Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 23 (2008) 269–297
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

Evaluating a feedback channel based transform domain Wyner–Ziv video codec

Catarina Brites[a], João Ascenso[b], José Quintas Pedro[a], Fernando Pereira[a,*]

[a]*Instituto Superior Técnico—Instituto de Telecomunicações, Av. Rovisco Pais, 1049-001 Lisbon, Portugal*
[b]*Instituto Superior de Engenharia de Lisboa—Instituto de Telecomunicações, R. Conselheiro Emídio Navarro, 1, 1959-007 Lisbon, Portugal*

Received 30 August 2007; received in revised form 13 January 2008; accepted 11 March 2008

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. In recent years, some practical WZ video coding solutions have been proposed with promising results. One of the most popular WZ video coding architectures in the literature uses turbo codes based Slepian–Wolf coding and a feedback channel to perform rate control at the decoder. This WZ video coding architecture has been first proposed by researchers at Stanford University and has been after adopted and improved by many research groups around the world. However, while there are many papers published with changes and improvements to this architecture, the precise and detailed evaluation of its performance, targeting its deep understanding for future advances, has not been made. Available performance results are mostly partial, under unclear and incompatible conditions, using vaguely defined and also sometimes architecturally unrealistic codec solutions.

This paper targets the provision of a detailed, clear, and complete performance evaluation of an advanced transform domain WZ video codec derived from the Stanford turbo coding and feedback channel based architecture. Although the WZ video codec proposed for this evaluation is among the best available, the main purpose and novelty of this paper is the solid and comprehensive performance evaluation made which will provide a strong, and very much needed, performance reference for researchers in this WZ video coding field, as well as a solid way to steer future WZ video coding research.

Keywords: Distributed video coding; Wyner–Ziv video coding; Turbo coding; Performance evaluation

1. Introduction

Although without even noticing, a growing percentage of the world population uses nowadays image, video and audio coding technologies on a rather regular basis. These technologies are behind the success and quick deployment of services and products such as digital pictures, digital television, DVDs and MP3 players. The main purpose of digital audiovisual coding technologies is to compress the original information into a much smaller number of bits, without affecting in an unacceptable way the decoded signal quality. Regarding video, the current coding paradigm is mostly based on four

*Corresponding author. Tel.: +351 218418460; fax: +351 218418472.
E-mail address: Fernando.Pereira@lx.it.pt (F. Pereira).

0923-5965/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2008.03.002

types of tools: (i) motion compensated temporal prediction between video frames to exploit the temporal redundancy; (ii) transform coding, typically using the Discrete Cosine Transform (DCT), to exploit the spatial redundancy; (iii) quantization of the transform coefficients to exploit the irrelevancy related to the human visual system limitations; and (iv) entropy coding to exploit the statistical redundancy of the created coded symbols. The quality of the decoded video is mainly controlled through the quantization process and may be adapted to the service needs or compression factors requested. Because the current video coding paradigm considers both the temporal (prediction) and frequency (DCT) domains, this type of coding architecture is well known as hybrid predictive or only predictive coding.

Since predictive coding has been the solution adopted in most available video coding standards, notably the ITU-T H.26x and ISO/IEC MPEG-x families of standards, this coding paradigm is nowadays used in hundreds of millions of video encoders and decoders. Because this video coding solution exploits the correlation between and within the video frames at the encoder, it typically leads to rather complex encoders and much simpler decoders, without much flexibility in terms of complexity budget allocation besides making the encoder less complex and thus less efficient. This approach fits well some application scenarios, such as broadcasting, using the so-called down-link model, where a few encoders typically provide coded content for millions of decoders; in this case, the decoder complexity is the real critical issue. Moreover, the temporal prediction loop used to compute the residuals to transmit, after the motion compensated prediction of the current frame, requires the decoder to run the same loop in perfect synchronization with the encoder. This means that, when there are channel errors, the temporal prediction synchronization is lost and errors propagate in time, strongly affecting the video quality until some Intra coding refreshment is performed.

With the wide deployment of wireless networks, there are a growing number of applications which do not fit well the typical down-link model but rather follow an up-link model where many senders deliver data to a central receiver. Examples of these applications are wireless digital video cameras, low-power video sensor networks, and surveillance systems. Typically, these emerging applications require light encoding or a flexible distribution of the codec complexity, robustness to packet losses, high compression efficiency and, many times, also low latency/delay; there is also a growing usage of multiview video content, which means the data to be delivered regards many (correlated) views of the same scene. The ideal case would be to find a video coding solution that could address all these requirements with the same coding efficiency as the best predictive coding schemes available, as well as with an encoder complexity and error robustness similar to the current Intra coding solutions; this would mean low complexity encoding and (almost) no error propagation due to the absence of the prediction loop.

To address some of these issues, some research groups decided, around 2002, to revisit the video coding problem at the light of an Information Theory theorem from the 70s: the Slepian–Wolf theorem [20]. This theorem addresses the case where two statistically dependent discrete random sequences independently and identically distributed, X and Y, are independently encoded, and not jointly encoded as in the largely deployed predictive coding solution. The Slepian–Wolf theorem says that the minimum rate to encode the two (correlated) sources is the same as the minimum rate for joint encoding, with an arbitrarily small error probability. This distributed source coding (DSC) paradigm is a very interesting result in the context of the emerging application challenges presented above, since it opens the doors to new coding solutions where, at least in theory, separate encoding does not induce any compression efficiency loss when compared to the joint encoding used in the traditional predictive coding paradigm. Slepian–Wolf coding is the term generally used to characterize coding architectures that follow this independent encoding approach. Slepian–Wolf coding is also referred in the literature as lossless distributed source coding since it considers that the two statistically dependent sequences are perfectly reconstructed at a joint decoder (neglecting the arbitrarily small probability of decoding error), thus approaching the lossless case. Slepian–Wolf coding has a deep relationship with channel coding. Since the sequence X is correlated with the sequence Y, it can be considered that a virtual "dependence channel" exists between sequences X and Y. The Y sequence is, therefore, a "noisy" ("erroneous") version of the original uncorrupted X sequence. Thus, the "errors" between the X and Y sequences can be corrected applying a channel coding technique to encode X

which will be conditionally (jointly) decoded using Y; this relationship was studied in the 70s by Wyner [23]. Thus, channel coding tools typically play a main role in the new distributed source coding paradigm.

Though, there was at this stage still a major constraint since the Slepian–Wolf theorem regards 'lossless coding' which is not the most useful case in practical video coding solutions. In fact, lossless coding achieves rather small compression factors since it does not eliminate the irrelevant video information, not perceived by the human visual system. However, in 1976, Wyner and Ziv studied a particular case of Slepian–Wolf coding corresponding to the lossy source coding of the X sequence considering that the Y sequence, known as side information (SI), is available at the decoder. These studies allowed to derive the so-called Wyner–Ziv (WZ) theorem [24] which states that when performing independent encoding with side information under certain conditions, i.e. when X and Y are jointly Gaussian sequences and a mean-squared error distortion measure is considered, there is no coding efficiency loss with respect to the case when joint encoding is performed, even if the coding process is lossy (and not 'lossless' anymore). Later, it would be shown that only the innovation, this means the $X-Y$ difference, needs to be Gaussian, relaxing the requirements on the joint X and Y statistics [18].

Together, the Slepian–Wolf and the Wyner–Ziv theorems suggest that it is possible to compress two statistically dependent signals in a distributed way (separate encoding, jointly decoding) approaching the coding efficiency of conventional predictive coding schemes (joint encoding and decoding). Based on this result, a new video coding paradigm, well known as distributed video coding (DVC), has emerged. DVC does not rely on joint encoding and thus, when applied to video coding, results on the absence of the temporal prediction loop (used in predictive video coding schemes) and lower complexity encoders. Therefore, DVC based architectures may provide the following functional benefits which are rather important for many emerging applications: (i) flexible allocation of the global video codec complexity; (ii) improved error resilience; (iii) codec independent scalability (since upper layers do not have to rely on precise lower layers); and (iv) exploitation of multiview correlation without cameras/encoders communicating among them. The functional benefits above can be relevant for a large range of emerging application scenarios such as wireless video cameras, low-power surveillance, video conferencing with mobile devices, disposable video cameras, visual sensor networks, distributed video streaming, multiview video systems, and wireless capsule endoscopy [11].

With the theoretical doors opened, practical design of WZ video codecs started around 2002, following important advances in channel coding technology, especially error correction codes with a capacity close to the Shannon limit, e.g. turbo and low-density parity-check (LDPC) codes. While theory suggests that WZ video coding solutions may be as efficient as joint encoding solutions, practical developments did not yet achieve that performance in any condition, especially if low complexity encoding is also targeted. For example, while the theory assumes that the encoder knows the statistical correlation between the two sources, X and Y, and the innovation $X-Y$ to be Gaussian, in real conditions this is often not true. Naturally, the better the encoder knows the statistical correlation between X and Y, the higher the compression efficiency; this highlights a main WZ coding paradox: although encoders may be rather simple, they may also need to become more complex to increase the compression efficiency and reach the limits set by the theory.

The first practical WZ solutions emerged around 2002, notably from Stanford University [1,10] and University of California, Berkeley [19]. The most popular practical architecture in the literature seems to be the Stanford architecture which is mainly characterized by turbo codes based Slepian–Wolf coding and a feedback channel to perform rate control at the decoder. This architecture has been after adopted and improved by many research groups around the world; hundreds of papers have been published with evolutions and variations of this WZ video coding solution. However, while there are many papers published with changes and improvements to this architecture, the precise and detailed evaluation of its performance, targeting its deep understanding for later advances, has not been made. Available performance results are mostly partial, e.g. typically only some rate-distortion (RD) performance results, under unclear and incompatible conditions, e.g. different sequences, different sets of frames for each sequence, different key frames coding, using vaguely defined and also sometimes architecturally unrealistic codecs, e.g. assuming the original frames available at the decoder, using side

information at the encoder, or a vague side information creation (SIC) process.

The lack of clear, credible, and complete WZ video coding performance references, at least for some key architectures, is affecting the understanding, comparison, and evolution of this research activity since it is difficult to benchmark new solutions with such an unclear landscape. In this context, this paper intends to contribute to overcome this *status quo* by providing a detailed, clear, and complete performance evaluation of an advanced transform domain WZ codec based on the Stanford architecture [1,10]. Although the WZ video codec proposed for this evaluation is among the most powerful available, the main purpose and novelty of this paper is precisely the solid and comprehensive performance evaluation made which will provide a strong, and very much needed, performance reference for experts in this field.

To reach the stated objective, this paper is organized in the following way: Section 2 provides a description of the WZ video codec developed by the authors and under evaluation; the description is complemented with a set of published papers which together define in detail the implemented and evaluated WZ video codec. This is essential to reach one of the main objectives of this paper: allow the readers to replicate without much difficulty the WZ video coding solution evaluated. Section 3 describes the evaluation conditions, while Sections 4–6 present and analyze the performance results related to the forward channel, i.e. key frames and WZ bits, the feedback channel, i.e. decoder requests bits, and the codec complexity. Finally, Section 7 concludes the paper.

2. The evaluated transform domain Wyner–Ziv video codec

The TDWZ (transform domain Wyner–Ziv) video codec developed and implemented by the authors, and evaluated in detail in this paper, is based on the pixel domain (PD) WZ codec developed at Instituto Superior Técnico (IST) designated as IST-PDWZ codec [3]; this basically means that the TDWZ codec reused the PDWZ codec tools whenever this was possible and appropriate, including in addition a spatial transform (in this case, the DCT) to exploit the spatial redundancy in the video data [10]. A detailed description of some of the modules in this TDWZ codec is available in [3,4,7,8]; in the following, a summary is presented.

The TDWZ coding architecture illustrated in Fig. 1 works as follows: a video sequence is divided into WZ frames and key frames. The key frames may be inserted periodically with a certain group of pictures (GOP) size or instead an adaptive GOP size selection process can be used to exploit the different amount of temporal correlation along the video sequence [4]; most results available in the literature use a GOP size of 2 which means that odd and even frames are key frames and WZ frames, respectively.

While the key frames are coded using a standard coding solution, e.g. H.264/AVC Intra [12], WZ frames are coded using a WZ coding approach.

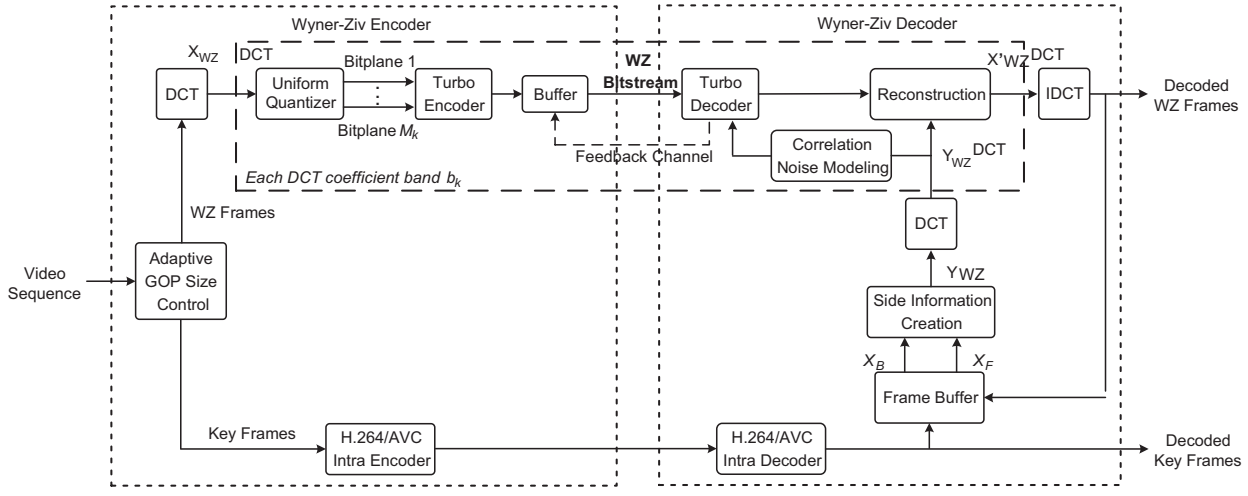

Fig. 1. TDWZ video codec architecture.

Over each WZ frame X_{WZ}, a 4×4 block-based DCT [7] is applied. The DCT coefficients of the entire frame X_{WZ} 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 quantization symbol stream (associated to the DCT coefficients band b_k), bitplane extraction is performed. For a given band, the quantization symbols bits of the same significance (e.g. the most significant bit, MSB) are grouped together, forming the corresponding bitplane arrays which are then independently turbo encoded.

