel in transform domain WZ video coding, notably through a number of metrics such as the frequency the feedback channel is used as well as its associated rate. Also a study is presented on the quality evolution of the decoded frames as more parity bits are requested via the feedback channel. Finally, the codec performance is also analyzed in terms of the compression factor at bitplane/band level. Analysing the behaviour of the metrics above is important not only to understand the feedback channel impact but also to design new tools to improve the overall codec performance.

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. Distributed video coding (DVC) also know as Wyner-Ziv (WZ) 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 sub-case, low encoding complexity; for more details in terms of applications potential benefits, see [2].

In the last 3-4 years, there has been a growing interest in developing DVC practical solutions. Some of the most relevant solutions are based on turbo coding and a feedback channel, both for pixel and transform domains [3][4][5]; the feedback channel is used by the decoder to request more (parity) bits to the encoder and thus successfully correct the errors in the so-called side information generated at the decoder as an estimation of the information to be coded.

The impact and behaviour of the feedback channel for a pixel domain WZ codec was already studied for the IST-PDWZ codec [6]. This paper intends to perform a similar study for an analogous transform domain Wyner-Ziv codec where a discrete cosine transform (DCT) has been introduced to exploit the spatial redundancy in the video data. The target is to get answers to questions such as: How many decoder requests are made to successfully decode a given band/bitplane? What is the evolution of the decoded frame quality with the (increasing) number of parity bits received? Since it is possible to sent an amount of parity bits greater than the number of bits in a bitplane, are there any situations where instead of having bitrate compression we have bitrate expansion? These questions are relevant because they may allow gathering information to further improve the TDWZ codec performance, e.g. by reducing the number of requests or by adopting different coding solutions for the less efficient bands/bitplanes.

The paper is organized as follows: Section 2 presents a brief overview of the IST-TDWZ codec. Section 3 describes in detail the metrics to be evaluated in this paper. Experimental results are presented and analysed in Section 4. Finally, conclusions and some future work topics are presented in Section 5.

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

The IST-TDWZ codec here presented uses as starting point the IST-PDWZ [4], i.e. it reuses the pixel domain tools whenever this is adequate. The overall IST-TDWZ coding architecture illustrated in Figure 1 works as follows: a video sequence is divided into Wyner-Ziv frames and key frames as in the IST-PDWZ solution, odd and even frames respectively. While the key frames are coded using a standard coding solution, e.g. H.264/AVC Intra, WZ frames are coded using a DVC approach. Over each Wyner-Ziv frame X_{2i}, it is applied a 4×4 block-based discrete cosine transform (DCT). The DCT coefficients of the entire frame X_{2i} are then grouped together, according to the position occupied by each DCT coefficient within the 4×4 blocks, forming the DCT coefficients bands. After the transform coding operation, each DCT coefficients band b_k is uniformly quantized with 2^{M_k} levels (where the number of levels 2^{M_k} depends on the DCT coefficients band b_k). Over the resulting quantized symbol stream (associated to the DCT coefficients band b_k), bitplane extraction is performed. For a given band, the quantized symbols bits of the same significance (e.g. the most significant bit) are grouped together, forming the corresponding bitplane array which is then independently turbo encoded.

The turbo coding procedure for the DCT coefficients band b_k starts with the most significant bitplane array, which corresponds to the most significant bits of the b_k band quantized symbols. The parity information generated by the turbo encoder for each bitplane is then stored in the buffer and sent in chunks upon decoder request, through the feedback channel.

[1] The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

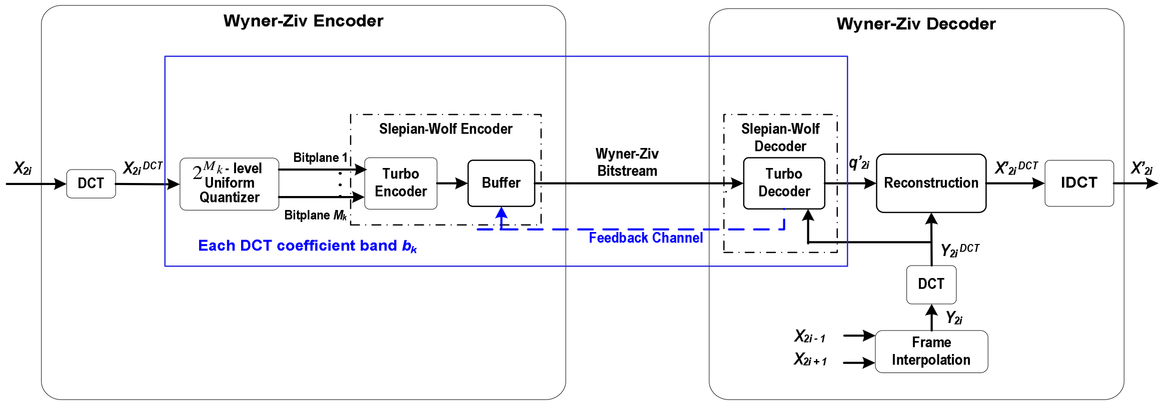


Figure 1 - The IST-TDWZ video codec architecture.

