

RESEARCH

Open Access

Encoder-driven rate control and mode decision for distributed video coding

Frederik Verbist^{1,2*}, Nikos Deligiannis^{1,2}, Shahid M Satti^{1,2}, Peter Schelkens^{1,2} and Adrian Munteanu^{1,2}

Abstract

To provide low-complexity encoding for video in unidirectional or offline compression scenarios, this paper proposes an efficient feedback-channel-free distributed video coding architecture featuring a novel encoder-driven rate control scheme in tandem with a designated mode selection process. To this end, the encoder features a novel low-complexity motion estimation technique to approximate the side-information (SI) available at the decoder. Then, a SI-dependent correlation channel estimation between the approximated SI and the original frames is used to derive the theoretically required rate for successful Slepian-Wolf (SW) decoding. Based on the evaluation of the expected trade-off between the estimated required coding rate and the estimated distortion outcome, a novel encoder-side mode decision module assigns a different coding mode to distinct portions of the coded frames. In this context, skip, intra and SW coding modes are supported. To reduce the effect of underestimation, the final SW rate is adjusted upwards using a novel rate formula. Additionally, a successive SI refinement technique is exploited at the decoder to decrease the number of SW decoding failures. Experimental results illustrate the benefit of the different coding options and show similar or superior compression performance with respect to the feedback-based DISCOVER benchmark system. Finally, the low-complexity encoding characteristics of the proposed system are confirmed, as well as the beneficial impact of the proposed scheme on the decoding complexity.

Keywords: Distributed video coding; Feedback channel suppression; Encoder-driven mode decision

1 Introduction

The fundamental work of Slepian and Wolf [1] proved that separate lossless encoding but joint decoding of independently and identically distributed (i.i.d.) discrete random sources X and Y can be as efficient as joint encoding and joint decoding. The former setting is known as Slepian-Wolf (SW) coding or distributed source coding. In a particular case of the former scenario, called *asymmetric SW coding*, one source, e.g. Y , is compressed to its proper entropy while the other source, X , is compressed separately to the conditional entropy $H(X|Y)$. At the decoder, source Y is restored after which X is decoded in the presence of Y , called the side-information (SI). Extending the *asymmetric SW coding* setup, Wyner and Ziv [2] established the achievable lower rate bound under a distortion constraint when a single source is independently encoded but decoded in the presence of SI.

The Wyner-Ziv (WZ) theorem states that in such a coding scenario, a rate loss generally occurs compared to the setting where the encoder also has access to the SI.

However, a loss in compression performance is acceptable with respect to the benefit brought by adopting WZ coding. Independent encoding of information sources enables low-complexity encoding architectures since the removal of inter-source redundancy is no longer an encoder task. Instead, the encoding operation is essentially reduced to quantization followed by asymmetric SW encoding, usually implemented using channel encoding which is of low complexity. WZ coding found its application in coding data under severe resource constraints [3], e.g. in terms of computational power or energy supply. Specifically, distributed video coding (DVC) [4,5] ‘essentially WZ coding for video’ offers low-complexity encoding architectures. In DVC, complex operations, like motion estimation and compensation, are performed at the decoder to create SI. As a result, DVC targets lightweight multimedia applications [6], e.g. wireless capsule endoscopy [7,8].

* Correspondence: fverbist@etro.vub.ac.be

¹Electronics and Informatics Department, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium

²Future Media and Imaging Department, iMinds, Gaston Crommenlaan 8 (b102), Ghent 9050, Belgium

In practical WZ coding of video, rate control poses a major challenge. Namely, what is the required SW or channel rate to ensure successful channel decoding in the presence of the SI? Because of the distributed nature of a WZ coding system, the encoder has no access to the SI, since it is generated at the decoder. Hence, a WZ encoder is not in a position to determine the required channel rate exactly, since the conditional entropy $H(X|Y)$ cannot be measured directly. Solely duplicating the operations performed at the decoder would provide the encoder an identical copy of the SI. However, this would involve complex SI generation operations at the encoder, which would (1) compromise the low encoding complexity benefit of DVC and (2) rather favour a traditional predictive coding approach from a compression performance point of view.

The majority of high-performance DVC systems make use of a feedback channel to solve the rate control problem. Such an approach, often referred to as *decoder-driven* rate control, sends non-systematic information in chunks [5]. Should the channel rate prove insufficient for proper decoding, the decoder is able to request a larger amount of non-systematic information from the encoder and attempt decoding anew. The process is repeated until decoding proves successful. In this way, the presence of a feedback channel not only guarantees decoding success but also ensures that this is achieved at a minimal channel rate. However, it is evident that a feedback-channel-based rate control scheme is incompatible with unidirectional application scenarios. Moreover, decoder-driven rate control links the encoding and decoding process. Consequently, feedback-channel-based DVC is unsuitable for offline applications, e.g. storage purposes, and may demonstrate excessive delay [9].

In feedback-channel-free (alias unidirectional) DVC architectures, the encoder is responsible for determining the required channel rate for successful decoding, which is referred to as *encoder-driven* rate control. However, estimating the necessary channel rate at the encoder is a delicate problem; underestimation leads to a poor decoding of the source while overestimation results in wasted rate. Hence, such DVC systems suffer a performance loss with respect to feedback-based schemes. The encoder's main obstacle to determine the necessary rate to guarantee decoding is the lack of access to the SI. Instead, the latter is approximated at the encoder, where special care must be taken as not to compromise the low-complexity encoding characteristics. Moreover, feedback-channel-free systems may suffer from inflated decoding complexity [4,10].

This paper introduces a novel feedback-channel-free transform-domain WZ (TDWZ) video coding architecture^a. The core system is an efficient hash-based WZ video codec [11]. To avoid feedback, the proposed system creates an encoder-side approximation of the SI

using a novel technique that mimics the SI generation executed by the decoder, without undermining low-complexity encoding. Based on the correlation between the original frame and the estimated SI, a novel encoder-driven rate allocation scheme assigns an appropriate SW rate. To increase compression performance and reduce the effect of SW rate underestimation, the proposed architecture also features a novel encoder-driven mode decision process. If the quality of the corresponding SI is expected to be high, parts of the original frames may not be coded at all but rather skipped and reconstructed as the SI. Alternatively, conventional entropy (intra) coding may be applied when failure of proper SW decoding is likely or would result in severe distortion. At the decoder, a successive SI refinement scheme is exploited to minimize the distortion associated to SW rate underestimation. At every SI refinement stage, a higher-quality version of the SI is generated. This creates the opportunity to reattempt to decode any SW coded information that failed to decode properly at the previous refinement stages.