The turbo coding procedure for the DCT coefficients band b_k starts with the MSB 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; for more details, please see Section 2.1. The usage of the feedback channel has implications beside the obvious need for that feedback channel to be available, notably: (i) this coding architecture can only be used for real-time applications scenarios; (ii) the application and the video codec must be able to accommodate the delay associated to the feedback channel; and (iii) the usage of the feedback channel simplifies the rate control problem since the decoder, knowing the available side information, can easily regulate the necessary bitrate. For a more detailed discussion on the feedback channel in the context of this WZ video coding architecture, please see [9].

The decoder creates the so-called side information for each WZ coded frame, Y_{WZ}, a good estimate of X_{WZ}, by performing a motion compensated frame interpolation process [4] using the previous and next decoded frames temporally closer to X_{WZ}; for more details, please see Section 2.2. The better is the estimation, the smaller the number of 'errors' the WZ turbo codec has to correct. A block-based 4×4 DCT is then carried out over Y_{WZ} in order to obtain Y_{WZ}^{DCT}, an estimate of X_{WZ}^{DCT}. The residual statistics between correspondent coefficients in X_{WZ}^{DCT} and Y_{WZ}^{DCT} is assumed to be modeled by a Laplacian distribution; the Laplacian parameter is estimated online at the decoder, for each DCT coefficient, based on the residual between the two motion compensated reference frames used to create Y_{WZ} [6]; for more details, please see Section 2.3. Once Y_{WZ}^{DCT} and the residual statistics for a given DCT coefficients band b_k are known, the decoded quantization symbol stream q'_{WZ} associated to the DCT band b_k can be obtained through an iterative turbo decoding (tDec) procedure. After successfully tDec the most significant bitplane array of the b_k band, the decoding process 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 tDec the M_k bitplanes associated to the DCT band b_k are grouped together to form the decoded quantization symbol stream associated to the b_k band. Once all decoded quantization symbol streams are obtained, the DCT coefficients, X'^{DCT}_{WZ} are reconstructed using an optimal mean squared error (MSE) estimate; for more details, see Section 2.4. For the DCT coefficients bands for which no WZ bits are transmitted (depends on the selected WZ quantization) the decoder uses the corresponding DCT bands of the side information, Y_{WZ}^{DCT}. After all DCT coefficients bands are reconstructed, a block-based 4×4 Inverse Discrete Cosine Transform (IDCT) is performed and the reconstructed X_{WZ} frame, X'_{WZ}, is obtained. To finally get the decoded video sequence, decoded key frames and WZ frames are conveniently mixed.

The WZ coding architecture here adopted may also include additional tools such as the encoder sending some auxiliary data for the WZ frames to help the decoder to create better side information, notably for the more motion critical parts of the frame. Since this auxiliary data may assume many forms, e.g. hash codes in [2], the performance evaluation proposed in this paper regards the case without encoder auxiliary data for the WZ frames, also trying to limit the encoder complexity. It is important to stress that the TDWZ codec adopted here does not include any of the architectural limitations often present in papers adopting the Stanford architecture; this means no original frames are used at the decoder to create the side information, to measure the bitplane error probability or to estimate the parameters for the turbo decoder correlation noise model.

2.1. TDWZ Slepian–Wolf codec

The Slepian–Wolf codec is a key module in WZ coding architectures since it is responsible to correct the errors in the side information, Y_{WZ} (the WZ frame estimation), generated at the decoder. Due to the relationship between Slepian–Wolf coding and channel coding (see Section 1), the Slepian–Wolf codec is usually constituted by an efficient channel codec. While the first WZ coding solutions using the architecture adopted in this paper have made use of turbo codes [1,10], it is also possible to use other channel coding solutions such as LDPC codes; for a comparison between turbo and LDPC codes in the context of this WZ video coding architecture, please see [17].

In the TDWZ video coding architecture (Fig. 1), the Slepian–Wolf encoder consists of a turbo encoder and a buffer. The turbo encoder encloses a parallel concatenation of two identical constituent recursive systematic convolutional (RSC) encoders of rate 1/2; a pseudo-random L-bit interleaver is employed to decorrelate the L-bit input sequence between the two RSC encoders. The pseudo-random interleaver length L corresponds to the DCT coefficients band size, i.e. the ratio between the frame size and the number of different DCT coefficients bands. Each RSC encoder outputs a parity stream and a systematic stream. After turbo encoding a bitplane of a given DCT coefficients band (starting with the most significant one), the systematic part (a copy of the turbo encoder input) is discarded and only the parity bits are stored in the buffer. Upon decoder request, the parity bits produced by the turbo encoder are transmitted according to a pseudo-random puncturing pattern. In each request, the WZ encoder sends one parity bit in each P parity bits from each RSC encoder parity stream; each parity bit is only sent once; P is often called the puncturing period. The location of the parity bits to be sent in each request is pseudo-random generated (it is known both at the encoder and decoder).

The Slepian–Wolf decoder encloses an iterative turbo decoder constituted by two soft-input soft-output (SISO) decoders. Each SISO decoder is implemented using the Logarithmic *Maximum A Posteriori* (Log-MAP) algorithm. The decoding process starts by converting the side information (DCT coefficients) into soft-input information (conditional bit probabilities) using a Laplacian distribution (with the corresponding parameter estimated online) to model the residual between the original frame DCT bands and the corresponding DCT bands of the side information. For a given bit representing a quantization symbol, the conditional bit probability is computed taking into account the side information and all the previously decoded bits, thus exploring the correlation between consecutive bitplanes; for more details, see [5]. These probabilities are used in the iterative tDec process, where the *extrinsic* information, computed by one SISO decoder, becomes the *a priori* information of the other SISO decoder. Through this information exchange procedure between the two SISO decoders, the *a posteriori* estimates of the data bits are updated until the maximum number of iterations allowed for iterative decoding is reached. A confidence measure based on the *a posteriori* probabilities ratio is used as error detection criterion to estimate the current bitplane error probability P_e for a given DCT band [14]. If P_e is higher than 10^{-3}, the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the bitplane tDec task is considered successful.

In the following, more details are provided for the WZ video codec tools which are more complex and critical for the RD performance.

2.2. TDWZ side information creation process

Several frame interpolation techniques can be employed at the WZ decoder to generate the side information, Y_{WZ}. The choice of the techniques used can significantly influence the codec RD performance. More accurate side information through frame interpolation means less errors (Y_{WZ} is more similar to X_{WZ}); therefore the decoder needs to request less parity bits from the encoder and the bitrate is reduced for the same quality. The TDWZ decoder uses the block-based frame interpolation framework proposed in [3] to generate accurate side information, see Fig. 2. The frame interpolation module generates the side information Y_{WZ}, an estimate of the X_{WZ} frame, based on two references, one temporally in the past (X_B) and another in the future (X_F), as follows:

1. For a GOP length of 2, X_B and X_F are the previous and the next temporally adjacent decoded key frames to the WZ frame being decoded. For other GOP lengths, the frame interpolation structure definition algorithm proposed in [4] indicates the decoding order and the

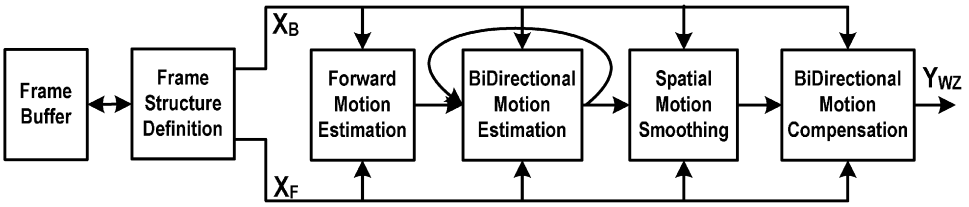

Fig. 2. Architecture of frame interpolation for side information creation.

reference frames to be used by the frame interpolation algorithm.

2. Both reference frames, X_B and X_F, are first low pass filtered to improve the reliability of the motion vectors. Then, a block-matching algorithm is used to estimate the motion between the X_B and X_F frames. In this step, full search motion estimation with modified matching criteria is performed [3]; the criteria include a regularized term that favors motion vectors that are closer to the origin.
3. The bidirectional motion estimation module refines the motion vectors obtained in the previous step with an additional constraint: the motion vector selected for each block has a linear trajectory between the next and previous reference frames and crosses the interpolated frame at the center of the blocks. This technique combines a hierarchical block size technique with a new adaptive search range strategy in a very efficient way [4]. The hierarchical coarse-to-fine approach tracks fast motion and handles large GOP sizes in the first iteration (block size 16×16) and then achieves finer detail by using smaller block sizes (block size 8×8).
4. Next, a spatial motion-smoothing algorithm based on weighted vector median filters [3] is used to make the final motion vector field smoother, except at object boundaries and uncovered regions.
5. Once the final motion vector field is obtained, the interpolated frame can be filled by simply using bidirectional motion compensation as defined in standard video coding schemes.

This type of advanced frame interpolation solution has a major contribution for the good RD performance of the selected WZ video codec since the quality of the generated side information has a crucial role in the overall codec performance.

2.3. TDWZ correlation noise modeling

To make good usage of the side information obtained through the frame interpolation framework, the decoder needs to have a reliable knowledge of the model that characterizes the correlation noise, $(X_{WZ}^{DCT} - Y_{WZ}^{DCT})$, between corresponding DCT bands of the WZ and side information frames. The Laplacian distribution is widely used to model the residual statistics between correspondent coefficients in X_{WZ}^{DCT} and Y_{WZ}^{DCT}, e.g. [7,10], and, thus, it is also adopted in this paper. The Laplacian distribution parameter α is estimated online at the decoder, for each DCT coefficient, using Eq. (1), based on the residual between the reference frames X_B and X_F used to create Y_{WZ} after motion compensation [6]. The α parameter estimation in Eq. (1) leads to a finer adaptation of the correlation noise model, both spatially (within a frame) and temporally (along the video sequence).

$$\hat{\alpha}_b(u,v) = \begin{cases} \hat{\alpha}_b & [D_b(u,v)]^2 \leqslant \hat{\sigma}_b^2 \\ \sqrt{\frac{2}{[D_b(u,v)]^2}} & [D_b(u,v)]^2 > \hat{\sigma}_b^2 \end{cases}. \quad (1)$$

In Eq. (1), $\hat{\alpha}_b(u,v)$ is the α parameter estimate for the DCT coefficient located at (u,v) position, $\hat{\sigma}_b^2$ and $\hat{\alpha}_b$ are, respectively, the estimates of the variance and the α parameter for the DCT band to which the DCT coefficient belongs to and D_b represents the distance between the (u,v) coefficient and the DCT coefficients band b average value; the distance D_b measures how spread the coefficient's values are regarding its corresponding DCT band average value [6].

2.4. TDWZ reconstruction

The turbo decoded bitplanes together with the side information and the residual statistics for each DCT coefficient band are used by the reconstruction (Rec) to obtain the decoded DCT coefficients matrix, X'^{DCT}_{WZ} as in [15]. Consider that the M_k bitplanes associated to each DCT coefficients band, for which WZ bits were received, are successfully decoded. For each band, the bitplanes are grouped and a decoded quantization symbol (bin) q' is obtained for each DCT coefficient, guiding the

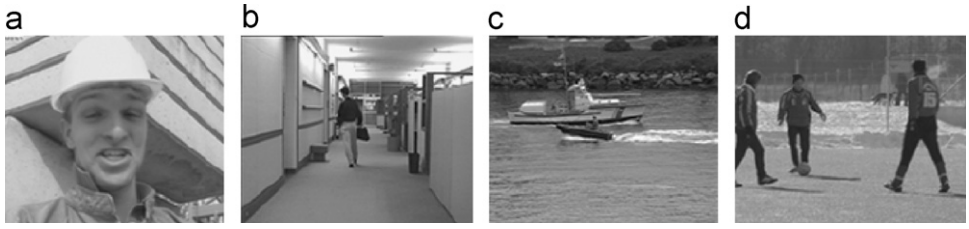

Fig. 3. Sample frames for test sequences: (a) Foreman (frame 80); (b) Hall Monitor (frame 75); (c) Coast Guard (frame 60); and (d) Soccer (frame 8).

decoder about where the original DCT coefficient value lies (an interval). The decoded quantization bin q' corresponds to the true quantization bin q, obtained at the encoder before bitplane extraction, if all errors in the decoded bitplanes were corrected (however, a small error probability is allowed). The Rec function is optimal in the sense that it minimizes the mean squared error of the reconstructed value, for each DCT coefficient, of a given band and is given by [15]:

$$x' = E[x|q', y] = \frac{\int_l^u x f_{X|y}(x|y)\,\mathrm{d}x}{\int_l^u f_{X|y}(x|y)\,\mathrm{d}x}, \tag{2}$$

where x' is the reconstructed DCT coefficient, y is the corresponding DCT coefficient of $Y_{\mathrm{WZ}}^{\mathrm{DCT}}$, $E[\cdot]$ is the expectation operator and l and u represent the lower and the upper bounds of q', respectively. In Eq. (2), the conditional probability density function $f_{X|y}(\cdot)$ models the residual statistics between corresponding coefficients in $X_{\mathrm{WZ}}^{\mathrm{DCT}}$ and $Y_{\mathrm{WZ}}^{\mathrm{DCT}}$; according to Section 2, $f_{X|y}(\cdot)$ is assumed to be a Laplacian distribution. After some analytical manipulations in Eq. (2), the reconstructed DCT coefficient can be obtained from Eq. (3) where Δ is the quantization bin size [15]; in Eq. (3), α is the Laplacian distribution parameter estimated online at the decoder for each DCT coefficient (see Section 2.3).

$$x' = \begin{cases} l + b & y < l \\ y + \dfrac{(\gamma + (1/\alpha))\,\mathrm{e}^{-\alpha\gamma} - (\delta + (1/\alpha))\,\mathrm{e}^{-\alpha\delta}}{2 - \mathrm{e}^{-\alpha\gamma} - \mathrm{e}^{-\alpha\delta}} & y \in [l, u[\\ u - b & y \geqslant u \end{cases} \quad \text{with } b = \frac{1}{\alpha} + \frac{\Delta}{1 - \mathrm{e}^{\alpha\Delta}}, \quad \gamma = y - l, \quad \delta = u - y \tag{3}$$

As it can be noticed from Eq. (3), the Rec function shifts the reconstructed DCT coefficient value towards the center of the decoded quantization bin. The DCT coefficients bands to which no WZ bits are sent are replaced by the corresponding DCT bands of the side information Y_{WZ}.

3. Performance evaluation conditions

Considering the main purpose of this paper, it is essential to precisely define the performance evaluation conditions and the relevant parameters necessary to control the WZ video codec selected. As usual in the WZ video coding literature, only the luminance component is coded and thus all rate and distortion results refer only to the luminance.