The decoder performs frame interpolation [4] using the previous and next temporally adjacent key frames of X_{2i} to generate an estimate of frame X_{2i}, Y_{2i}, the so-called side information. A block-based 4×4 DCT is then carried out over Y_{2i} in order to obtain Y_{2i}^{DCT}, an estimate of X_{2i}^{DCT}. The residual statistics between correspondent coefficients in X_{2i}^{DCT} and Y_{2i}^{DCT} is assumed to be modelled by a Laplacian distribution; the Laplacian parameter is estimated online and at different granularity levels, notably band and coefficient level. Once Y_{2i}^{DCT} and the residual statistics for a given DCT coefficients band b_k are known, the decoded quantized symbol stream q'_{2i} associated to the DCT band b_k can be obtained through an iterative turbo decoding procedure. After successfully turbo decoding the most significant bitplane array of the b_k band, the turbo decoder proceeds in an analogous way to the remaining M_{k-1} bitplanes associated to that band. Once all the bitplane arrays of the DCT coefficients band b_k are successfully turbo decoded, the turbo decoder starts decoding the b_{k+1} band. This procedure is repeated until all the DCT coefficients bands for which WZ bits are transmitted are turbo decoded. After turbo decoding the M_k bitplanes associated to the DCT band b_k, the bitplanes are grouped together to form the decoded quantized symbol stream associated to the b_k band; this procedure is performed over all the DCT coefficients bands to which WZ bits are transmitted. Once all decoded quantized symbol streams are obtained, it is possible to reconstruct the matrix of DCT coefficients, $X'_{2i}{}^{DCT}$. For some DCT coefficients bands, no WZ bits are transmitted; at the decoder, those DCT coefficients bands are replaced by the corresponding DCT bands of the side information, Y_{2i}^{DCT}. After all DCT coefficients bands are reconstructed, a block-based 4×4 inverse discrete cosine transform (IDCT) is performed and the reconstructed X_{2i} frame, X'_{2i}, is obtained. To finally get the decoded video sequence, decoded key frames and WZ frames are mixed conveniently.

3. MEASURING THE FEEDBACK CHANNEL BEHAVIOR

In the adopted TDWZ coding architecture, the feedback channel has the role to adapt the bitrate to the changing statistics between the side information (an estimation of the frame to be encoded) and the frame to be encoded, i.e. to the quality (or accuracy) of the frame interpolation [4]. Therefore, contrary to conventional codecs, it is the decoder's responsibility to perform rate control and in this way to guarantee that only a minimum of parity bits are sent to correct the mismatches/errors present in each side information bitplane, thus the use of a channel coding tool like turbo coding.

Since this decoder rate control operation based on the feedback channel is central in the IST-TDWZ architecture, it is important to be aware of its behaviour and impact in order to design more efficient WZ video coding algorithms. Thus, in the following subsections, some relevant metrics will be defined for later evaluation and analysis.

3.1 Measuring the Number of Requests

During the decoding of a given bitplane of a given band b_k, 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 b_k band 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.

To have an insight on how the number of requests varies with the temporal correlation of the video sequence (and thus with the quality of the side information), it is proposed here to measure, at the bitplane level of each band, 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 at frame, D_Q, and bitplane, D_{Qij}, levels for a certain quality rank, Q, is computed by (1) and (2):

$$D_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N} r_{ijl}}{N} \quad (1) \qquad D_{Qij} = \frac{\sum_{l=1}^{N} r_{ijl}}{N} \quad (2)$$