A version of the feedback-free DVC system proposed in this paper, excluding the encoder-driven mode decision process, was presented in [12]. The experimental results presented in this work illustrate the benefit of the different coding modes available to the proposed system and clarify their influence on the compression performance. Additionally, in contrast to [12], the experimental results include an analysis of the impact of the low-complexity SI approximation methods on the overall RD performance and the encoding complexity. The compression performance of the proposed feedback-free system is compared to a collection of alternative feedback-based DVC systems, including the benchmark DISCOVER [13] codec. The experimental results show that the proposed feedback-free architecture achieves similar or superior compression performance with respect to DISCOVER [13], which is noteworthy considering that most feedback-channel-free systems in the literature are significantly falling behind DISCOVER [14-16]. A last set of experiments confirms the low-complexity encoding characteristics and the beneficial effect of the proposed scheme on the decoding complexity.

The rest of this paper is organized as follows. Section 2 offers an overview of related work and highlights the novel features included in the proposed architecture. Section 3 details the proposed system broken down in its primary components. Experimental results are provided in Section 4, and finally, Section 5 concludes the paper.

2 Related work and contributions

2.1 Related work

2.1.1 Feedback-free DVC solutions

The unidirectional pixel-domain DVC codec in [17] used two parallel WZ encoders where the original image

is scattered by interleavers prior to encoding. The architecture was further enhanced [18] with an iterative decoding scheme where the SI was gradually updated using spatio-temporal predictions. In both schemes, the rate was an input parameter. An alternative encoder-driven rate control scheme for pixel-domain DVC was put forward in [16]. A coarse approximation of the SI is generated at the encoder by averaging the key frames in a group of pictures (GOP) of 2, after which the correlation noise is modelled by a zero-mean Laplacian distribution. The Laplacian correlation noise model serves as a basis to derive the bit-plane error probability, which is mapped to a bit-rate using functions trained offline. The probabilities are calculated without taking any previously decoded bit-planes into account. The final rate calculation was modified in [19], where the offline module was replaced by machine learning. Considering a multi-user scenario, the feedback channel was removed from a pixel-domain architecture in [20].

However, the compression performance of pixel-domain DVC lags behind that of TDWZ architectures [5]. In [21], a feedback-channel-free transform-domain architecture was designed, where a coarse version of the SI was generated by averaging the key frames in a GOP size of 2. Then, the SW rate is derived from the coarse SI based on empirical results obtained in offline experiments.

The first motion estimation algorithm to generate an approximation of the SI at the encoder was proposed in [14] and integrated in a TDWZ architecture. In essence, the algorithm constitutes a low-complexity variant of the motion-compensated interpolation (MCI) method employed by the decoder to generate SI. The technique performs MCI for a limited number of blocks based on the sum of absolute differences (SAD) criterion, while the SI approximation for the other blocks is the average of the co-located blocks in the reference frames. The resulting approximation of the correlation noise instantiates a Laplacian correlation noise model per frequency band, based on which a closed-form formula determines the required SW rate. The conditional error probabilities are computed as in [22], where any already decoded bit-planes are taken into account. The scheme was extended in [15], where additional care was taken to reduce the probability of failed channel decoding. When a bit-plane is not error-free after the maximum number of decoding runs has been reached, the log-likelihood ratios (LLRs) for the bits that are most likely to be erroneous are flipped, after which channel decoding is attempted anew.

In [23], further tools were introduced to increase the performance. At the encoder, the quantized symbols of original WZ frames as well as the coarse SI frames undergo Gray mapping [24] prior to SW coding. Additionally, an updated form of the closed-form formula in

[14] yields the estimated SW rate. At the decoder, the reconstruction [25] of the coefficients was modified to cope with any bit-planes of the quantization indices that failed to decode. The final reconstruction is the weighted sum of the centroids of every individual bin, where the weights are assigned according to the bit-plane LLRs after Turbo decoding [26]. Finally, [23] used a SI refinement stage after all frequency bands have been decoded based on overlapped block motion estimation (OBME). Based on the refined SI, new attempts are made to decode any erroneous bit-planes and reconstruction is performed again.

2.1.2 Coding modes and DVC architectures

When integrating multiple coding modes in DVC, a fundamental decision is the locality of the mode decision process, namely, at the encoder or at the decoder. Given the distributed nature of the system, optimized mode selection is challenged by the fact that the original signal is only available to the encoder, while the SI is only present at the decoder.

Regarding decoder-driven mode decision, a pixel-domain DVC architecture which skips bit-planes based on a rate-distortion model was presented in [27]. However, the skip mode only improved performance on low and medium motion sequences. The work in [28] includes a feedback-based TDWZ architecture with decoder-driven skip, intra and SW modes. The skip mode is selected based on the trade-off between rate and distortion derived from the SI and the virtual correlation channel. The decision whether to apply intra or SW coding is made by applying both modes to the co-located bit-plane in the previous decoded frame, after which the mode yielding the lowest rate is selected. In this way, the complexity of the decoder is increased since encoding and decoding is duplicated at the decoder.

A decoder-driven block-based mode decision scheme was proposed in [29]. By evaluating the linearity of the motion vectors, blocks with linear motion vectors were skipped, while blocks with highly non-linear motion vectors were supported with additional hash information to help improve the SI quality. Alternatively, the block-based DVC architecture in [28] allowed individual blocks to be skipped, intra or WZ coded, based on the estimated accuracy of the SI. This is achieved by assessing the mean squared error between the past and the future reference blocks. The decision between intra and WZ coding is determined on an RD basis, by selecting the mode with the lowest rate at equal distortion.

In all the above codecs, the outcome of the mode selection process must be signalled to the encoder via a feedback channel, rendering them unsuitable in a unidirectional context. An encoder-side mode selection approach was followed in PRISM [4]. A total of 16

different coding modes or classes are available for every 8×8 block, where each class corresponds to a different distribution of skipped, intra or SW coded bits. PRISM assigns coding modes based on thresholding the squared error difference between every block and its co-located block in the previous frame. In [30], blocks were coded either WZ or with a combination of WZ and intra coding. The encoder selects the mode that yields the lowest estimated rate. The hybrid coding mode sends a low-quality intra-coded version of the block, which helps to improve the quality of the SI. In [31], a block-based skip mode was integrated in a TDWZ codec. SW coding, using Turbo [26] codes, was used to compress the entire frame, where the Turbo decoder was modified as to cope with the skipped blocks. Both codecs in [30] and [31] employ a feedback channel for rate control.

2.2 Contributions

In contrast to existing systems, this is the first work to introduce an encoder-driven mode decision at the bit-plane and frequency band level in a feedback-free hash-based DVC system. To support unidirectional operation, the proposed codec includes several new features at the encoder and decoder.

First, the SI is approximated at the encoder by a low-complexity emulation of the hash-based OBME and compensation (OBMEC) technique used to generate the SI at the decoder. This approach conceptually differs from the fast MCI scheme used in [23], which coarsely matches the motion-compensated interpolation technique of [23] to generate SI at the decoder.

Second, the SI approximation at the encoder is used to compute the theoretical required SW rate, namely the conditional entropy $H(X|Y)$, to represent the quantized source X . The pursued approach varies from existing schemes by relying on a SI-dependent (SID) correlation channel [11,32] to capture the dependency between the coarse SI and the WZ frame. To limit the likelihood that bit-planes fail to decode due to SW rate underestimation, the final SW rate is adjusted using a novel formula that takes the significance of the bit-plane into account.