3.1. Test material

The video test material used and some relevant test conditions are described in the following:

- *Sequences*: Foreman (with the Siemens logo), Hall Monitor, Coast Guard, and Soccer; these sequences represent different types of content.
- *Frames for each sequence*: All frames; this means 299 frames for Foreman, 329 frames for Hall Monitor, 299 frames for Coast Guard, and 299 frames for Soccer (one sample frame of each test sequence at 30 Hz is plotted in Fig. 3).
- *Spatial resolution*: QCIF.
- *Temporal resolution*: 15 and 30 Hz (this means 7.5 or 15 Hz for the WZ frames when GOP = 2 is used).
- *GOP length*: 2 if not otherwise indicated; sometimes, also 4 or 8.

3.2. Quantization

Different RD performance can be achieved by changing the M_k value for the DCT band b_k. In this

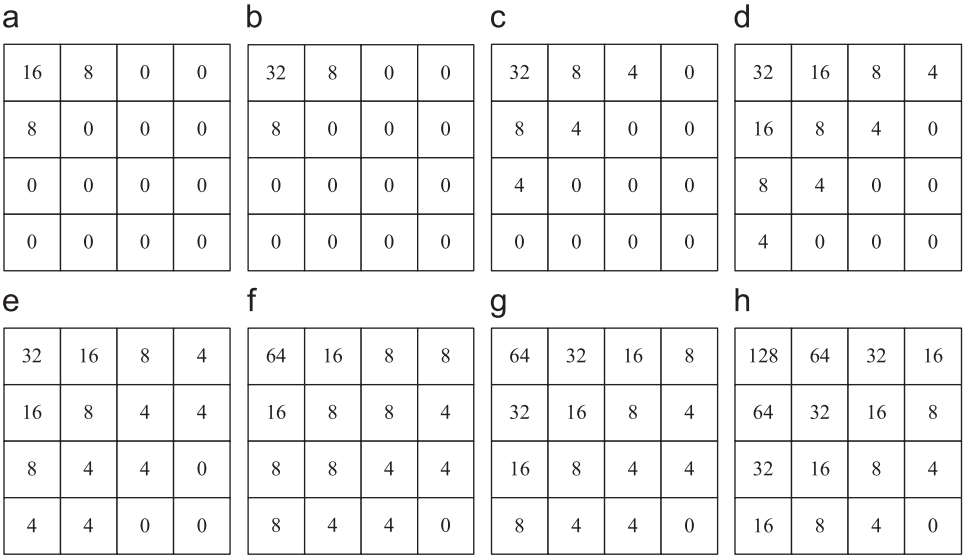

Fig. 4. Eight quantization matrices associated to different RD performances.

paper, eight RD points are considered corresponding to the various 4×4 quantization matrices depicted in Fig. 4. The first seven matrices in Fig. 4 are similar to the ones used in [7] while the last matrix is proposed by the authors to evaluate the TDWZ RD performance for higher bitrates. Within a 4×4 quantization matrix, the value at position k in Fig. 4 indicates the number of quantization levels associated to the DCT coefficients band b_k; the value 0 means that no WZ bits are transmitted for the corresponding band. In the following, the various matrices will be referred as Q_i with $i = 1, \ldots, 8$; when Q_i increases, the bitrate and the quality also increase.

3.3. Turbo codec

For the turbo encoder defined in Section 2.1, each rate 1/2 RSC encoder is represented by the generator matrix: $\left[1 \quad \frac{1+D+D^3+D^4}{1+D^3+D^4}\right]$; the trellis of the second RSC encoder is not terminated. The puncturing period P is 48, which allows a fine control of the bitrate.

3.4. Side information creation process

For the frame interpolation algorithm in the side information creation process, the smallest block size is 8×8 and ± 32 pixels are used for the forward motion estimation search range. All techniques operate at full-pel precision and the low pass filter is the mean filter with a 3×3 mask size.

Table 1
Key frames quantization parameters for the various RD points for QCIF at 15 Hz

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
Foreman	42	40	39	36	35	33	31	26
Hall Monitor	37	36	35	33	32	31	29	25
Coast Guard	39	38	38	35	34	33	31	27
Soccer	45	44	42	38	38	35	31	26

3.5. Key frames coding

The key frames are always encoded with H.264/AVC Intra (Main profile) [12] since this is among the best performing standard Intra coding solutions available. The key frames are coded with the quantization parameters (QPs) as defined in Tables 1 and 2. The QPs have been found through an iterative process which stops when the average quality (PSNR) of the WZ frames is equal to the quality of the key frames (H.264/AVC Intra encoded); the values in Tables 1 and 2 have been obtained for GOP 2, QCIF, 15 and 30 Hz. The selection of these QP values for the key frames was made with the target to have almost constant decoded video quality for the full set of frames (key frames and WZ frames) since this is important from the subjective impact point of view. Allocating

Table 2
Key frames quantization parameters for the various RD points for QCIF at 30 Hz

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
Soccer	43	41	40	36	35	33	29	24
Foreman	37	35	35	32	31	29	28	23

the same total bitrate in a different way between WZ and key frames may lead to a better RD performance, for example by allocating more bits to the key frames, at the cost of a less stable video quality (which is not desirable); anyway, this is a choice to be made by the encoder which needs to be taken into account when comparing RD performances.

The performance evaluation presented in the next sections includes only *a small but representative set of the full amount of results available*, notably in terms of video sequences, spatial resolutions, frame rates, GOP sizes, and RD points (Q_i), which due to space constraints cannot be included here.

4. Forward channel performance evaluation

This section targets the performance evaluation of the forward channel, which, in the described transform domain WZ video codec, regards the bits associated to the key frames and WZ frames. Since the forward channel bits are sent to reach the maximum video decoded quality, this section mainly regards the evaluation of how efficiently the resources—key frames and WZ bits—are used to maximize the decoded quality. In this case, and due to length constraints, error free channels are considered but it is well acknowledged that WZ video coding solutions are especially interesting when error resilience is a critical requirement [16].

4.1. Measuring the overall rate-distortion performance

Although many metrics are relevant to evaluate the forward channel performance, it is recognized that the most used metric is the average PSNR over all coded frames of a sequence using a certain quantization matrix for the DCT coefficients, as defined in Section 3. When this metric is represented as a function of the used bitrate—in this case, including the WZ and key frames bits for the luminance component—very important performance charts are obtained since they allow to easily comparing the overall RD performance with other coding solutions, e.g. the largely well known and used standard coding solutions.

In this paper, the RD performance of the described WZ solution is compared with the corresponding performance of three standard coding solutions which have in common the fact that the expensive motion estimation task at the encoder is not performed for any of them. The three standard video coding solutions used for benchmarking are:

- *H.263+ Intra* [13]—Coding with H.263+ without exploiting temporal redundancy; this 'Intra comparison' is still the one that appears the most in the WZ video coding literature but H.263+ Intra is clearly not the best standard Intra coding available; thus obtaining better RD results than H.263+ Intra is much easier than for H.264/AVC Intra.
- *H.264/AVC Intra* [12]—Coding with H.264/AVC in Main profile without exploiting temporal redundancy. This type of Intra coding is among the most efficient standard Intra coding solutions, even more than JPEG 2000 for many conditions; thus, it is much more difficult to defeat than H.263+ Intra. Note that the spatial correlation in H.264/AVC Intra is efficiently exploited with several 4×4 and 16×16 Intra modes (a feature missing in the TDWZ codec) and the CABAC arithmetic encoder, of course at the cost of some additional complexity.
- *H.264/AVC (Inter) No Motion* [12]—Coding with H.264/AVC in Main profile exploiting temporal redundancy in a IB…BI structure but without performing any motion estimation which is the most computationally expensive encoding task; this comparison is not typically provided in most WZ coding published papers because it is still very difficult to defeat this RD performance with the current WZ video coding solutions.

Fig. 5 shows the RD performance for the TDWZ codec (GOP 2) and the three standard-based solutions above presented for the eight quality levels, previously defined in Section 3. From the charts, it is possible to extract the following conclusions:

- At 15 Hz, the TDWZ codec has a better or similar RD performance to H.264/AVC Intra for

all sequences except for the Soccer sequence; this behavior is not consistent for the Foreman sequence, since TDWZ performs better for the lower bitrates and worse for the higher bitrates. The worst RD performance for the Soccer sequence is justified by the fact that this sequence has rather high and complex motion, making difficult for the decoder frame interpolation to create good side information, especially at 15 Hz, where the key frames are temporally farther apart. For Soccer, the TDWZ codec just has a RD performance equivalent to the H.263 + Intra codec, defeating it only for the lower bitrates.

- At 15 Hz, the TDWZ codec already manages to defeat the H.264/AVC No Motion codec for the Coast Guard sequence since its motion is especially uniform and well behaved and thus the decoder frame interpolation manages to create good side information. This is also starting to happen for the Hall Monitor sequence, although only for the lower bitrates.

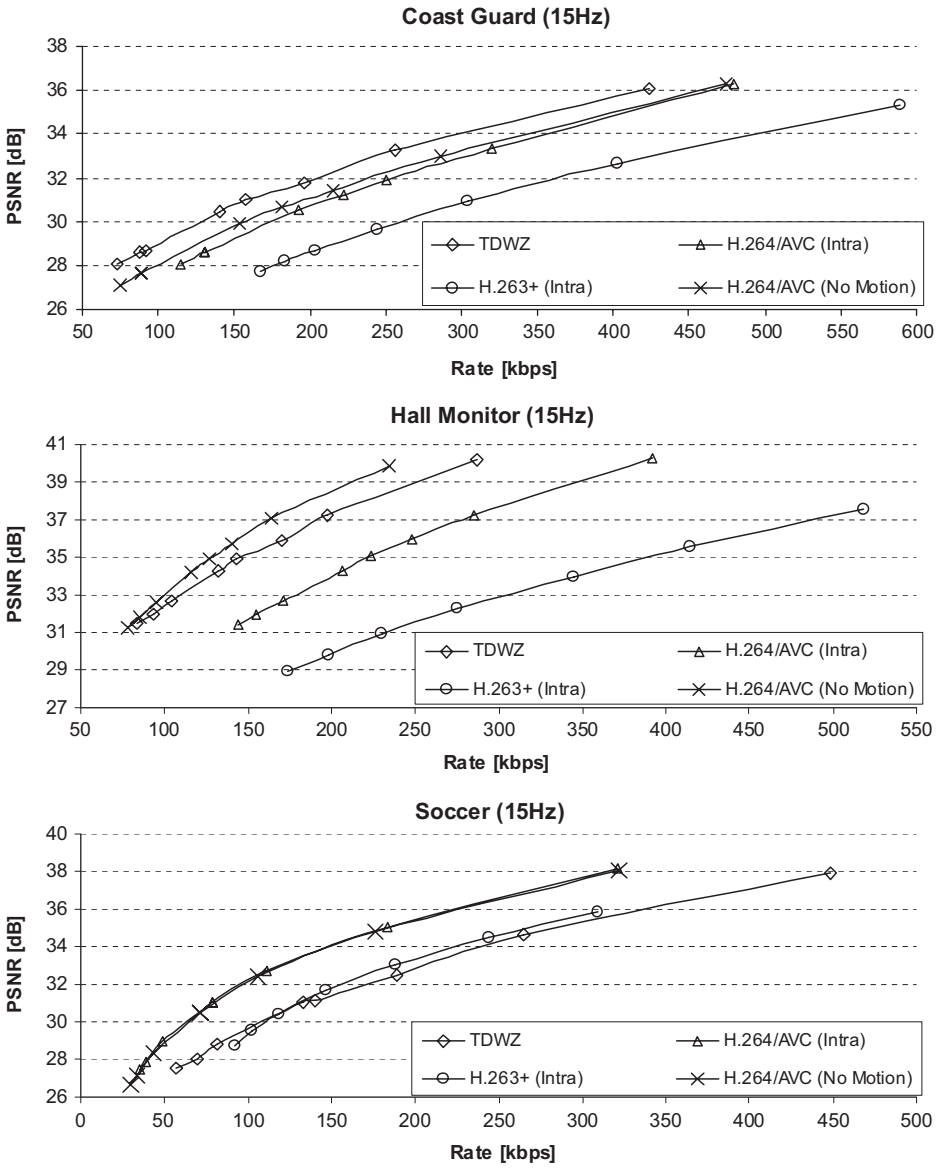

Fig. 5. RD performance for Coast Guard (QCIF at 15 Hz), Hall Monitor (QCIF at 15 Hz), Soccer (QCIF at 15 and 30 Hz), and Foreman (QCIF at 15 and 30 Hz), for GOP 2.

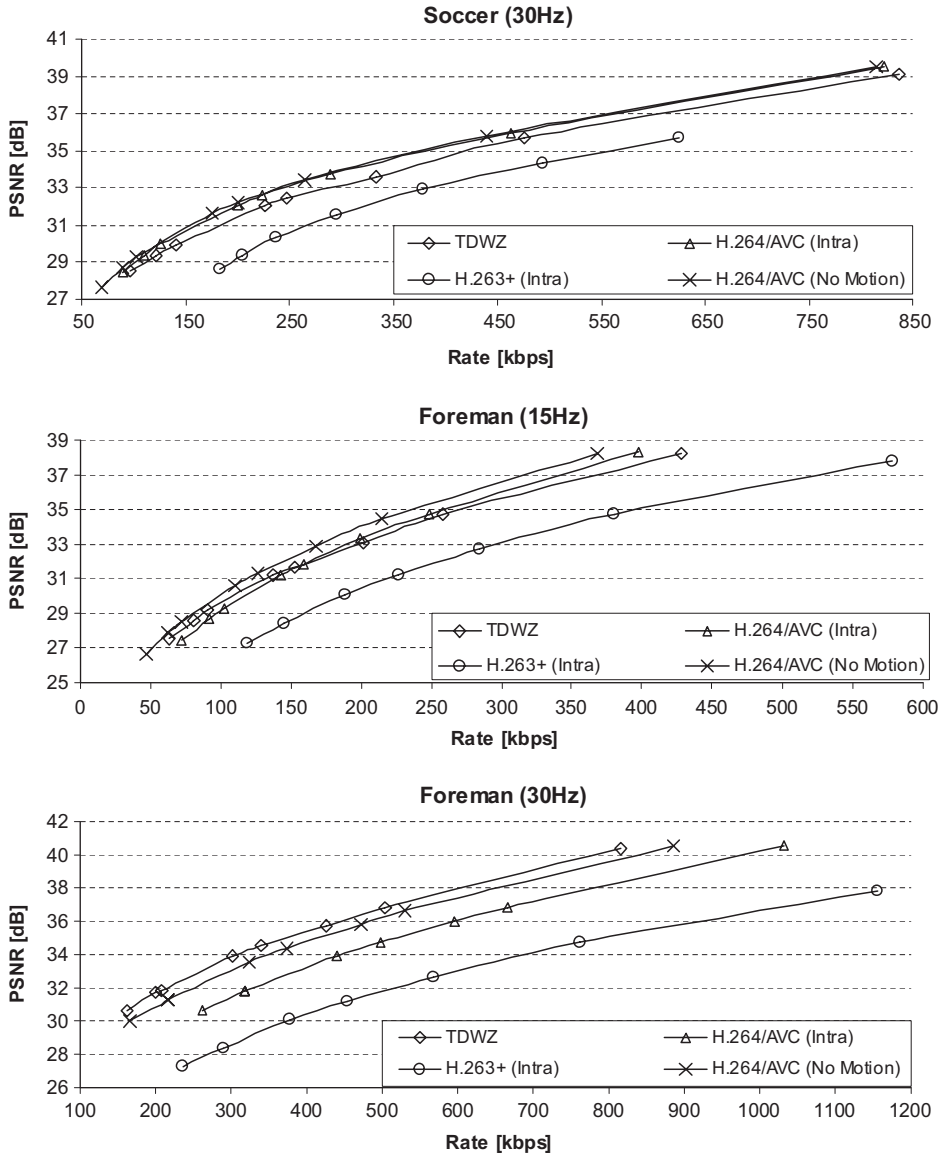

Fig. 5. (*Continued*)

- At 30 Hz, this means with frames closer in time and thus with less complex motion between frames, the TDWZ codec can outperform all codec competitors for the Foreman sequence (and also Coast Guard and Hall Monitor), but it still does not manage to defeat any H.264/AVC codec for the Soccer sequence which is clearly a though sequence for WZ video coding. However, at 30 Hz, the TDWZ codec clearly defeats the H.263+ Intra codec for any sequence.