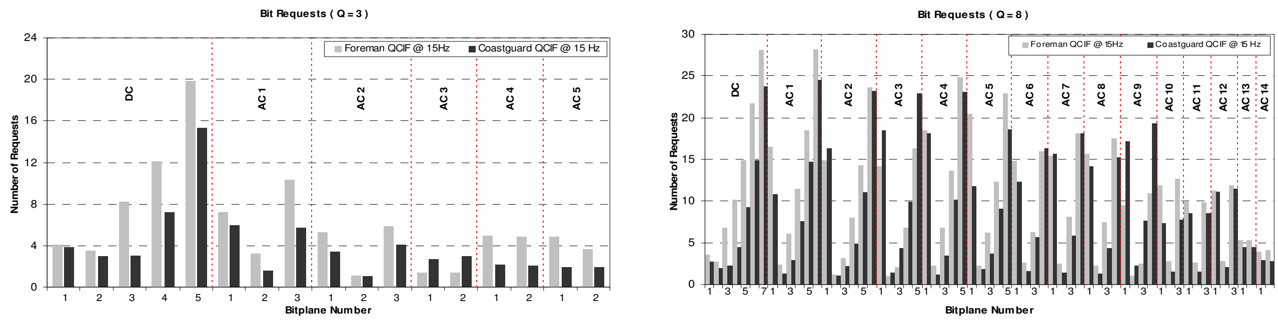
where r_{ijl} is the number of requests made via the feedback channel for WZ frame l for bitplane j at band i; N is the total number of WZ frames coded, M_i is the number of bitplanes in each band b and B_Q is the number of bands for a certain quality rank Q. D_{Qij} is a partial result of equation (1) representing the average number of decoder requests per frame for bitplane j of band b and i and for quality rank Q.

3.2 Measuring the Feedback Channel Rate

After the average number of requests per band and bitplane is known, it is possible to measure the feedback channel rate for each band and 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. Since only one bit is transmitted via the feedback channel for each decoder request, the total feedback channel rate at frame R_Q, and bitplane R_{Qij} levels for a certain quality rank, Q, can be obtained from (3) and (4) respectively:

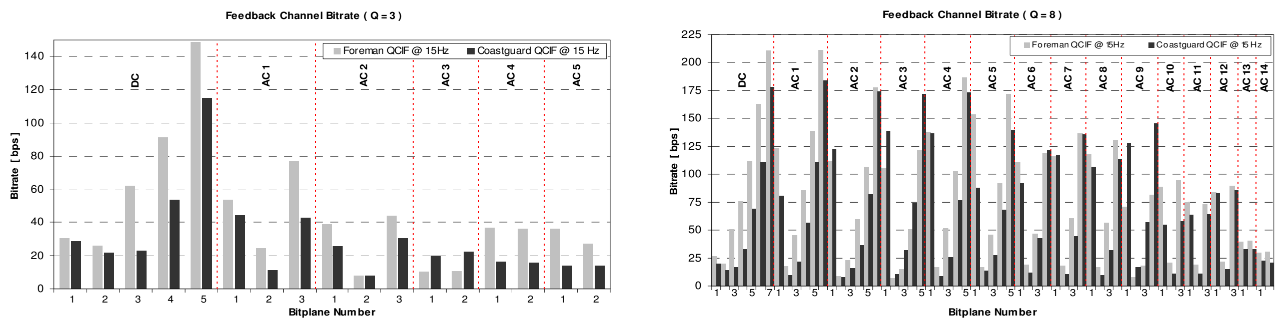
$$R_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N} n_{ijl}}{N} \times f \quad (3) \qquad R_{Qij} = \frac{\sum_{l=1}^{N} n_{ijl}}{N} \times f \quad (4)$$

In (3) and (4), f is the WZ frame rate and n_{ijl} is the number of bits sent via the feedback channel for WZ frame l, for bitplane j at band i; N is the total number of WZ frames, M_i is the number of bitplanes of each band i and B_Q is the number of bands considering a certain quality



a) b)

Figure 2- Average number of requests for *Coastguard* and *Foreman* QCIF, 15 Hz sequences using: a) Q=3 and b) Q=8.



a) b)

Figure 3 - Average feedback channel rate for *Coastguard* and *Foreman* QCIF, 15 Hz sequences using: a) Q=3 and b) Q=8.

rank, Q. R_{Qij} is a partial result of equation (1) representing the average feedback channel rate per frame for bitplane j of band i for a certain quality rank Q.

3.3 Measuring the Quality Evolution of WZ Decoded Frames

Since the parity bits are successively requested to improve the decoded quality, it is important to know how the WZ frames quality evolves with the number of bit requests in order to design more adequate request strategies since this also has a significant impact on the decoder complexity. The algorithm to obtain the WZ decoded frames quality 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 for the current band;
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;
   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 returns to step 1);
   b) If $P_e \leq 10^{-3}$, the current bitplane turbo decoding task is considered successful and the turbo decoding of the next 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 or band (if there are no more bitplanes to decode for the current band) should be transmitted (thus returning back to step 1); if there are no more bitplanes and bands, the decoding of another WZ frame starts.

For both situations 3.a) and 3.b), the PSNR of the decoded frame is recorded thus providing not only the final but also the intermediate decoded frame quality and thus the quality evolution with the number of decoder bit requests.