Third, based on the approximation of the SI and correlation channel model, a novel mode decision process is executed. Three coding modes are supported, namely, skip, intra and SW coding. The skip mode is applied per frequency band whereas intra and SW modes are assigned per bit-plane. Since the proposed WZ architecture is feedback-free, the full responsibility for selecting and signalling the coding modes falls to the encoder.

The proposed architecture also features specific measures at the decoder to reduce the suffered distortion due to any SW coded information that fails to decode properly. For this purpose, the principles of successive SI refinement [33] are adopted. The WZ frames are

decoded in distinct stages called refinement levels [34], where at each stage, a higher quality version of the SI is generated. Given the improved SI at every refinement level, SW decoding is reattempted for all SW coded information that failed to decode at previous levels, when only a poorer version of the SI was available. The proposed decoding process thereby merges SI refinement, SW decoding and reattempts at SW decoding of any SW coded information that failed to decode at previous refinement levels. In contrast, the approach in [23] first decodes an entire WZ frame, after which additional SI updates create new opportunities to attempt decoding bit-planes that failed to SW decode successfully.

Finally, the proposed feedback-free system is thoroughly evaluated. In this context, preliminary results of the proposed feedback-free architecture without any encoder-side mode decision were presented in [12]. The experimental results presented in this work, however, show the benefit of the different coding modes available to the proposed system and clarify their influence on the compression performance. Additionally, the compression performance of the proposed feedback-channel-free DVC architecture is compared to the benchmark systems in DVC, that is, the DISCOVER codec [13] and H.264/AVC Intra [35], as well as our previous hash-based DVC system with feedback from [7]. In addition, compression results obtained using the proposed system including the presented mode decision process but configured with decoder-driven feedback-channel-based rate allocation are included as well. The evaluation for a GOP size of 2, 4 and 8 shows comparable or superior performance compared to the DISCOVER [13] codec. Despite the additional tools, experimental results confirm that the proposed system maintains low encoding complexity.

3 Proposed feedback-free distributed video coding architecture

The block diagram of the proposed feedback-free WZ codec is presented in Figure 1.

3.1 The encoding procedure

Building on the architecture presented in [34], the encoder divides an input video sequence into GOPs. Every GOP contains a key frame I , which is coded using H.264/AVC Intra [35], and WZ frames X , which are WZ coded in the 4×4 discrete cosine transform (DCT) domain. For the latter purpose, the quantization matrices from [36,37] define uniform and double-deadzone quantizers for the DC and AC frequency bands, respectively. The resulting quantization indices are grouped per band and organized into bit-planes, ready for SW coding based on LDPCA [38] codes. Additionally, the encoder creates a hash frame for every WZ frame, according to

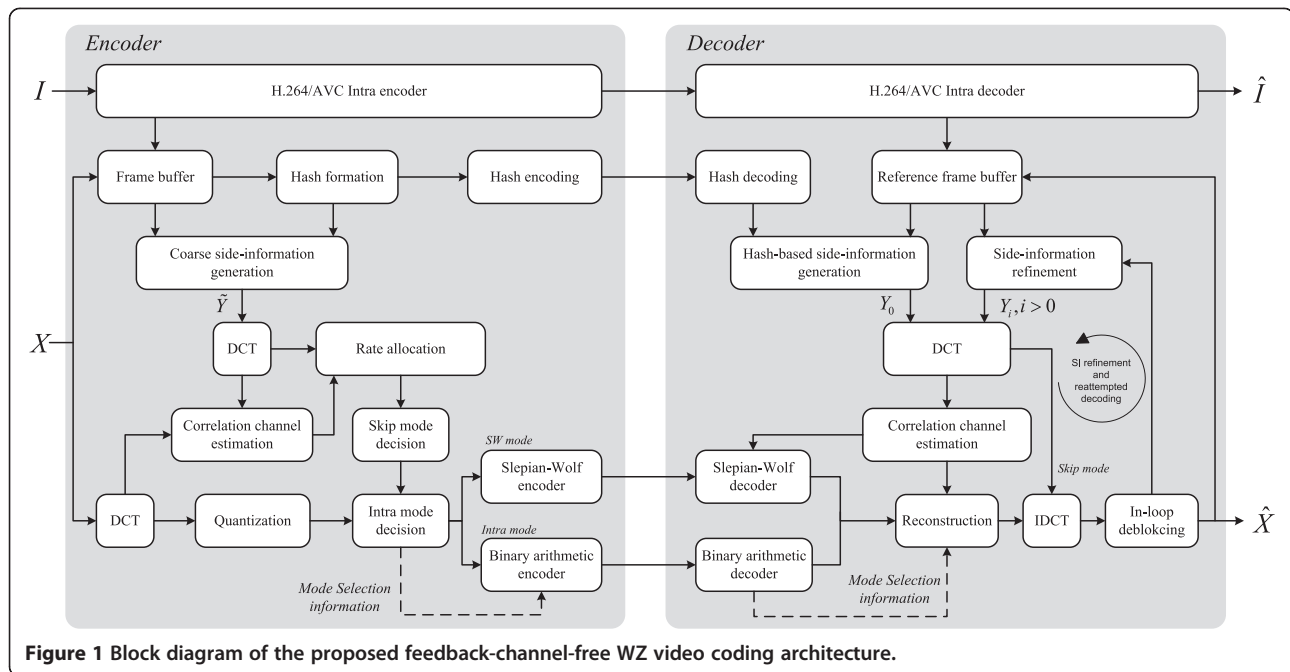


Figure 1 Block diagram of the proposed feedback-channel-free WZ video coding architecture.

the technique presented in [11,34] to enable hash-based SI generation at the decoder.

Since a feedback channel is not present, the encoder is forced to estimate the required channel rate for successful decoding per bit-plane in every frequency band. For this purpose, the encoder generates a coarse approximation of the SI available to the decoder using a low-complexity SI generation technique that emulates the hash-based OBMEC used at the decoder. Using the approximated SI, the encoder computes the theoretical rate, that is, the conditional entropy given the coarse SI, for every bit-plane in every quantized frequency band of the WZ frames. Finally, a rate formula is used to compute the final channel rate estimate from the conditional entropy in order to compensate for the mismatch between the SI estimated at the encoder and the real SI generated at the decoder.

Encoder-driven rate control is a sensitive process since underestimation leads to failed channel decoding and poorly reconstructed samples while overestimation wastes rate and no longer reduces the distortion level. To counter the effect of over- and underestimation on the overall RD performance, the proposed system includes additional tools.

To mitigate the effect of overestimation, the encoder first applies a band-level mode decision process, referred to as skip mode selection. Skipped DCT bands are not actually coded but are substituted by the corresponding band in the SI at the decoder. The bit-planes of the bands that are not skipped are additionally subjected to a second mode decision process. The encoder decides

whether a particular bit-plane is SW encoded or encoded in intra mode using a binary arithmetic entropy coder. When a specific coding mode has been assigned to every bit-plane, the bit-plane is fed to the appropriate encoder (unless the band the bit-plane belongs to is skipped). The resulting syndrome bits and binary arithmetic coded data are multiplexed with the hash bit-stream, as well as the mode signalling information, and sent to the decoder or stored for offline decoding.