Finally, Fig. 6 shows the comparative RD performance for various GOP sizes, highlighting that:

- The TDWZ RD performance is still, typically, the best for the smallest GOP size, this means GOP 2. This exposes the lack of capability of the motion interpolation process to create side information which is good enough when the key frames are more distant, especially if the motion content is less well behaved.

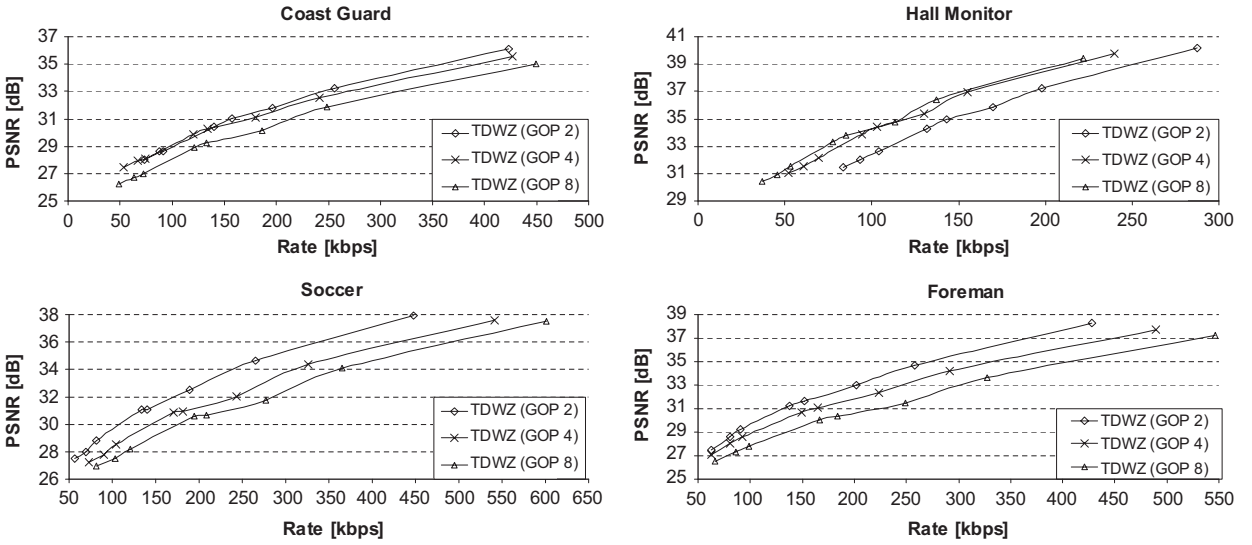

Fig. 6. RD performance for Coast Guard, Hall Monitor, Soccer and Foreman for GOP sizes 2, 4, and 8 (QCIF at 15 Hz).

- For very stable content, e.g. from video surveillance, like the Hall Monitor sequence the best TDWZ RD performance is achieved for a GOP size of 8 (very similar to a GOP size of 4). This highlights again that, if the motion interpolation manages to model better the motion, even when the key frames are more distant, better side information is produced, and thus better RD performance is achieved for longer GOP sizes. This points towards a very important WZ video coding research direction which is motion modeling for side information generation in order to obtain more efficient RD performance [11] for longer GOP sizes. Another possible improvement is to have the encoder using Intra coding modes for the parts of the frame where there is less temporal correlation and thus the frame interpolation process cannot produce good side information.

While the overall RD performance results may not seem as good as in other papers in the literature, attention should be paid to the test conditions and anchors, e.g. not always QCIF at 30 Hz, not always comparing with H.263+ Intra, and not only low motion (and thus well temporal correlated) sequences. Notice that if RD results were shown only for QCIF at 30 Hz in comparison with H.263+ Intra, the TDWZ video codec would clearly win for all sequences and RD points tested.

4.2. Measuring the quality evolution of WZ decoded frames

Since the WZ (parity) bits are requested to improve the quality of the side information, and thus to obtain a higher 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; this will have a significant impact on the decoder complexity since the turbo decoder has to run after each request. The algorithm to obtain the WZ decoded frames quality as the number of requests increases is presented in the following:

1. After a given chunk of parity bits is received, the turbo decoder decodes the current bitplane for the current band;
2. After the tDec 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 [5]:
   (a) If $P_e > 10^{-3}$, the decoder requests for more parity bits from the encoder, and returns to step 1;
   (b) If $P_e \leqslant 10^{-3}$, the current bitplane tDec task is considered successful and the tDec of the next bitplane starts in step 1. If there are no more bitplanes and bands for the current frame, the decoding of the next WZ frame starts.

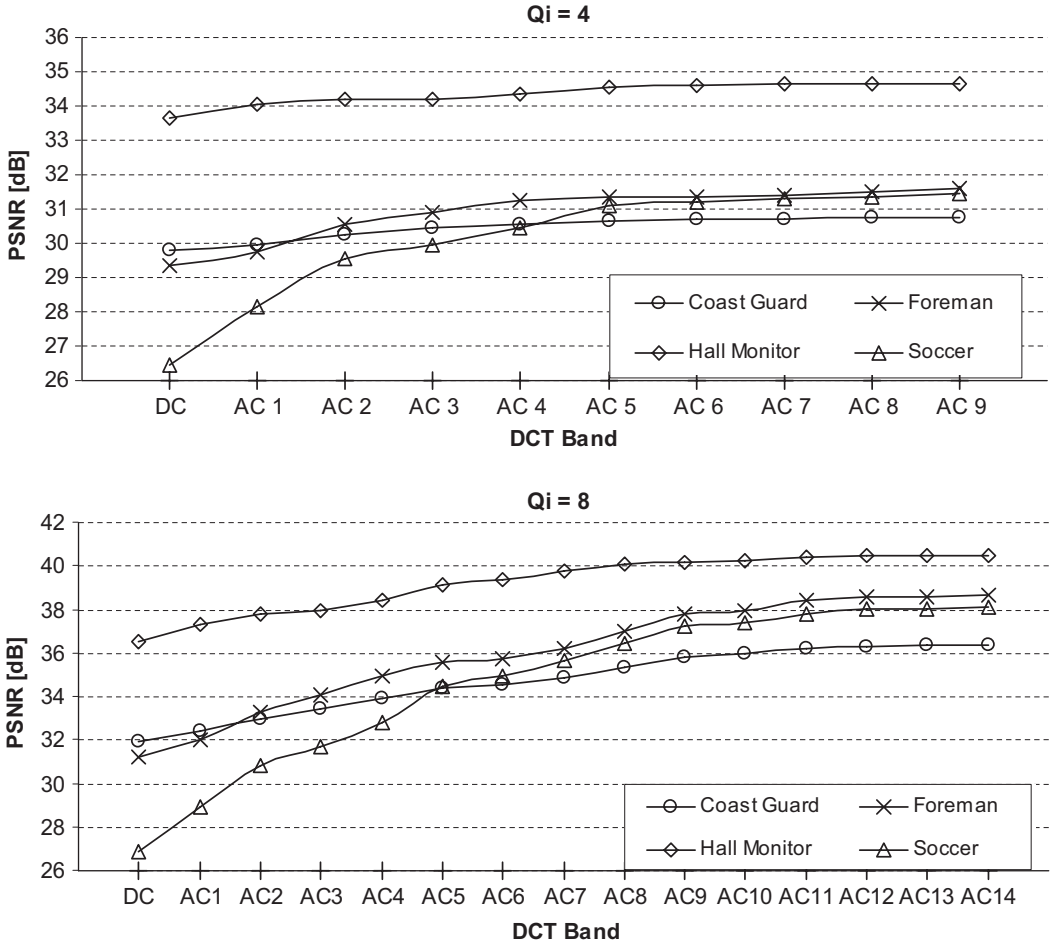

Fig. 7. WZ decoded quality evolution with the number of decoded DCT bands for Coast Guard, Hall Monitor, Soccer, and Foreman sequences for $Q_i = 4$ and 8 (QCIF at 15 Hz).

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

Fig. 7 shows the evolution of the average decoded PSNR as a function of the DCT band for $Q_i = 4$ and 8, highlighting the intrinsic quality scalability provide by the TDWZ video codec: the more DCT coefficients are decoded, the higher the decoded quality. It is important to notice that, for different frames of the same sequence, the same DCT band does not have to correspond to the same number of requests since this depends on the side information quality, which mainly depends on the motion content. Notice that the side information quality is not the same for the two Q_i because the quality of the key frames changes for the various Q_i. The main conclusions are:

- As expected, the more complex the sequences, the lower the quality of the side information and the higher the PSNR improvement with the WZ bits.
- The higher is the Q_i, the higher the target quality and thus more bands and bitplanes for each band will be WZ decoded and more requests will be made (according to Fig. 4). Notice that the higher is the band and the lower is the bitplane, typically, the lower the correlation between the side information and the original frame.
- The more textured and motion complex is the video sequence, the higher will be the PSNR improvement for the higher Q_i since detailed texture; for example, the water in Coast Guard, is very hard to estimate well. This type of PSNR improvement is typically

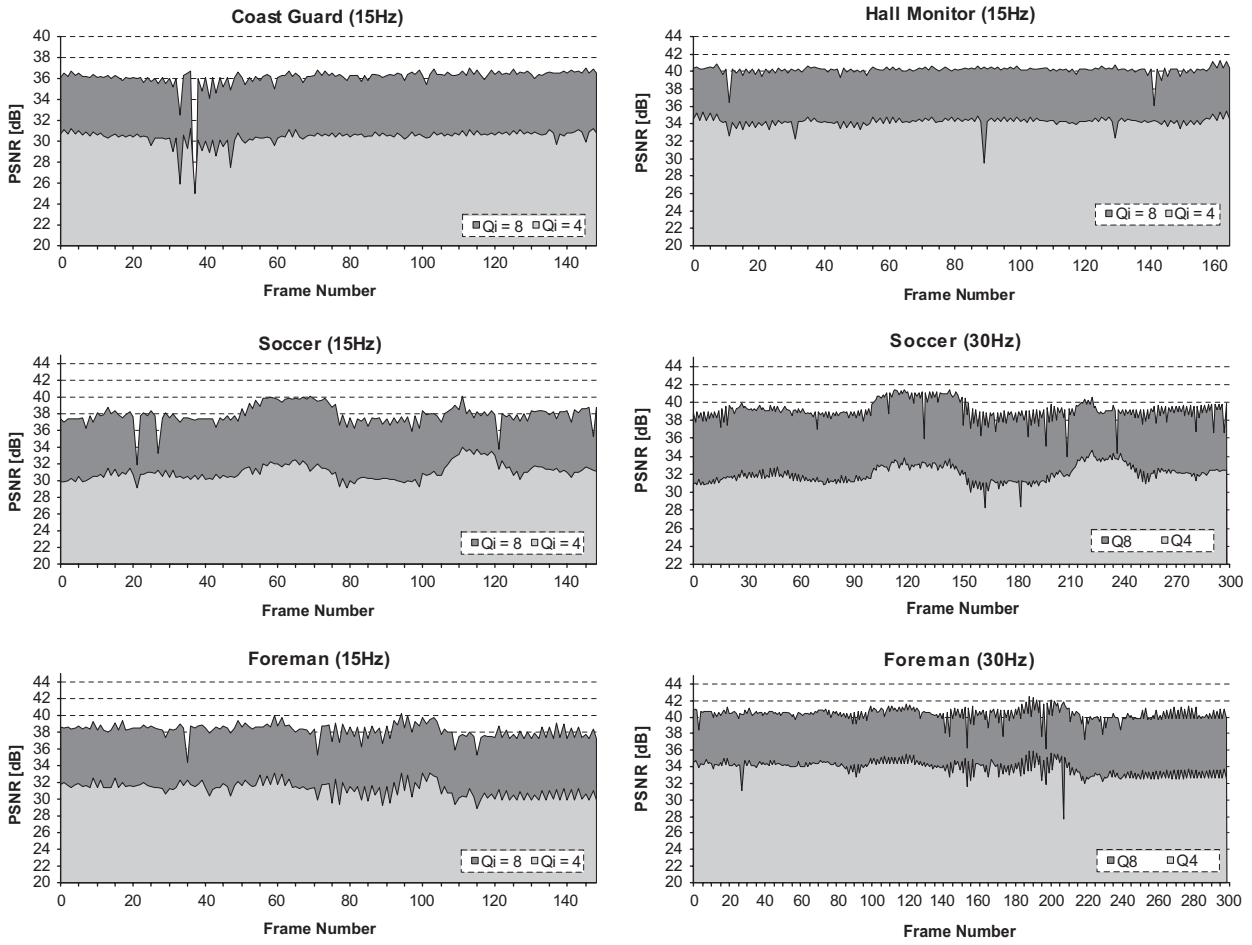

Fig. 8. PSNR temporal evolution of the decoded frames for Coast Guard (15 Hz), Hall Monitor (15 Hz), Soccer (15 and 30 Hz), and Foreman (15 and 30 Hz), for $Q_i = 4$ and 8.

rather costly in bitrate because it regards the coding of information for which the correlation is rather small; thus, the WZ coding approach is rather inadequate justifying other coding options for this information such as DCT Intra and entropy coding [19]. By definition, the WZ approach is only efficient when there is high correlation between side information and original frames (for high motion and complex texture, this correlation may be quite low); otherwise, it becomes rather inappropriate and thus inefficient which means that efficient encoders should, in principle, intelligently perform mode decision between the WZ and conventional approaches. Some initial work was already done in the context of adaptive GOP size [4] choosing which frames should be Intra coded (the remaining are WZ), and intra mode decision at the block level, i.e. deciding which blocks should be Intra coded [21] in a WZ frame.