3.4 Measuring the Bitplane Compression Factor

As described in [4], the turbo encoder encloses two recursive systematic convolutional (RSC) encoders of rate 1/2, which means that the total number of parity bits per bitplane created by the RSCs is twice the number of the input bitplane bits. This way it is possible to have situations where the number of parity bits sent is bigger (maximum twice bigger) than the original bitplane itself; of course, this is an undesirable situation that must be avoided. The total average compression factor at frame, CF_Q, and bitplane, CF_{Qij}, levels for a certain quality rank, Q, is given by (5) and (6):

$$CF_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N}\frac{C_{ijl}}{w_{ijl}}}{N} \quad (5) \qquad CF_{Qij} = \frac{\sum_{l=1}^{N}\frac{C_{ijl}}{w_{ijl}}}{N} \quad (6)$$

where M_i is the number of bitplanes of each band b, N the total number of WZ frames, C_{ijl} the number of bits in each original coefficient bitplane j of each band i and w_{ijl} is the amount of parity bits sent for each bitplane j of band i at frame l, B_Q is the number of bands considering the quality rank, Q. CF_{Qij} given by (6) represents the average compression factor at bitplane j of band i for a certain quality rank Q.

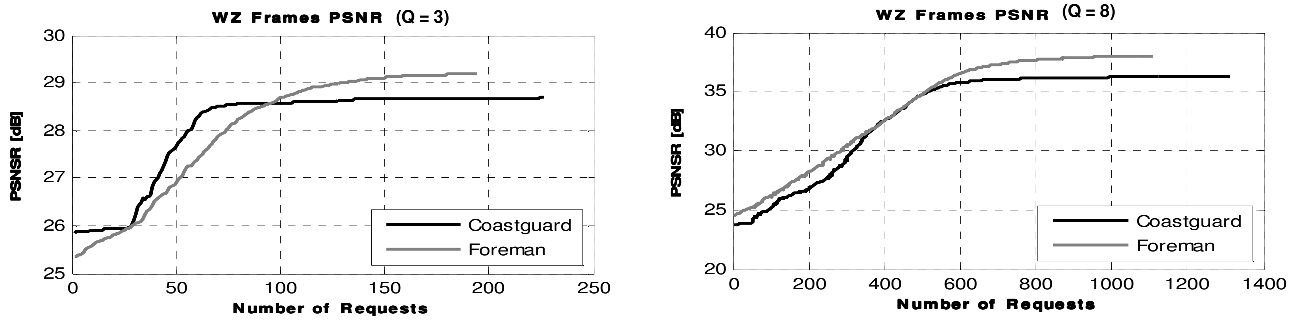
4. EXPERIMENTAL RESULTS

This section will present and analyse the results regarding the metrics proposed in the previous section. These results will allow having a better knowledge of the reality of the feedback channel in the context of transform domain 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 150 frames, at 15 frames per second (fps). The test conditions for the frame interpolation and motion compensated reconstruction are the same as in [6]. Due to the lack of space, only two rate-distortion points (the 3rd and the 8th) represented by Q, i.e. two quantization matrices as defined in [5] have been used for this evaluation; the higher is Q, the higher are the bitrate and the quality. It was checked that, for other Q values, the metrics behaviour is the same as for the Q values used here.

The key frames are always encoded with H.264/AVC Intra. For the *Foreman* QCIF sequence, at 15 frames per second, the key frames are encoded with a quantization parameter (QP) equal to 39, and 26, for Q=3 and Q=8, respectively. For the *Coastguard* sequence, and for the same Q, the QP used is equal to 38 and 27, respectively. Using these QP values for the key frames allows having almost constant decoded video quality for the full set of frames (key frames and WZ frames). Since a GOP length of 2 has been used, the Wyner-Ziv frame rate is



a) b)

Figure 4 - Wyner-Ziv decoded frames quality with the number of decoder requests for: a) Q=3 and b) Q=8.

7.5 frames per second; as usual for WZ coding, only the luminance component is coded and thus results refer only to the luminance.

4.2 Analysis of Results

Analysing the Number of Requests

The average number of decoder requests per bitplane per band is shown in Figure 2a) and 2b) for WZ frames encoded using Q=3 and Q=8, for the *Coastguard* and *Foreman* QCIF sequences at 15 Hz (the bands are separated by the red lines). The amount of correlation between the side information and the WZ frame is the main reason for the evolution of the decoder requests; the higher is the correlation, the less parity bits are requested by the decoder. It can be noticed that the decoder always makes fewer requests for the second bitplane of each band than for the first, due to a strong correlation with the first bitplane. However, that correlation becomes weaker as more bitplanes the decoder has to decode, and more parity bits are needed. The use of a variable puncturing period, smaller for less significant bitplanes, may reduce the usage of the feedback channel and thus the latency.