3.1.1 Low-complexity side-information generation

Reference frame averaging is a simple low-complexity technique to generate a coarse SI signal at the encoder to approximate the true SI. The result may resemble the SI generated at the decoder rather well for low-motion sequences and small GOP sizes, e.g. GOP2. However, since mere averaging of the reference frames is incapable of capturing motion patterns, the SI estimate will significantly deviate from the true SI at the decoder when motion content or GOP size increase. Therefore, the proposed system features an alternative option to estimate the SI at the encoder, namely, a coarse approximation of the hash-based SI generation technique employed at the decoder. In other words, the encoder carries out a substantially simplified version of bidirectional OBMEC.

In detail, OBME is carried out on downscaled versions of the frames at the encoder in a hierarchical temporal prediction structure, similar to the prediction structure used in H.264/SVC [39]. Let $\xi = 2^k$, $k \in \mathbb{N}$ be the down-scaling factor applied at the encoder side, resulting in

frames with dimensions $W' = W/\xi$, $H' = H/\xi$. High downscaling factors ξ reduce the motion estimation complexity at the cost of reduced accuracy. Next, mimicking the SI generation process at the decoder [11,34], the encoder divides every downsampled WZ frame into overlapping blocks β of $B \times B$ pixels with an overlap step size of ε pixels, $1 \leq \varepsilon < B$.

For every such block, the best matching block is found in the reference frames R_n , $n \in \{0, 1\}$ within search range sr . To this end, the Hamming distance calculated from the most significant bit of the pixels values in the blocks is minimized. In other words, best matching blocks have a maximum number of co-located pixel values, for which the most significant bit is equal. To reduce the number of block matching operations per motion-estimated block, the search range sr is kept low. Since, the downscaling process does not include filtering; actual down-sampling is not required from an implementation point of view. Instead, the block matching process can only take samples located at the preserved row and column positions into account.

After OBME, every motion vector per overlapping block is upsampled by a factor ξ , after which the upsampled motion field is used to motion-compensate blocks $\xi\beta$ of size $\xi B \times \xi B$ from the original reference frames. Since these blocks are overlapping as well, every pixel position in the predicted frame belongs to a number of overlapping blocks $\xi\beta$, each of them linked with their best matching blocks in each of the reference frames. The pixels in these best matching blocks act as temporal predictors for the co-located pixels in the predicted block. In this way, every pixel in the predicted frame is linked to a set of candidate predictor pixels. Finally, the pixel values in the predicted WZ frame are calculated as the average of the candidate temporal predictors at every pixel position.

To reduce the involved computational complexity, particular measures are taken. Namely, (1) motion estimation is carried out on downsampled versions of the original frames, (2) the size of the overlapping blocks together with the overlap size is chosen to be large as to substantially reduce the number of motion-estimated blocks, (3) the motion search range is kept small and (4) overlapping blocks with low-motion characteristics are skipped and replaced by the average of the co-located blocks in the reference frames.

To reduce the complexity at the encoder, a block skip function is included. Namely, prior to motion estimation of a particular overlapping block, the Hamming distance between the co-located blocks in the reference frames is checked first. If the Hamming distance is smaller than a specific threshold T_H , motion estimation is skipped and zero motion vectors are used instead. In this way, the parameter T_H influences the number of skipped blocks and thereby the motion estimation complexity.

3.1.2 Determining the conditional entropy

Let X , \tilde{Y} denote the random variables representing the transform-domain samples, that is, the transform coefficients, in the original WZ frame and the coarse SI frame, respectively. Duplicating the rationale at the decoder, the correlation between the transformed source X and coarse SI \tilde{Y} is expressed as an additive noise channel $X = \tilde{Y} + \tilde{Z}$, where \tilde{Z} is the random variable representing the samples of the estimated correlation noise.

Since both X and \tilde{Y} are available at the encoder, the correlation noise \tilde{Z} can be computed directly based on the histogram. Similar to the noise model established by the decoder, the SID correlation channel concept [11] is adopted. Specifically, the channel output X is modelled by a Laplacian distribution centred on the particular realization of the SI \tilde{y} with a standard-deviation $\sigma(\tilde{y})$ that depends on \tilde{y} . The different standard deviations are obtained using the offline SID correlation channel estimation (CCE) procedure, described in [11].

Once the correlation channel has been modelled, the conditional entropy of every bit-plane in every frequency band (that is coded according to the QM) is determined. For simplicity, the presentation in the following is narrowed down to the coefficients X and \tilde{Y} that belong to a specific frequency band β . Not to overload the notation, the suffix β is omitted. Let M be the total number of bit-planes used to represent the coefficients in band β . Denote by x_n , \tilde{y}_n the n th coefficient in the bands of the WZ and SI frame, respectively. Also, denote by q_n the M -bit quantization index corresponding to x_n . Finally, let $b_n^0, b_n^1, \dots, b_n^{M-1}$ represent the bits composing the binary representation of index q_n , where b_n^0 is the most significant bit. With these notations, the conditional probability p_n^m of bit m of the n th quantized coefficient in X is calculated according to [22],

$$p_n^m = p(b_n^m | \tilde{y}, b_n^0, b_n^1, \dots, b_n^{m-1}) \quad (1)$$

$$= \frac{p(b_n^0, b_n^1, \dots, b_n^{m-1}, b_n^m | \tilde{y})}{p(b_n^0, b_n^1, \dots, b_n^{m-1} | \tilde{y})},$$

where $p(b_n^0, b_n^1, \dots, b_n^{m-1}, b_n^m | \tilde{y})$ and $p(b_n^0, b_n^1, \dots, b_n^{m-1} | \tilde{y})$ are evaluated using the SID correlation channel model estimated at the encoder. Then, the conditional entropy $H_{X|\tilde{Y}}^m$ of the entire bit-plane m , given the transform-domain coarse SI \tilde{Y} , is computed as [22]

$$H_{X|\tilde{Y}}^m = \frac{1}{N} \sum_{n=0}^{N-1} [-p_n^m \log_2(p_n^m) - (1-p_n^m) \log_2(1-p_n^m)], \quad (2)$$

where N is the total number of coefficients in frequency band β .