Fig. 8 shows the PSNR temporal evolution for the four test sequences for $Q_i = 4$ (medium bitrate and thus medium quality) and $Q_i = 8$ (high bitrate and thus high quality). The charts show that:

- The quality variations in the decoded frame quality are, in general, rather small, since the QP for the Intra frames was manually adjusted to have an overall constant quality.
- However, since the QP selection was not done locally but rather on average for all frames, maintaining a fixed QP over all the sequence does not guarantee a fully constant PSNR. When the motion interpolation fails, and thus the side information quality is poorer, the bands and bitplanes which are not WZ coded for a certain Q_i remain with poor side information estimation, and thus with a significant amount of errors left to correct, which causes a lower decoded quality

when compared to the adjacent key frames and the average PSNR of the whole sequence. This effect is notorious for the Coast Guard sequence when a strong tilt-up occurs (around frame 35) and for the Foreman sequence when a pan-right occurs (around frames 70–100 for 15 Hz). For the Soccer sequence, the side information generation fails more often and, therefore, more variations in the decoded frame quality occur for the same reason.

4.3. Measuring the bitplane compression factor

As described in Section 2.1, the turbo encoder encloses two 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 that, for some specific cases, 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 since the compression factor would be lower than 1. The total average compression factor at frame, CF_Q, and bitplane, CF_{Qij}, levels for a certain quality rank, Q, is given by Eqs. (4) and (5):

$$CF_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N}\frac{C_{ijl}}{w_{ijl}}}{N}, \tag{4}$$

$$CF_{Qij} = \frac{\sum_{l=1}^{N}\frac{C_{ijl}}{w_{ijl}}}{N}, \tag{5}$$

where M_i is the number of bitplanes of band i, N is the total number of WZ frames, C_{ijl} is the number of bits in each original coefficient bitplane j of each band i at frame l 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 Eq. (5) represents the average compression factor at bitplane j of band i for a certain quality rank Q.

Fig. 9 shows the average bitplane compression factor per band for $Q_i = 4$ (medium bitrates) and $Q_i = 8$ (high bitrates) for the four test video sequences at 15 Hz (similar results were obtained for 30 Hz). The following conclusions can be inferred:

- In a general way, the least significant bitplanes (LSBs) have lower compression factors when compared to the most significant bitplanes (MSB and MSB-1). This is because the number of errors in the side information increases when the bitplane number index increases (i.e. for LSB), especially for the bands for which WZ bits are sent for a high number of bitplanes (e.g. DC and AC1-2). Note that it is possible that the compression factor is lower than 1 for the LSBs since their correlation is lower. In Fig. 9, which presents average results for the whole sequences, this happens with the Soccer sequence for the seventh bitplane of the DC coefficient with $Q_i = 8$; however, at frame level, this happens for more cases.
- In general, for the AC bands, the highest compression factor is achieved for the MSB-1 bitplane. Since the MSB-1 and MSB bitplanes are quite correlated and that correlation is exploited during the tDec operation (see Section 2.1), it is possible to achieve higher compression factors. The correlation between consecutive bitplanes decreases as the bitplane significance level decreases and, therefore, lower compression ratios are achieved.
- The last AC bands usually have high compression ratios for both bitplanes since they only have four wide bins (two bitplanes) and, therefore, the correlation between the corresponding DCT bands of the side information and WZ frames is high and large compression ratios can be achieved. When more bitplanes are considered in each band, the number of bins increases (and thus they are smaller) and more errors between the side information and the WZ occur (i.e. the mismatch between the corresponding quantization DCT coefficients bins of the side information and the WZ frames increases), decreasing the compression factor.

The very low, in fact sometimes lower than 1, compression factors achieved for the LSB with less correlation indicate the need for better coding solutions, notably: (i) usage of conventional DCT Intra coding for the bands and bitplanes with low correlation such as in the PRISM codec [19]; or (ii) usage of channel codes which guarantee that compression factors lower than 1 are not achieved such as the LDPC codes [22].

4.4. Measuring the decoded quality versus the side information

This section evaluates the PSNR improvement obtained after WZ decoding with respect to the side information. This improvement is defined as the

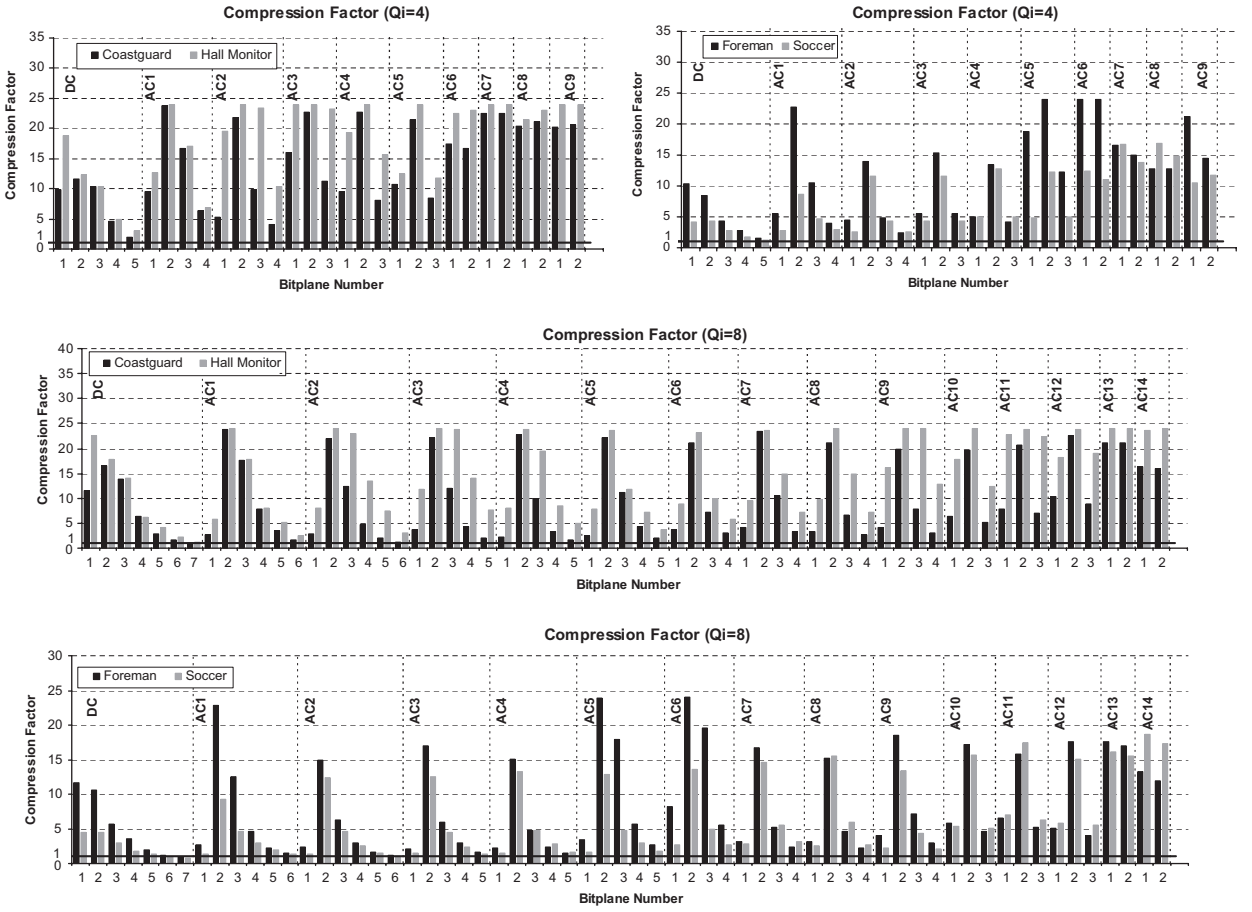

Fig. 9. Bitplane compression factor for Coast Guard, Hall Monitor, Soccer, and Foreman sequences, for $Q_i = 4$ and 8 (QCIF at 15 Hz).

difference between the PSNR of the WZ frames and the PSNR of the corresponding side information. While it is obvious that this gain is always positive, it is important to understand how big these quality gains are and for which conditions.

Fig. 10 shows the decoded PSNR versus the side information PSNR for all the WZ frames of the four test video sequences selected, for $Q_i = 4$ (medium bitrates) and $Q_i = 8$ (high bitrates). The following conclusions can be taken:

- As expected, the quality of the WZ decoded frames is always above the quality of the side information which means that the parity bits sent always improve the quality of the side information.
- The ideal behavior for the WZ decoded PSNR plots would be a horizontal line with all points lying on it; this would mean that, independently of the side information quality (i.e. on how good the frame interpolation is), the decoded frame quality would remain always similar. Of course, this "ideal line" would lie on different positions (i.e. with different decoded PSNR) depending on the Q_i used. Therefore, for $Q_i = 4$, since some bands/bitplanes correspond to the side information (no WZ bits are sent to those bands/bitplanes), a drift from the expected behavior is expected. On the other hand, for $Q_i = 8$, since most of the bands/bitplanes are corrected, the WZ decoded quality is more constant.
- Another factor that influences these results is the amount of errors in the side information; if the side information has high quality, the variations in the decoded frame quality would be minimal (since there would be fewer errors in the bands/bitplanes that are not WZ decoded). On the other hand, if some side information frames have a higher amount of errors, they will be propagated for some bands/bitplanes of the WZ decoded

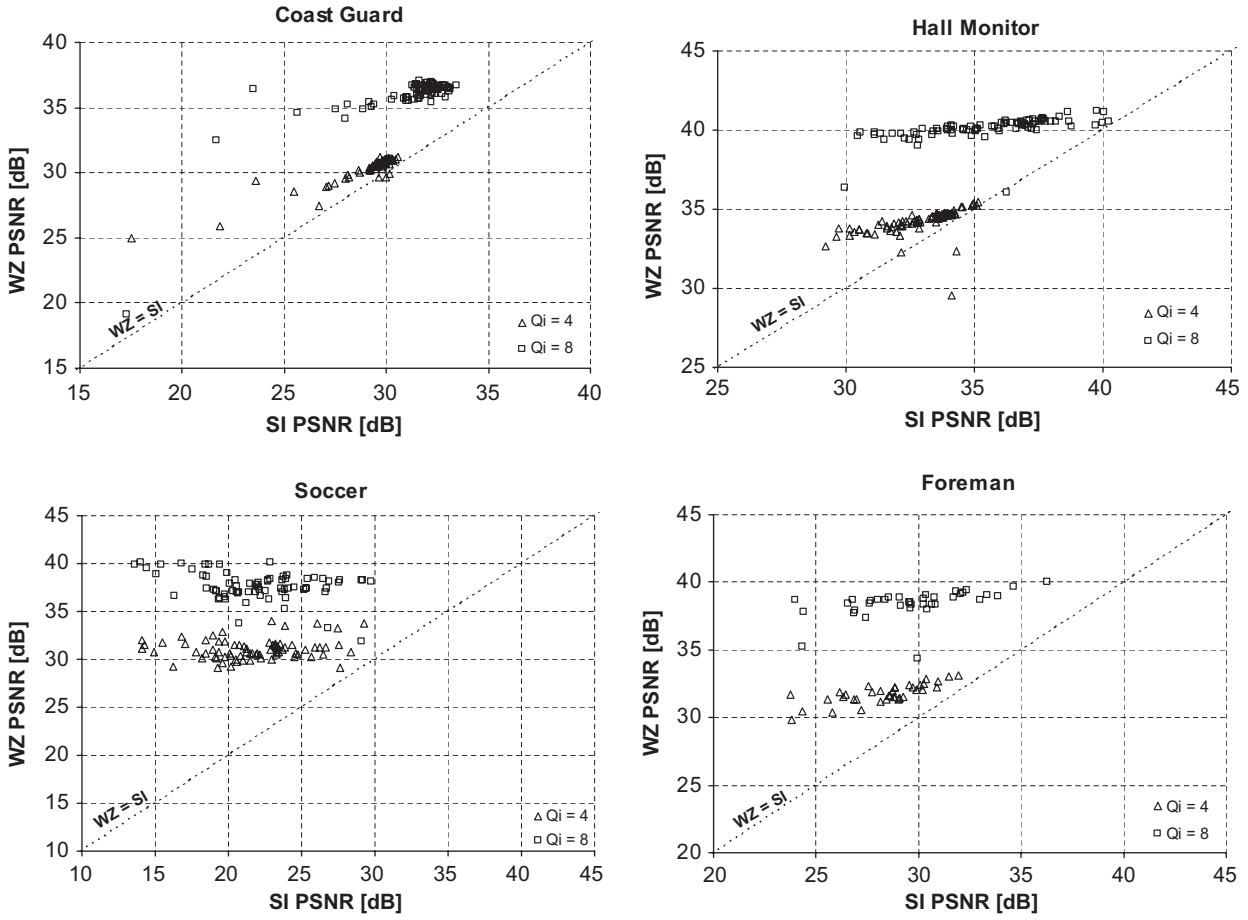

Fig. 10. WZ frames PSNR versus side information PSNR for $Q_i = 4$ and 8, for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz).

frames and will cause variations in the decoded quality (as observed mainly for the Soccer and Foreman sequences).

Fig. 11 and Table 3 show now the average WZ decoded PSNR versus the average side information PSNR for the four test video sequences selected for the eight Q_i tested. The main conclusions are:

- The more complex is the motion in the scene sequence, the higher the PSNR improvements regarding the side information. For very stable sequences, like Hall Monitor, the side information has a rather high quality and thus the gap to the target PSNR is lower, reducing significantly the observed PSNR improvements.
- For the same sequence, the higher is the frame rate, the lower the PSNR improvement since the better is the side information. When the temporal gap between the reference frames used for motion interpolation decreases, the side information tends to have a better quality.
- The PSNR improvements regarding the side information increase with the RD point meaning that the quality of the side information does not increase in the same way that the final target quality increases; the higher is the Q_i the higher is the PSNR improvement since WZ bits are sent for most of the bands/bitplanes of the side information.
- The less textured is the sequence, the less is the bitrate needed to achieve a certain PSNR improvement.