Analysing the Feedback Channel Rate

Figures 3a) and 3b) show the average feedback channel rate for Q=3 and Q=8, for the *Foreman* and *Coastguard* QCIF sequences at 15 Hz. As expected, the number of bits through the feedback channel varies proportionally to the number of requests. For the *Foreman* and *Coastguard* sequences, the feedback channel rate represents 0.041% and 0.028% of the total rate (110 kbps and 108 kbps), for Q=3, and 0.014% and 0.012% of the total rate (552 kbps and 554 kbps), for Q=8, respectively. In spite of the percentage being smaller for Q=8, the feedback channel bitrate is around twice the feedback channel bitrate for Q=3. These results show that the major impact of the feedback channel is not related to the feedback rate since it is rather negligible.

Analysing the Quality Evolution of WZ Decoded Frames

Figures 4a) and 4b) present the quality evolution of the reconstructed WZ frames, after each request, regardless the error probability of the bitplane. The evolution of the WZ frames quality reconstructed after each request may be irregular because the number of errors in a bitplane does not always decrease with the number of requests (this is a well-known turbo coding feature); however, it is the final quality that matters when assessing the rate-distortion performance. While for Q=3 (lower final quality and thus lower quality key frames), the *Coastguard* sequence shows better performance for the initial requests, the opposite happens for Q=8; this may be due to the better behaved motion of *Coastguard*. However, the final quality is always higher for *Foreman*, very likely due to the complex texture, i.e. water, of *Coastguard*.

Analysing the Bitplane Compression Factor

The average compression factor per bitplane per band, for Q=8, for both sequences, is illustrated in Figure 5. Generally, good compression factors are achieved per bitplane, since most of them are bigger than 1; however, as more bitplanes are considered in each band, less correlation between bitplanes exists and thus, consequently, more requests are needed (see Figure 5), decreasing the compression factor. For the *Coastguard* and *Foreman* sequences, and for Q=3, there are 0.3% and 1.19% of the total bitplanes where bitrate expansion occurs, although the compression factor is always close to 1 (not much below). For Q=8, bitrate expansion occurs in 4.16% for *Coastguard*, and 7.79% for *Foreman*, of the total number of bitplanes; in some cases, a compression factor of 0.5 is achieved, which means that twice the size of the original bitplane is transmitted. The more affected bitplanes, as illustrated in Figure 5, are the less significant bitplanes, particularly the last bitplane of each band, where the correlation between the side information and the WZ frames is especially weaker, notably when a lot of bitplanes to describe each band are used.

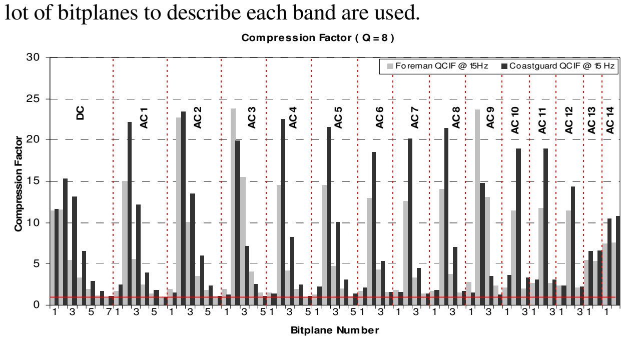


Figure 5 – Average compression factor per bitplane per band for *Coastguard* and *Foreman* QCIF, 15 Hz sequences using: a) Q=3 and b) Q=8.

5. FINAL REMARKS

This paper presents a study of the behaviour of the feedback channel for a transform domain Wyner-Ziv codec. The analysis of the metrics evaluated will be useful for the design of improved techniques, e.g. rate control minimizing the use of the feedback channel and variable size of the parity bits package to decrease the latency and decoder complexity by avoiding too many turbo decoder runs.

REFERENCES

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

[2] C. Brites and F. Pereira, "Distributed Video Coding: Bringing New Applications to Life", *5th Conference on Telecommunications - ConfTele*, Tomar, Portugal, April 2005.

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

[4] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Smolenice, Slovak Republic, June 2005.

[5] C. Brites, J. Ascenso and F. Pereira, "Improving Transform Domain Wyner-Ziv Coding Performance", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006.

[6] C. Brites, J. Ascenso and F. Pereira; "Feedback Channel In Pixel Domain Wyner-Ziv Video Coding: Myths and Realities", *14th European Signal Processing Conference*, Florence, Italy, September 2006.