3.1.3 Proposed coding mode selection process

Concerning mode decision, similar to [28], the skip mode is selected on a frequency band basis, in which case, entire frequency bands from the SI are substituted in the reconstructed WZ frames. Such an approach is advantageous in the sense that high-frequency components are often less important and can be replaced by the corresponding components at a relatively small distortion penalty while no rate is spent. Moreover, skipping entire frequency bands creates more consistent reconstructed coefficients compared to, for instance, a bit-plane-based skip where potentially erroneous bits are introduced. Such erroneous bits undermine the successful decoding of any less significant SW bit-planes since during the creation of the soft-input information every already decoded bit-plane is assumed to be correct. What is more, even when decoding of subsequent bit-planes proves successful, be it intra or SW, any errors in more significant skipped bit-planes would push the reconstructed coefficient value into the wrong quantization bin which increases the incurred distortion. Moreover, the rate spent on any subsequent less significant bit-planes is used sub-optimally. On the other hand, both intra and SW modes are assigned on a bit-plane basis. Intra coding is an attractive alternative when SW coding is expected to be inefficient due to poor SI. Under this condition, SW decoding failure is a potential risk. In this context, intra coding is favoured for bit-planes with higher significance as to further reduce the danger of distortion due to significant SW coded bit-planes that fail to decode.

The proposed mode selection is performed on the fly at the encoder and the selected modes are signalled to the decoder. The mode signalling information (MSI) is compiled per frame and organized into a binary map. For every frequency band considered in the relevant QM, the MSI indicates whether the band is skipped or not, while for every bit-plane in a coded band the MSI signals whether intra or SW coding is applied. The binary MSI string is compressed using binary arithmetic coding and the resulting bitstream is multiplexed with the intra, SW and hash bitstream.

3.2 Skip mode selection

For clarity of the ensuing discussion, the source data is confined to the transform coefficients of a single frequency band β . The problem to solve is whether the coefficients in band β should be skipped or not. On the one hand, skipping a band does not spend any rate at the cost of the distortion incurred by using the SI as reconstruction. On the other hand, coding a band consumes rate with the benefit of reduced distortion. Such a balancing act can be expressed as a Lagrangian cost.

The Lagrangian cost function C_{Skip} , when skipping band β , is given by:

$$C_{\text{Skip}} = R_{\text{Skip}} + \lambda D_{\text{Skip}}, \quad (3)$$

where R_{Skip} , D_{Skip} are, respectively, the required rate and suffered distortion. In a complementary manner, the cost function C_{NoSkip} , when frequency band β would be coded, is given by:

$$C_{\text{NoSkip}} = R_{\text{NoSkip}} + \lambda D_{\text{NoSkip}}, \quad (4)$$

where R_{NoSkip} is the rate for coding the bit-planes of the band and D_{NoSkip} corresponds to distortion from quantization, under the assumption that all bit-planes are correctly decoded. The Lagrange multiplier λ in Equations (3) and (4) controls the relative importance of rate versus distortion in the total cost.

In case the frequency band is skipped, no data is actually coded. Hence, it is trivial that $R_{\text{Skip}} = 0$. When the band is not skipped, the rate is approximated by the sum of the theoretically required SW rates for coding the bit-planes composing the quantization indices of the quantized transform coefficients in the band. The conditional bit-plane entropy $H_{X|\tilde{Y}}^m$ of bit-plane m , given the coarse SI \tilde{Y} is given in Equation (1). Hence, the estimated total rate is the sum of the conditional entropies over all M bit-planes, that is,

$$R_{\text{NoSkip}} = \sum_{m=0}^{M-1} (H_{X|\tilde{Y}}^m). \quad (5)$$

Regarding the computation of the distortion contributions in the Lagrangian cost functions, the distortion suffered from skipping the band is due to the reconstruction at the co-located SI values. However, the true SI Y is not available at the encoder, where only its coarse approximation \tilde{Y} is present. Hence, the mean square error (MSE) distortion D_{Skip} is estimated using the coarse SI \tilde{Y} as:

$$D_{\text{Skip}} = E[(X-Y)^2] \approx E[(X-\tilde{Y})^2] \approx \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \tilde{y}_n)^2, \quad (6)$$

where N is the number of coefficient samples in the frequency band and x_n , \tilde{y}_n are the n th sample value of the original coefficients X and \tilde{Y} , respectively.

In case the frequency band is not skipped, the expected MSE distortion D_{NoSkip} between the original coefficients X and their reconstruction \hat{X} is expressed by:

$$D_{\text{NoSkip}} = E[(X-\hat{X})^2] \approx \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \hat{x}_n)^2, \quad (7)$$

where \hat{x}_n is the reconstruction of the n th sample value x_n . Mimicking the decoder operation under the assumption

that all bit-planes representing X are properly decoded, the reconstruction points at the encoder are derived from coarse SI values \tilde{y}_n and the encoder-side correlation channel statistics $f_{X|\tilde{Y}}(x|\tilde{Y} = \tilde{y}_n)$. In particular, the reconstruction of the n th sample is approximated by:

$$\hat{x}_n = \frac{\int_{l(q_n)}^{u(q_n)} x f_{X|\tilde{Y}}(x|\tilde{Y} = \tilde{y}_n) dx}{\int_{l(q_n)}^{u(q_n)} f_{X|\tilde{Y}}(x|\tilde{Y} = \tilde{y}_n) dx}, \quad (8)$$

where $u(q_n)$, $l(q_n)$ are the upper and lower bound of the quantization interval defined by q_n , respectively.

An appropriate λ configuration was obtained as a result of offline experimentation on (1) a set of medium and high-motion sequences different from the ones reported in Section 4 and (2) over the entire rate range per sequence. The λ parameter is calculated according to the form:

$$\lambda = \lambda_1 e^{-\lambda_2(1-Q_m)}, \quad (9)$$

where the parameter Q_m is the sequence number, ranging from 1 (lowest quality) to 8 (highest quality), of the QM [36,37] used for quantizing the WZ frames and λ_1 , λ_2 are the model parameters. Then the final decision whether to skip frequency band β is made by comparing both Lagrangian cost functions C_{Skip} and C_{NoSkip} , the smaller of the two indicating the selected coding mode.

3.3 Intra mode decision

From a compression point of view, intra coding is theoretically less efficient given that the entropy H_X ‘the theoretical lower bound for intra coding’ is always higher than or equal to the conditional entropy $H_{X|\tilde{Y}}$ the theoretical limit for the SW mode. Nevertheless, the option of an intra coding mode is very attractive in the context of the proposed feedback-free WZ architecture to reduce the suffered distortion. Indeed, intra decoding success is independent of the quality of the SI and does not depend on any encoder-side rate estimation.

As before, the binary representation of every quantization index q is composed by M bits, b^0, b^1, \dots, b^{M-1} , with b^0 the most significant bit. Then, the entropy H_X^m of bit-plane $m = 0, 1, \dots, M - 1$ is

$$H_X^m = -p(b^m) \log_2(p(b^m)) - (1-p(b^m)) \log_2(1-p(b^m)), \quad (10)$$

where the bit probabilities $p(b^m)$ are obtained directly from the histogram of the quantization indices q composing frequency β band.

The decision process whether to apply SW or intra coding to bit-plane m is based on a comparison between the bit-plane entropy H_X^m and the conditional bit-plane entropy $H_{X|\tilde{Y}}^m$, specifically,

$$\mu(m) H_{X|\tilde{Y}}^m < H_X^m, \quad (11)$$

with $\mu(m) \geq 1$ and of the form,

$$\mu(m) = 1 + \mu_1 e^{-\mu_2 m}, \quad (12)$$

where μ_1 and μ_2 are the model parameters. If Equation (11) is true, bit-plane m is SW coded, otherwise intra coding is applied.