In general, the results show the importance of the SIC process, which strongly determines the RD performance of the codec.

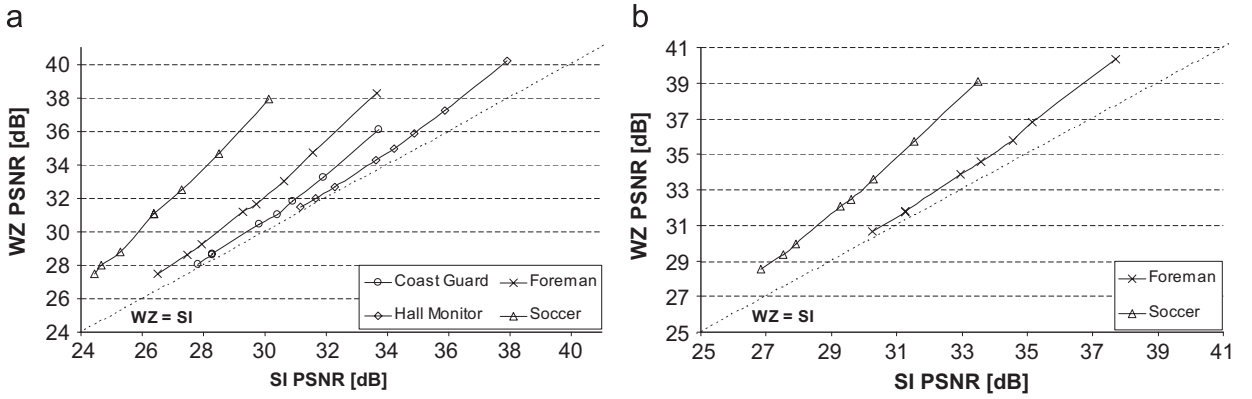

Fig. 11. Overall WZ quality versus overall side information quality for (a): Coast Guard, Hall Monitor, Soccer, and Foreman sequences at 15 Hz and (b) Soccer and Foreman sequences at 30 Hz.

Table 3
PSNR improvement (dB) regarding the side information and corresponding WZ bitrate

Q_i	Coast Guard (QCIF at 15 Hz)		Hall Monitor (QCIF at 15 Hz)		Soccer (QCIF at 15 Hz)		Foreman (QCIF at 15 Hz)		Soccer (QCIF at 30 Hz)		Foreman (QCIF at 30 Hz)	
	PSNR gain (dB)	WZ rate (kbps)	PSNR gain (dB)	WZ rate (kbps)	PSNR gain (dB)	WZ rate (kbps)	PSNR gain (dB)	WZ rate (kbps)	PSNR gain (dB)	WZ rate (kbps)	PSNR gain (dB)	WZ rate (kbps)
1	0.21	15.18	0.28	11.03	3.05	38.95	1.00	26.87	1.71	51.18	0.41	30.21
2	0.31	21.96	0.34	15.50	3.34	50.28	1.16	35.00	1.86	66.83	0.48	40.04
3	0.36	25.83	0.38	17.96	3.55	56.88	1.28	39.36	2.03	76.96	0.55	48.63
4	0.60	43.78	0.65	28.08	4.67	93.77	1.94	65.68	2.79	125.87	0.92	81.86
5	0.64	45.56	0.72	30.56	4.72	100.75	1.97	72.18	2.84	134.69	0.99	89.00
6	0.89	70.65	0.95	44.93	5.20	132.61	2.38	101.49	3.30	187.55	1.20	127.38
7	1.34	95.14	1.38	54.42	6.15	172.27	3.15	133.15	4.17	243.25	1.67	170.34
8	2.37	182.09	2.27	90.35	7.79	286.02	4.59	227.83	5.62	424.43	2.68	298.08

5. Feedback channel performance evaluation

In the adopted TDWZ video coding architecture, the feedback channel has the role to adapt the used WZ bitrate to the changing statistics between the side information (an estimation of the frame to be encoded) and the WZ frame to be encoded, i.e. to the quality (or accuracy) of the frame interpolation process [3]. 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.

Since this decoder rate control operation based on the feedback channel is central in the TDWZ architecture, notably determining the decoder complexity, it is important to be aware of its behavior and impact in order to design more efficient WZ video coding solutions. Thus, in the following subsections, some feedback channel relevant metrics will be defined and analyzed.

5.1. Measuring the number of requests

During the decoding of a given bitplane of a given band b_k, the decoder may send a request to the encoder one or more times asking 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 corresponding DCT bands of 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 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 from Eqs. (6) and (7):

$$D_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N} r_{ijl}}{N}, \tag{6}$$

$$D_{Qij} = \frac{\sum_{l=1}^{N} r_{ijl}}{N}, \tag{7}$$

where r_{ijl} is the number of requests made via the feedback channel for bitplane j of band i at WZ frame l; N is the total number of WZ frames coded, M_i is the number of bitplanes of band i and B_Q is the number of bands for a certain quality rank Q. D_{Qij} (Eq. (7)) is a partial result of Eq. (6) representing the average number of decoder requests per frame for bitplane j of band i and for quality rank Q.

Fig. 12 shows the average number of requests per bitplane per band, for the four selected video sequences (QCIF at 15 Hz), for GOP 2; the DCT bands are separated by the vertical dashed lines. Results are provided for the 4th and the 8th RD points, represented by $Q_i = 4$ and 8, respectively. The higher is Q_i, the higher the bitrate and the quality. For other Q_i values, the metrics behavior is the same as for the Q_i values used here. Note that the AC bands are numbered in a zigzag scanning order. The following conclusions may be drawn:

- The main reason for the evolution of the number of decoder requests is the amount of correlation between the side information and the (quantized) WZ frame; the higher is the correlation, the less parity bits are requested by the decoder.
- In many cases, the decoder makes less requests for the second bitplane of each band than for the first (MSB), due to a strong correlation with the first bitplane. Since the correlation between consecutive bitplanes is explored in the tDec operation (Section 2.1), it is possible to reduce the number of decoder requests for the second bitplane when compared to the MSB bitplane number of requests. However, that correlation between consecutive bitplanes becomes weaker as the number of bitplanes to be decoded for a given band increases, and more parity bits are needed.
- The highest number of requests happens for the LSB of the lower frequency bands since there are more bitplanes to decode there (due to the smaller quantization bin size) and the correlation between consecutive bitplanes is lower (Section 4.3), leading to a higher number of requests. For the higher frequency bands, the number of requests does not increase much simply because less bitplanes are coded (due to the larger quantization bin size). At 30 Hz, the conclusions are similar.

An efficient way to reduce the number of requests, and thus also the feedback channel rate and the decoding complexity, is by having the encoder making a conservative estimation of the number of bits needed to correct each bitplane, for each band [14]. In this case, the decoder needs to request fewer parity bits, reducing the number of requests, the delay involved and also the decoder complexity; note that the turbo decoder has to be run for each additional set of WZ bits received. Another possibility is to have an encoder rate control technique where the encoder estimates the number of bits necessary for successful decoding and sends them at once to the decoder without making use of the feedback channel [5]; however, since at the encoder the side information is not available, the encoder may under estimate or over estimate the necessary bitrate (regarding the decoder rate control case) resulting in video artifacts or some RD performance loss.

5.2. Measuring the feedback channel rate

After the average number of requests per band and bitplane is known, it is possible to estimate the feedback channel rate for each band and bitplane. For this, it is assumed in a rather simplistic way 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

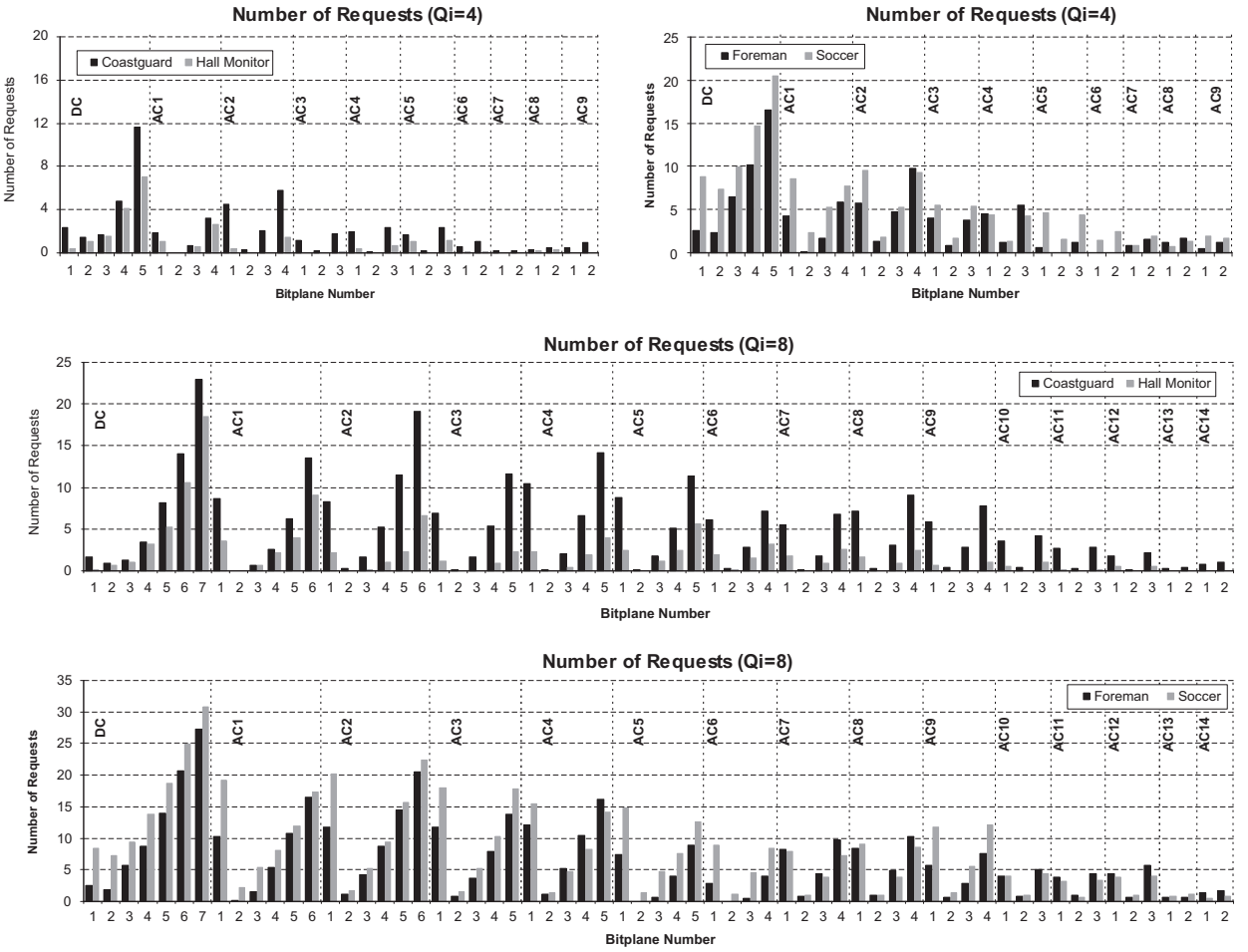

Fig. 12. Number of requests per bitplane per band for $Q_i = 4$ and 8, for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz).

for a certain quality rank, Q, can be obtained from Eqs. (8) and (9), respectively.

$$R_Q = \frac{\sum_{i=1}^{B_Q}\sum_{j=1}^{M_i}\sum_{l=1}^{N} n_{ijl}}{N} \times f \tag{8}$$

$$R_{Qij} = \frac{\sum_{l=1}^{N} n_{ijl}}{N} \times f \tag{9}$$

In Eqs. (8) and (9), f is the WZ frame rate and n_{ijl} is the number of bits sent via the feedback channel for bitplane j of band i at WZ frame l; N is the total number of WZ frames, M_i is the number of bitplanes of band i and B_Q is the number of bands considering a certain quality rank, Q. R_{Qij} is a partial result of Eq. (8) representing the average feedback channel rate per frame for bitplane j of band i for a certain quality rank Q. If more bits have to be sent for each request, it is easy to scale the rate computed above by the convenient factor.

Table 4 shows the total feedback rate for the selected test sequences for each Q_i. It can easily be seen that the feedback rate is rather negligible; the maximum feedback channel rate for each sequence happens for $Q_i = 8$ and ranges, at 15 Hz, from about 400 bps for Hall Monitor to about 1900 bps for Soccer. As could be expected, Soccer needs more requests because has worse side information due to the complex and erratic motion.

Regarding the relative distribution of the feedback channel rate by bitplane and DCT band, this can be seen in Fig. 12 since the feedback channel rate corresponds to the number of requests per bitplane multiplied by the number of frames per second. As it can be observed, the number of bits sent through the feedback channel, for each band,

increases with the number of bitplanes to be decoded; the reduced correlation between the side information and the WZ frame is the main reason for that behavior.

5.3. Measuring the number of errors versus number of requests

Since WZ coding is very much about correcting 'errors' in the side information estimation, it is interesting to evaluate the amount of requests needed to correct a certain amount of 'errors'. With this purpose in mind, this section evaluates the number of errors corrected with the number of requests made for each bitplane.

Fig. 13 shows the average number of errors versus the average number of requests per bitplane for $Q_i = 4$, for the four selected QCIF video sequences.

Table 4
Overall feedback channel rate (bps)

Q_i	QCIF at 15 Hz				QCIF at 30 Hz	
	Coast Guard	Hall Monitor	Soccer	Foreman	Soccer	Foreman
1	74.75	42.78	252.84	161.99	305.10	147.25
2	121.79	72.78	333.88	217.59	415.10	213.80
3	125.94	66.46	358.58	226.34	441.80	228.70
4	209.09	90.81	580.62	373.18	707.64	375.55
5	197.39	84.58	606.87	398.63	721.34	378.00
6	350.83	157.49	817.21	583.02	1055.14	598.79
7	516.78	210.53	1098.45	803.06	1436.33	886.44
8	1124.15	432.91	1905.31	1467.88	2707.77	1752.43

It can be noticed that:

- The number of errors increases as the bitplanes significance becomes lower since the correlation between the corresponding bitplanes of corresponding DCT bands of the side information and the WZ frames becomes weaker.
- Moreover, the lower is the correlation, the higher the amount of parity bits necessary for successful decoding. For example, for the Soccer sequence, the number of decoder requests is higher when compared to the Coast Guard sequence since each bitplane of the side information has a higher number of errors when compared to the Soccer sequence due to the lower quality side information.
- The highest number of requests for the highest number of errors happens for the last bitplane of the DC band in the Soccer sequence, where about 20 requests are needed to correct about 500 errors (32% of errors). This is expected since the correlation between the original frames and the corresponding side information is lower for the least significant bitplanes of the lower frequency coefficients; note that these coefficients have more bitplanes to code due to the smaller quantization bin size.

5.4. Measuring the number of requests versus side information quality

Since the parity bits are successively requested to correct the side information errors and improve the decoded quality, it is important to know how the

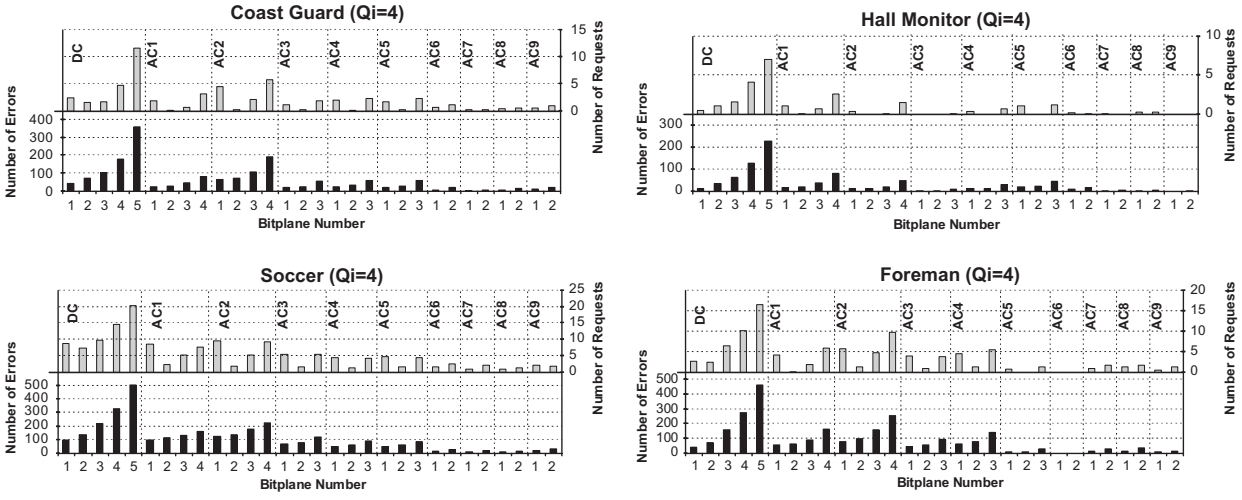

Fig. 13. Number of errors versus number of requests at the bitplane level for $Q_i = 4$, for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz).

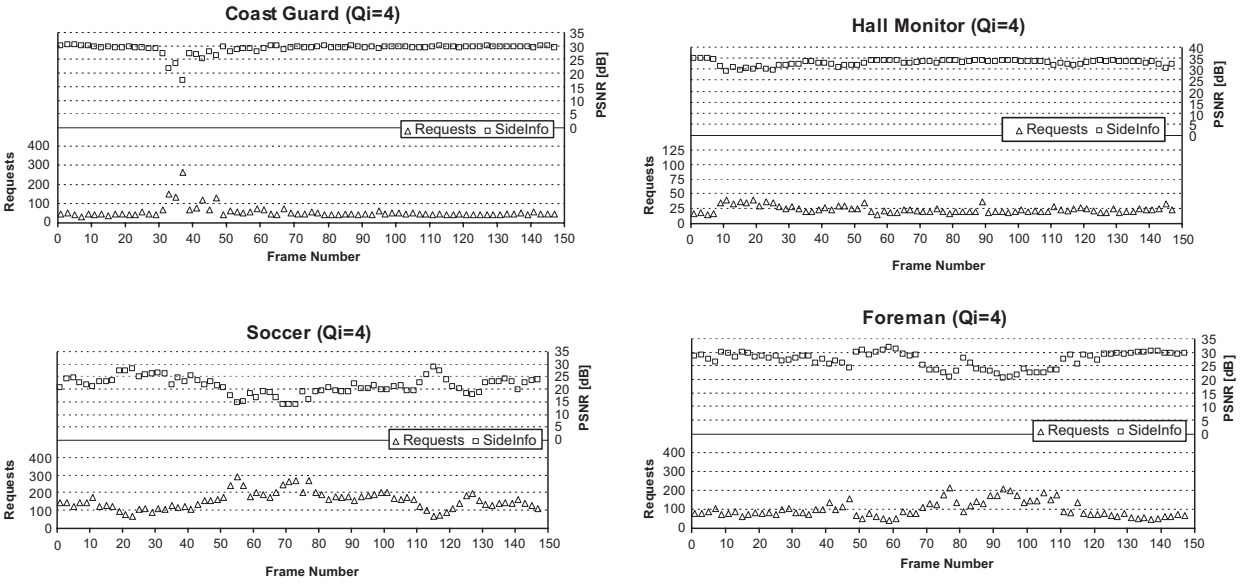

Fig. 14. Number of requests versus side information quality for $Q_i = 4$, for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz).

number of requests varies with the side information quality. This will allow designing more adequate request strategies since the number of requests has a significant impact on the decoder complexity.

Figs. 14 and 15 show the number of requests versus the side information quality at the frame level for $Q_i = 4$ and 8, for the four selected video sequences. It can be observed that:

- The maximum number of decoder requests is around 900 for $Q_i = 8$. This happens for the Coast Guard sequence (around frame 35) where a strong tilt-up occurs.
- For video sequences characterized by well defined motion, like the Hall Monitor sequence, the frame interpolation algorithm employed at the decoder generates high quality side information, i.e. few errors exist between the side information and the original WZ frames. Since there are few errors to be corrected by the turbo decoder, only a few decoder requests are needed. This concept is confirmed in Fig. 14 along time: the higher the side information PSNR, the lower the number of decoder requests.
- The inconstant and complex motion that characterizes the Foreman and Soccer sequences justifies the fluctuations on the side information quality and the corresponding fluctuations on the number of decoder requests per frame.

6. Complexity performance evaluation

Because evaluating the RD performance addresses only one side of the codec evaluation problem, this section intends to perform an evaluation of the complexity performance for the TDWZ video codec. Although it is commonly claimed that WZ video encoding complexity is 'low' and WZ video decoding complexity is 'high', not much solid and exhaustive complexity evaluation results are available in the literature. This section intends to make some steps further in bringing clarification to these issues.

6.1. Encoding complexity

This section targets the WZ video encoding complexity evaluation. The encoder complexity includes two major components which are the key frames and the WZ frames coding parts. The larger is the GOP size the less key frames are present in the bitstream and thus the lower will be the share of the key frames.

While it is possible to measure the encoding complexity in many ways, some of them rather sophisticated, it is also possible to get a rather good estimation of relative complexities using rather simple complexity metrics. In this paper, the encoding complexity will be measured by means of

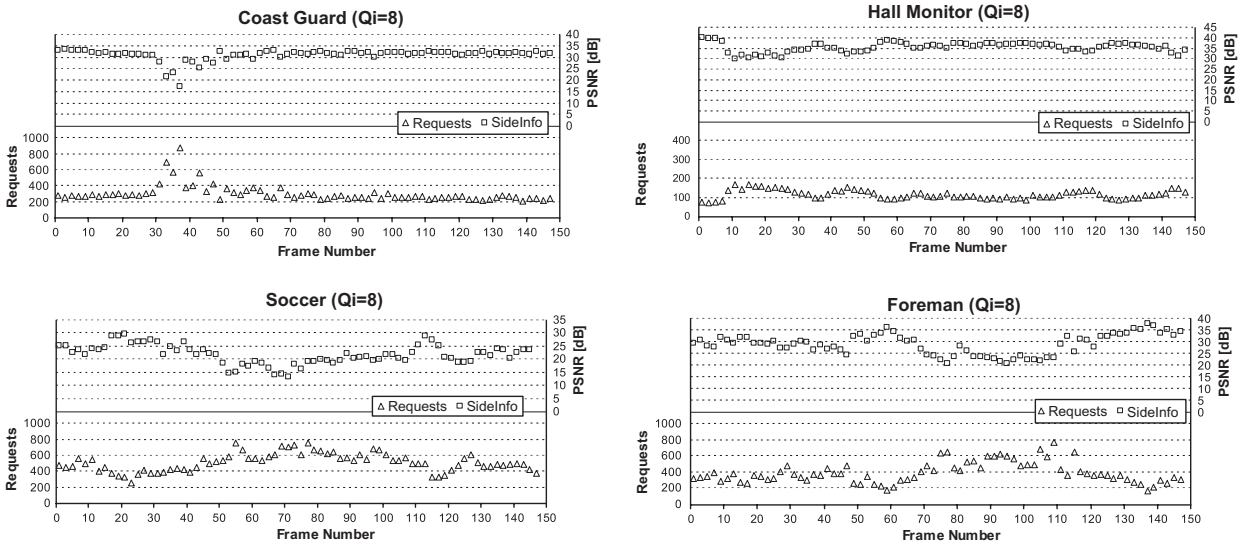

Fig. 15. Number of requests versus side information quality for $Q_i = 8$, for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz).

the encoding time for the full sequence, in seconds, under controlled conditions. It is well known that the encoding (and decoding) times are highly dependent on the used hardware and software platforms. For the present results, the hardware used was an x86 machine with a dual core Pentium D processor at 3.4 GHz with 2048 MB of RAM. Regarding the software conditions, the results were obtained with a Windows XP operating system, with the C++ code written using version 8.0 of the Visual Studio C++ compiler, with optimizations parameters on, such as the release mode and speed optimizations. Besides the operating system, nothing was running in the machine when gathering the performance results to avoid influencing them. Under these conditions, the results have a relative and comparative value, in this case allowing comparing the TDWZ codec with alternative solutions, e.g. H.264/AVC, running in the same hardware and software conditions. While the degree of optimization of the software has an impact on the running time, this is a dimension that was impossible to fully control in this case and thus will have to be kept in mind when dealing with the performance results.

Fig. 16, Tables 5 and 6 show the encoder complexity results for GOP 2 measured in terms of encoding time, distinguishing between the key frames (dotted pattern bars) and WZ frames (diagonal pattern bars) encoding times. The results allow concluding that:

- The TDWZ encoding complexity is always much lower than the H.264/AVC encoding complexity, both for the H.264/AVC Intra (about 60–70% higher complexity) and H.264/AVC No Motion solutions (for GOP 2, similar to H.264/AVC Intra).
- While the H.264/AVC Intra encoding complexity does not vary with the GOP size and the H.264/AVC No Motion encoding complexity is also rather stable with a varying GOP size, the TDWZ encoding complexity decreases with the GOP size. If encoding complexity is a critical requirement, the results in this section together with the RD performance results previously shown indicate that the TDWZ codec with GOP 2 is already a rather credible practical solution since it has a rather low complexity and defeats H.264/AVC Intra in terms of RD performance for most content cases.
- For the TDWZ codec, the WZ encoding complexity is negligible when compared to the key frames encoding complexity, even for GOP 2. Although not shown in the charts, for longer GOP sizes, the overall encoding complexity would decrease with the increase of the WZ frames share since the key frames share decreases, although their encoding complexity is

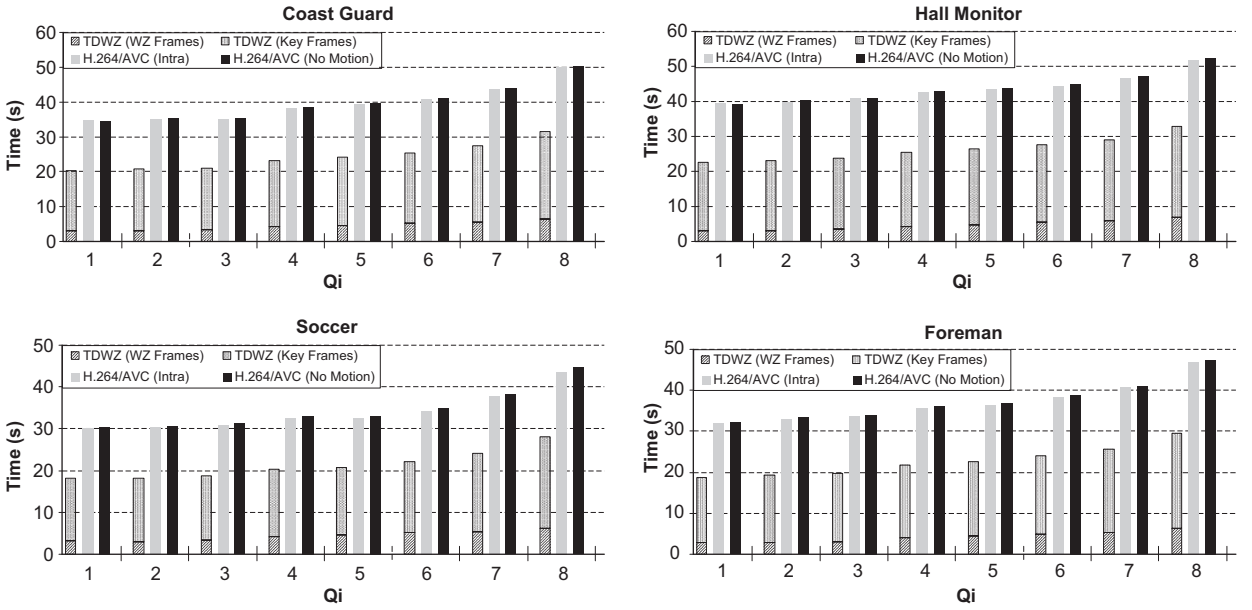

Fig. 16. Encoding complexity measured in terms of encoding time for Coast Guard, Hall Monitor, Soccer, and Foreman sequences (QCIF at 15 Hz, GOP 2).

Table 5
Encoding time (s) comparison between the TDWZ and H.264/AVC codecs for the Coast Guard and Hall Monitor sequences (QCIF at 15 Hz, GOP 2)

Q_i	Coast Guard				Hall Monitor			
	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	Ratio Intra/TDWZ	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	Ratio Intra/TDWZ
1	34.73	34.73	20.36	1.71	39.48	39.36	22.73	1.74
2	35.38	35.38	20.72	1.71	40.05	40.27	23.05	1.74
3	35.47	35.47	20.98	1.69	40.84	40.97	23.73	1.72
4	38.53	38.53	23.19	1.66	42.67	42.95	25.58	1.67
5	39.72	39.72	24.22	1.64	43.55	43.72	26.37	1.65
6	41.23	41.23	25.45	1.62	44.64	44.89	27.72	1.61
7	43.94	43.94	27.37	1.61	46.64	47.14	29.03	1.61
8	50.50	50.50	31.45	1.61	51.95	52.39	32.81	1.58

still the predominant part. This means that if higher quality side information can be produced for longer GOP sizes using more advanced techniques, the TDWZ encoding complexity would be further reduced without a significant RD performance penalty.