Significant bit-planes that are SW coded but fail to decode properly due to channel rate underestimation introduce large reconstruction errors, even when subsequent bit-planes of lesser significance are successfully decoded. Additionally, the generation of the soft-input information to initialize the LDPCA-decoder for decoding a particular bit-plane supposes that all previous bit-planes have been correctly restored. Erroneously decoded bits distort the soft input information, that is, the LLRs are calculated from erroneous data, which could result in failure to decode even though the assigned channel rate would prove sufficient when the soft input information were derived under error-free conditions.

Gradually decreasing $\mu(m)$ from the most to the least significant bit-plane tends to concentrate the likelihood of intra coding at the more significant bit-planes. As a result, the chance of reconstructing coefficients with large distortion at the decoder is reduced, since proper intra decoding is guaranteed. Moreover, channel decoding of later SW coded bit-planes is less disrupted by soft-input information derived from already decoded bit-planes that contain errors.

3.3.1 Finalizing the Slepian-Wolf rate

For those bit-planes $m \in \{0, \dots, M - 1\}$ that are SW coded, the theoretically required channel rate for successful decoding given the coarse SI signal \tilde{Y} , i.e. the conditional entropy $H_{X|\tilde{Y}}^m$, is adjusted. This compensates for the mismatch between the SI \tilde{Y} approximated at the encoder and the SI Y generated at the decoder and the fact that the estimated virtual correlation channel, identified by $f_{X|\tilde{Y}}(x|\tilde{y})$ is not identical to the one actually used at the decoder, governed by $f_{X|Y}(x|y)$.

Based on $H_{X|Y}^m$, where $0 \leq H_{X|Y}^m \leq 1$, the following simple yet effective rate formula is used to calculate the final rate R_{SW}^m for the SW-coded bit-plane m in band β ,

$$R_{SW}^m = \left(H_{X|Y}^m \right)^{g(m)}, \quad (13)$$

where $g(m)$ is a linearly increasing function given by $g(m) = b + (a - b)/M \cdot (m - 1)$, where $a, b \in [0, 1]$, $a > b$ are the model parameters and M is the total number of bit-planes in β . Under these conditions, $0 < g(m) \leq 1$ and thereby, $R_{SW}^m \geq H_{X|Y}^m$ holds for every bit-plane m . The rationale behind Equation (13) is the following. The incurred distortion due to decoding failure is higher for more significant bit-planes. Therefore, the exponent in Equation (13) increases with m , thus compensating more rate for more significant bit-planes. Finally, after the rate has been adjusted according to Equation (13), the closest supported syndrome for bit-plane m , that is, with length closest to $\text{ceil}[R_{SW}^m N]$ is sent to the decoder.

3.4 The decoding procedure

In the decoder, the key frames are decoded, reconstructed and stored in a buffer to serve as reference frames for motion estimation. The hash is decoded as well. Then, every WZ frame is decoded in distinct stages, called SI refinement levels (SIRLs), where after each stage, a higher quality version of the SI is generated by the decoder. Every SIRL is built around frequency bands of the 4×4 DCT aggregated along the diagonal, as introduced in our previous work [34]. The proposed feedback-channel-free DVC architecture takes advantage of the presence of the SI refinement scheme to reduce the distortion incurred at the decoder.

Figure 2 shows an overview of the proposed method. Suppose there is a total number of L refinement levels SIRL_l , $l = \{1, 2, \dots, L - 1\}$. At the first level, SIRL_0 , the decoder creates the initial SI Y_0 , using the designated hash-based OBMEC with sub-sampled matching (OBMEC/SSM) technique from [11]. The 4×4 DCT is applied and the coding MSI for the bit-planes in the frequency bands that belong to SIRL_0 , that is, the DC band, is addressed.

When the skip mode was selected for the DC frequency bands, no decoding, and as a consequence, no CCE or reconstruction takes place and the coefficients in the DC band of the current version of the SI are copied into the partially decoded frame in the transform domain. When the band was not skipped, every bit-plane composing the band is passed to the appropriate decoder, as dictated by the MSI, with the understanding that SW coded bit-planes might fail to decode while the intra-coded bit-planes are guaranteed to decode successfully. The success of SW decoding is determined as in [40].

As more bit-planes of a band are processed, the bit-plane-per-bit-plane progressively refined CCE algorithm of [11] simultaneously updates the correlation channel estimate for that particular frequency band. For decoding SW bit-planes, the last update of the correlation channel estimate serves as basis to generate the soft-input for the LDPCA decoder. When the bit-plane fails to decode, the erroneous bit-plane is still used to update the correlation channel estimate. For those bit-planes that require binary arithmetic decoding, CCE is irrelevant. However, after decoding, these bit-planes are valuable to the CCE algorithm to further refine the estimate.

When all bit-planes have been decoded, the coefficients are reconstructed at the centroid calculated over all quantization bins that match the correctly decoded bit-planes, using the available SI and CCE result.

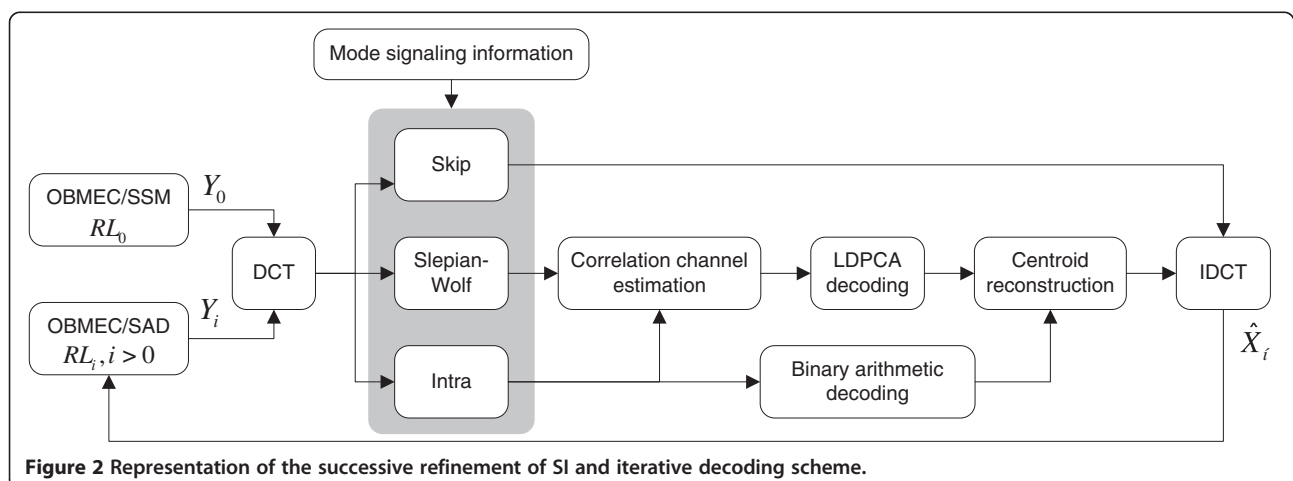


Figure 2 Representation of the successive refinement of SI and iterative decoding scheme.

Specifically, the n th coefficient in a decoded band is reconstructed as:

$$\hat{x}_n = \frac{\sum_{q \in I_n} \int_{l(q)}^{u(q)} x f_{X|Y}(x|y_n) dx}{\sum_{q \in I_n} \int_{l(q)}^{u(q)} f_{X|Y}(x|y_n) dx}, \quad (14)$$

where I_n is the set of quantization indices q that agree with the successfully decoded bit-planes in the band and $u(q)$, $l(q)$ are the respective upper and lower edge of the interval designated by index q .

The frequency bands that have not yet been processed are substituted by their co-located counterparts in the SI and the application of the IDCT yields the partially reconstructed frame \hat{X}_0 in the spatial domain. The applied reconstruction technique is optimal in the MSE sense given the SI, even when SW coded bit-planes failed to decode and no unique quantization bin in which to reconstruct is available.

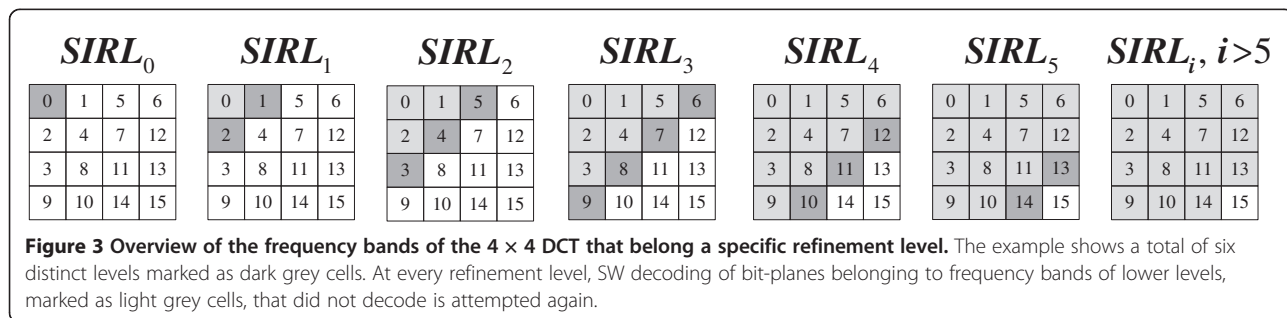
The primary objective, however, should be to minimize the total number of SW decoding failures. Therefore, with the intention of minimizing the number of bit-planes that fail to SW decode successfully, the proposed WZ system exploits the successive SI refinement loop at the decoder, by reattempting to decode any bit-planes that failed to decode at previous SIRLs using the updated SI available at the current refinement level. At the same time, the distortion due to skipped frequency bands can be mitigated by substituting the co-located frequency bands in the latest version of the SI into the partially decoded frame at every successive SIRL.

In detail, for every SIRL _{i} , $i > 0$, the partially decoded frame \hat{X}_{i-1} , created at the previous level, is used in another round of SI generation by means of OBMEC [34], where the SAD criterion is used as the error metric during block matching since the hash frame is no longer involved in the motion estimation. The resulting motion-compensated frame serves as a new version of the SI information and is converted to DCT domain.

Then, a new correlation channel estimate is executed for all the coded bands belonging to any previous SIRL _{j} , $j < i$. The result is then used, together with the SI Y_i to create the soft-input information for any bit-planes of SIRL _{j} , $j < i$, that failed to decode and reattempt SW decoding given the fixed number of received syndrome bits. These bands are indicated by the light grey cells in Figure 3 for every level in a configuration using six SIRLs. Although successful decoding is still not assured, any bit-planes that actually are decoded successfully reduce the distortion without additional rate. Next, the frequency bands that actually belong to SIRL _{i} are handled (i.e. the dark grey cells in Figure 3).

The mode selection information determines whether a band is skipped or not and in the latter case decides whether a bit-plane is passed to the LDPCA or binary arithmetic decoder. For every bit-plane, a CCE is performed to generate soft-input information to enable SW decoding or simply to update the CCE algorithm for the next bit-plane. When all bit-planes have been processed, the coefficients in the band are reconstructed given the current SI Y_i . Due to the CCE at every refinement level using the updated SI, the reconstruction of the coefficients in the already completed SIRL is further improved as well. At last, the coefficients of the bands of the processed SIRLs are assembled with the SI coefficients belonging to the as of yet not decoded bands, which after the IDCT yields the partially decoded WZ frame \hat{X}_i . To sum up, the proposed decoding process merges SI refinement, CCE updating, SW decoding and reattempting to decode any SW-coded bit-planes that failed to decode at previous refinement levels. This contrasts the approach proposed in [23], where an entire WZ frame is first decoded completely after which additional SI updating and CCE runs enable new opportunities to attempt decoding bit-planes that failed to SW decode successfully.

Yet, when all SIRLs have been terminated, the proposed decoder architecture still does not guarantee all SW coded bit-planes have been decoded properly. Therefore, similar to [23], additional SIRLs are added. These supplemental SIRLs, that is, the SIRL _{i} , $i > 5$ in the configuration depicted in Figure 3, solely consist of OMBEC-based SI generation,



CCE, reattempting to decode any failed bit-planes and reconstructing the coefficients in all frequency bands given the updated SI and correlation channel model. The number of these additional refinement levels is controlled by a fixed parameter or are skipped when all bit-planes happen to decode properly.

4 Experimental results

4.1 Experimental setup and codec configuration

The proposed feedback-channel-free WZ codec is configured using the following settings. The parameters governing the hash formation process, as well as the hash-based SI generation method used to create the initial version of the SI at the decoder, are identical to the configuration used in [11]. A total number of seven SIRLs are considered. During the first five levels new frequency bands are decoded, while the last two SIRLs only serve to reduce the number of SW coded bit-planes that fail to decode.

Regarding the configuration of the encoder-side components for rate control and coding mode selection, the following parameters are used. The coarse SI generation module uses a downscaling factor $\xi = 4$. The size of the overlapping blocks is $B = 8$ with an overlap step size of $\varepsilon = 4$. The search range is put to $sr = 4$ pixels. The resulting motion vectors are upscled by a factor $\xi = 4$ to motion-compensate overlapping blocks of size $\xi B \times \xi B$ pixels, i.e. 32×32 pixels, from the original sized reference frames. Additionally, the threshold T_H that controls whether a block is skipped during motion estimation is set to $T_H = 12$. Namely, a block is skipped if the number of unequal bits at the same position in the two co-located blocks in the reference frame is lower than 12, which is equivalent to a pixel error ratio of $12/(8 \times 8) \approx 0.18$. Concerning the mode selection modules, the parameters that control the Lagrange multiplier λ in Equation (9) are fixed to $\lambda_1 = 0.03$ and $\lambda_2 = 0.5$. Similarly, the parameters μ_1, μ_2 in Equation (12) to derive $\mu(m)$, which governs the intra mode decision process, are set to $\mu_1 = 0.5$ and $\mu_2 = 2.0$. Finally, the model parameters of the exponent $g(m)$ in the rate formula of Equation (13) are put to $b = 0.4$ and $a = 1.0$.