- The TDWZ encoding complexity does not increase significantly when the Q_i increases (i.e. when the bitrate increases); typically, a 50% complexity increase exist from the first to the last RD point.
- Finally, the TDWZ encoding complexity is rather similar for the various video sequences, independently of their complexity.

6.2. Decoding complexity

This section targets the WZ video decoding complexity evaluation. Again, the decoder complexity includes two major components which are the key frames and the WZ frames decoding parts. The larger the GOP size, the fewer the key frames

Table 6
Encoding time (s) comparison between the TDWZ and H.264/AVC codecs for the Soccer and Foreman sequences (QCIF at 15 Hz, GOP 2)

Q_i	Soccer				Foreman			
	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	Ratio Intra/ TDWZ	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	Ratio Intra/ TDWZ
1	30.13	30.45	18.06	1.67	31.88	32.14	18.70	1.70
2	30.33	30.63	18.11	1.67	32.92	33.27	19.31	1.70
3	30.81	31.25	18.72	1.65	33.59	33.94	19.78	1.70
4	32.39	33.06	20.36	1.59	35.56	36.02	21.78	1.63
5	32.42	33.14	20.63	1.57	36.39	36.77	22.62	1.61
6	34.23	34.80	22.11	1.55	38.30	38.78	24.03	1.59
7	37.77	38.34	24.09	1.57	40.58	40.81	25.56	1.59
8	43.72	44.70	28.02	1.56	46.78	47.28	29.56	1.58

Table 7
Decoding time (s) comparison between the TDWZ and H.264/AVC codecs for the Coast Guard and Hall Monitor sequences (QCIF at 15 Hz)

Q_i	Coast Guard			Hall Monitor		
	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2
1	1.50	1.53	295.15	1.77	1.66	240.67
2	1.55	1.53	408.77	1.79	1.68	322.45
3	1.58	1.55	475.67	1.80	1.69	371.00
4	1.66	1.66	794.65	1.85	1.72	564.22
5	1.69	1.70	820.48	1.86	1.74	605.84
6	1.69	1.73	1257.99	1.89	1.75	869.16
7	1.78	1.81	1668.90	1.94	1.79	1055.59
8	1.92	2.02	3090.83	2.06	1.91	1712.81

present in the bitstream; therefore, the lower the complexity associated to the key frames. Following the options used for the encoding complexity evaluation, the decoding complexity evaluation will be measured using an equivalent metric for the decoder, this means the decoding time for the full sequence, in seconds, under the same conditions and software/hardware platform.

Tables 7 and 8 show the decoder complexity results for GOP 2, measured in terms of decoding time. No charts are presented here for the decoding time because only the WZ decoding times would be visible since they are significantly higher than the H.264/AVC decoding times. The following conclusions can be inferred from the results:

- The TDWZ decoding complexity is always much higher than the H.264/AVC decoding complexity, both for the H.264/AVC Intra and H264/AVC No Motion solutions.
- While the H.264/AVC Intra decoding complexity does not vary with the GOP size and the H.264/AVC No Motion decoding complexity is also rather stable with a varying GOP size, the TDWZ decoding complexity increases with the GOP size (since there are more WZ frames to decode).
- For the TDWZ codec, the key frames decoding complexity is negligible regarding the WZ frames decoding complexity, even for GOP 2 (when there are as many key frames as WZ frames). This confirms the well known WZ coding trade-off where the encoding complexity benefits are paid with an increased decoding complexity.
- The TDWZ decoding complexity increases significantly when Q_i increases (i.e. when the bitrate

Table 8
Decoding time (s) comparison between the TDWZ and H.264/AVC codecs for the Soccer and Foreman sequences (QCIF at 15 Hz)

Q_i	Soccer			Foreman		
	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2	H.264/AVC Intra	H.264/AVC No Motion, GOP 2	TDWZ GOP 2
1	1.44	1.44	723.66	1.54	1.44	508.56
2	1.45	1.47	897.62	1.57	1.53	634.59
3	1.47	1.47	1010.78	1.59	1.55	709.72
4	1.50	1.52	1651.63	1.64	1.56	1178.80
5	1.52	1.56	1771.77	1.66	1.64	1290.25
6	1.57	1.58	2316.64	1.70	1.66	1802.88
7	1.67	1.69	2949.72	1.75	1.75	2328.32
8	1.81	1.83	4775.63	1.90	1.92	3873.75

increases) since the number of bitplanes to turbo decode is higher and the turbo decoder (and the number of times it is invoked) is the main responsible for the high decoding complexity.

Although the TDWZ decoding time is extremely high regarding the H.264/AVC decoding times, it is possible to reduce these times in several ways: (i) optimization of the turbo decoder software; (ii) estimation at the encoder of a conservative number of bits to be initially sent for each bitplane of each band, reducing significantly the number of times the turbo decoder has to be run (but not significantly affecting the RD performance); and (iii) improving the side information quality to reduce the number of errors to be corrected and thus the number of requests and number of WZ bits needed. Of course, if no feedback channel exists, and a pure encoder rate control solution is used, the decoding complexity will be much smaller since no requests are made.

As mentioned in Section 1, the motion estimation task is the main responsible for the high encoder complexity in predictive video coding. In the same way, and since most the complexity has been moved to the decoder, it is important to know how the main decoder modules contribute to the TDWZ decoding complexity, notably the SIC module, which includes motion estimation, as well as the correlation noise modeling (CNM), the tDec and the Rec modules. From Table 9 which shows the percentage of the decoding time (for the full sequence) associated with the SIC, CNM, tDec, and Rec modules for the eight RD points previously defined, the following conclusions may be derived:

- The results in Table 9 reveal that the most significant complexity burden is associated to the turbo decoder and the repetitive request-decode operation (95.08% and 98.59% of the TDWZ decoding time, at maximum, for the lowest and the highest RD points, respectively), typical of the WZ video codec architecture adopted. This highlights how important it is to reduce the number of decoder requests, not only by improving the side information quality, but mainly by adopting adequate rate control strategies such as efficient hybrid encoder–decoder rate control [14].
- As observed and expected, the SIC time share decreases when Q_i increases since the number of bitplanes to turbo decode is higher; on the contrary, the tDec time increases with Q_i when compared to the remaining modules time share. At maximum, the SIC time corresponds to 5.25% and 0.76% of the TDWZ decoding time, for the lowest and the highest RD points, respectively.
- Comparing RD points with the same number of WZ coded DCT bands, e.g. RD points pair (1, 2) and the triplet (6, 7, 8) (see Section 3), the higher is the RD point index, the higher the tDec time percentage since for that pair/triplet the number of bitplanes to turbo decode increases with the RD point index. This tDec time percentage increase comes along with a decrease in the CNM time percentage (for the same RD points). There are, however, two RD points (3 and 5) for which the tDec time percentage slightly decreases when compared to the corresponding previous RD point, i.e. RD points 2 and 4. On the RD points transitions 2 to 3 and 4 to 5, the number of DCT bands WZ coded increases by three (Section 3) and, since the CNM is performed at the DCT coefficient level for each DCT band

Table 9
Percentage of the decoding time associated with the side information creation (SIC), correlation noise modeling (CNM), turbo decoding (tDec) and reconstruction (Rec) modules for the tested sequences (QCIF at 15 Hz, GOP 2)

Q_i	Coast Guard (%)				Hall Monitor (%)				Soccer (%)				Foreman (%)			
	SIC	CNM	tDec	Rec	SIC	CNM	tDec	Rec	SIC	CNM	tDec	Rec	SIC	CNM	tDec	Rec
1	3.85	7.36	87.41	**0.03**	**5.25**	**10.78**	82.01	0.02	1.56	2.78	**95.08**	0.02	2.22	4.06	92.88	0.01
2	2.73	5.31	90.82	0.02	3.94	8.09	86.52	0.04	1.26	2.24	95.95	0.01	1.80	3.29	94.23	0.02
3	2.35	6.50	89.96	0.03	3.45	10.46	84.47	0.04	1.12	2.62	95.66	0.02	1.61	3.96	93.65	0.02
4	1.43	4.85	92.69	0.03	2.28	8.41	87.68	0.05	0.69	1.90	96.88	0.01	0.98	2.91	95.40	0.02
5	1.41	5.09	92.33	0.05	2.11	8.64	87.53	0.06	0.64	1.94	96.84	0.01	0.89	2.84	95.49	0.02
6	0.92	3.46	94.70	0.03	1.48	6.27	90.81	0.04	0.50	1.52	97.45	0.01	0.65	2.10	96.60	0.02
7	0.69	2.66	95.89	0.02	1.22	5.26	92.24	0.03	0.39	1.24	97.90	0.01	0.50	1.67	97.26	0.02
8	0.38	1.53	97.58	0.01	**0.76**	**3.39**	94.88	**0.02**	0.25	0.82	**98.59**	0.01	0.30	1.06	98.21	0.01

(Section 2.3), the CNM time percentage also increases, as expected, causing a reduction in the tDec time percentage. In those RD point transitions, the CNM time increase is more significant than the tDec time increase due to the DCT bands that exist in RD points 3 and 5 and do not exist in RD points 2 and 4, respectively.

- According to Table 9, and as expected, the time percentage associated with the Rec operation is rather negligible when compared to the remaining modules in the WZ decoder. Although it could be expected the SIC decoding time share to follow after the (dominating) tDec share, in practice it is the CNM share that always comes in second.

7. Final remarks

This paper presents a detailed performance evaluation of an advanced feedback channel and turbo coding based WZ video codec considering several types of metrics, notably in terms of rate, quality and complexity. This exhaustive evaluation, not yet available in the literature, allows not only to identify the strengths of this type of WZ video coding, e.g. the low encoding complexity, but also its weaknesses, notably the still much to improve RD performance, especially for longer GOP sizes, and the high decoding complexity. This evaluation is not only an important benchmarking for researchers in the field, since it was performed under very clear and precise conditions, but it is also relevant as a steering factor since it allows identifying WZ coding problems in need for an effective solution.

Finally, it is important to highlight that WZ video coding already proposes rather competitive solutions for application scenarios where encoding complexity is the main critical requirement since its RD performance is already the best for rather low encoding complexity, notably for more stable content, e.g. video surveillance material, in comparison with alternative standards based solutions.

References

[1] A. Aaron, R. Zhang, B. Girod, Wyner-Ziv coding of motion video, in: Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, November 2002.
[2] J. Ascenso, F. Pereira, Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding, in: International Conference on Image Processing, San Antonio, TX, USA, September 2007.
[3] J. Ascenso, C. Brites, F. Pereira, Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding, in: 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, June 2005.
[4] J. Ascenso, C. Brites, F. Pereira, Content adaptive Wyner-Ziv video coding driven by motion activity, in: International Conference on Image Processing, Atlanta, USA, October 2006.
[5] C. Brites, F. Pereira, Encoder rate control for transform domain Wyner-Ziv video coding, in: IEEE International Conference on Image Processing, San Antonio, TX, USA, September 2007.
[6] C. Brites, F. Pereira, Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding, IEEE Transactions on Circuits and Systems for Video Technology, in press.
[7] C. Brites, J. Ascenso, F. Pereira, Improving transform domain Wyner-Ziv coding performance, in: IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, May 2006.

[8] C. Brites, J. Ascenso, F. Pereira, Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding, in: International Conference on Image Processing, Atlanta, USA, October 2006.
[9] C. Brites, J. Ascenso, F. Pereira, Feedback channel in pixel domain Wyner-Ziv video coding: myths and realities, in: 14th European Signal Processing Conference, Florence, Italy, September 2006.
[10] B. Girod, A. Aaron, S. Rane, D. Rebollo Monedero, Distributed video coding, Proc. IEEE 93 (1) (January 2005) 71–83.
[11] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, J. Ostermann, Distributed monoview and multiview video coding, IEEE Signal Process. Mag. 24 (5) (September 2007) 67–76.
[12] Information Technology—Coding of Audio-Visual Objects—Part 10: Advanced Video Coding, ISO/IEC Std 14496-10, 2003.
[13] Video coding for low bitrate communication, ITU-T Recommendation H.263 Version 2, 1998.
[14] D. Kubasov, C. Guillemot, A hybrid encoder/decoder rate control for Wyner-Ziv video coding with a feedback channel, in: IEEE International Workshop on Multimedia Signal Processing, Crete, Greece, October 2007.
[15] D. Kubasov, J. Nayak, C. Guillemot, Optimal reconstruction in Wyner-Ziv video coding with multiple side information, in: IEEE International Workshop on Multimedia Signal Processing, Crete, Greece, October 2007.
[16] J. Pedro, L. Ducla Soares, C. Brites, J. Ascenso, F. Pereira, C. Bandeirinha, S. Ye, F. Dufaux, T. Ebrahimi, Studying error resilience performance for a feedback channel based transform domain Wyner-Ziv video codec, in: Picture Coding Symposium, Lisbon, Portugal, November 2007.
[17] F. Pereira, J. Ascenso, C. Brites, Studying the GOP size impact on the performance of a feedback channel-based Wyner-Ziv video codec, in: IEEE Pacific Rim Symposium on Image Video and Technology, Santiago, Chile, December 2007.
[18] S.S. Pradhan, J. Chou, K. Ramchandran, Duality between source coding and channel coding and its extension to the side information case, IEEE Trans. Inf. Theory 49 (5) (May 2003) 1181–1203.
[19] R. Puri, K. Ramchandran, PRISM: a new robust video coding architecture based on distributed compression principles, in: 40th Allerton Conference on Communication, Control and Computing, Allerton, USA, October 2002.
[20] J. Slepian, J. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inf. Theory 19 (4) (July 1973) 471–480.
[21] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, F. Pereira, Embedding a block-based intra mode in frame-based pixel domain Wyner-Ziv video coding, in: International Workshop on Very Low Bitrate Video, Sardinia, Italy, September 2005.
[22] D. Varodayan, A. Aaron, B. Girod, Rate-adaptive codes for distributed source coding, Signal Processing, Special Section: Distributed Source Coding 86 (11) (November 2006) 3123–3130.
[23] A. Wyner, Recent results in the Shannon theory, IEEE Trans. Inf. Theory 20 (1) (January 1974) 2–10.
[24] A. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Trans. Inf. Theory 22 (1) (January 1976) 1–10.