The parameters were derived heuristically based on offline experimentation on a training set, excluding the sequences reported in the experimental results. The values were selected to achieve good RD performance for various degrees of motion while not compromising the complexity at the encoder. In this context, the RD performance could be further optimized for different motion profiles. For instance, in case of high-motion sequences, the SI approximation module could be configured using a less strict, that is, lower, threshold T_H . Skipping a lower number of blocks during coarse motion estimation would increase the accuracy of the

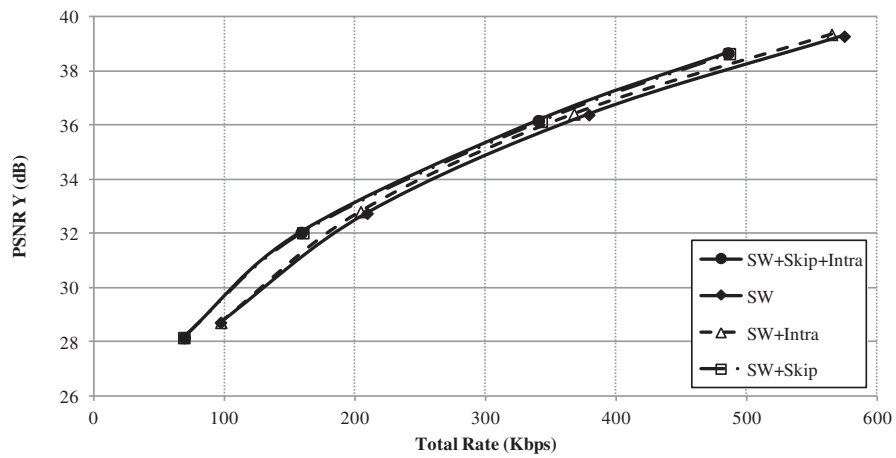
resulting SI approximation, in particular when the motion content is high. Additionally, a smaller overlap step size ε and/or a smaller downscaling factor ξ would increase the accuracy of the temporal prediction. However, all these measures have to be applied carefully since these would increase the SI approximation complexity at the encoder. On the other side, in case of low-motion sequences, more overlapping blocks could be skipped during the SI approximation, without significantly undermining the temporal prediction accuracy. Analogously, the number of overlapping blocks may be decreased. In this regard, a low-motion profile would impose less complexity on the encoder.

Further room for optimization may be achieved by tuning the parameters controlling the mode decision processes. Indeed, a high-motion parameter profile may put less stress on skipping frequency bands but rather put more emphasis on intra-coded bit-planes. Conversely, low-motion profiling should be more advantageous towards skipping frequency bands while penalizing the intra coding mode. However, the single parameter profile presented in this work was determined to (1) achieve good RD performance over all motion profiles, while (2) containing the additional complexity imposed on the encoder such that the low-complexity encoding characteristics are not compromised.

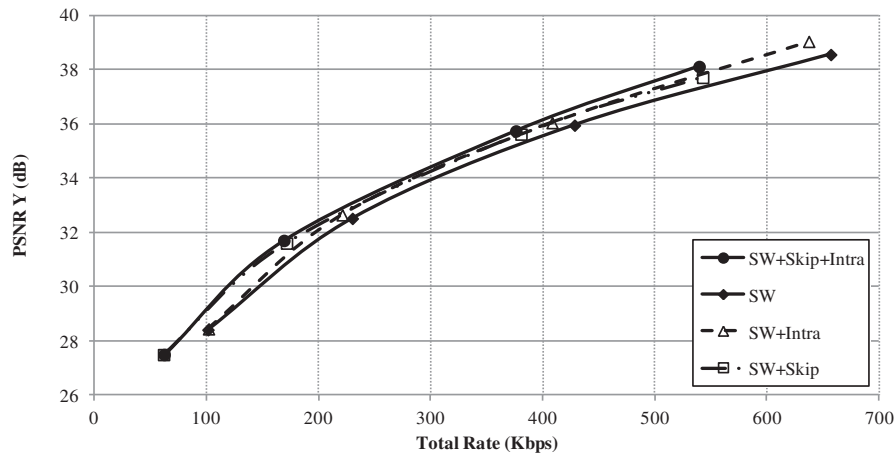
4.2 Mode selection evaluation

In the first set of experiments, the influence of the different coding modes is illustrated. To this end, the compression performance of four versions of the proposed system is assessed. The first version only supports the SW coding mode and essentially corresponds to our previous system presented in [12]. The second and third versions support an additional skip (SW+Skip) or intra (SW+Intra) coding mode, respectively. The final version of the system features all three coding modes (SW+Skip+Intra). Figures 4 and 5 show the compression performance of the proposed feedback-free DVC system with the four configurations on Foreman and Soccer QCIF 15Hz, GOP 2, 4 and 8. For the system featuring all three coding modes, Table 1 reports the percentage of bit-planes assigned to each mode for every GOP size at each considered RD point. Table 2 zooms in on the skip mode and provides insight to which frequency bands are skipped at the lowest and the highest RD point, corresponding to QM 1 and 8, respectively. Table 2 presents the frequency bands similar to the presentation of a QM, where frequency bands that are not coded according to the QM are marked not applicable (na).

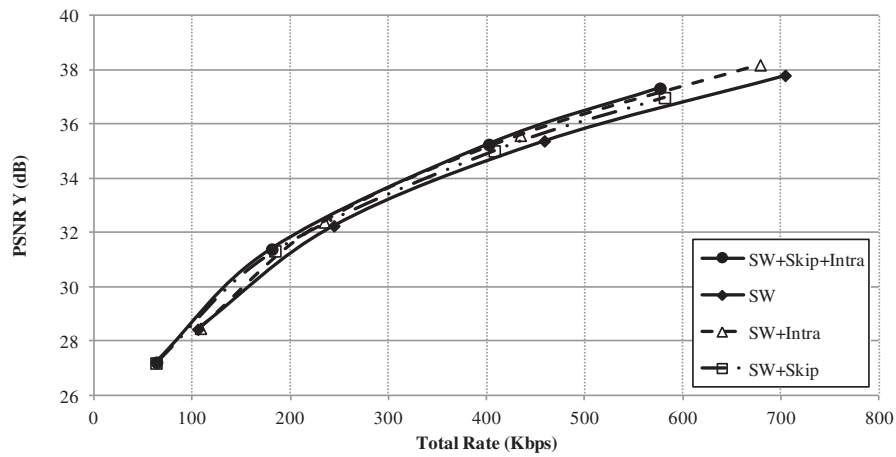
Regarding the results obtained on the Foreman sequence, a medium-motion sequence with complex facial expressions, SW coding plus the skip outperforms SW coding with the intra mode option in a GOP of 2. The



a

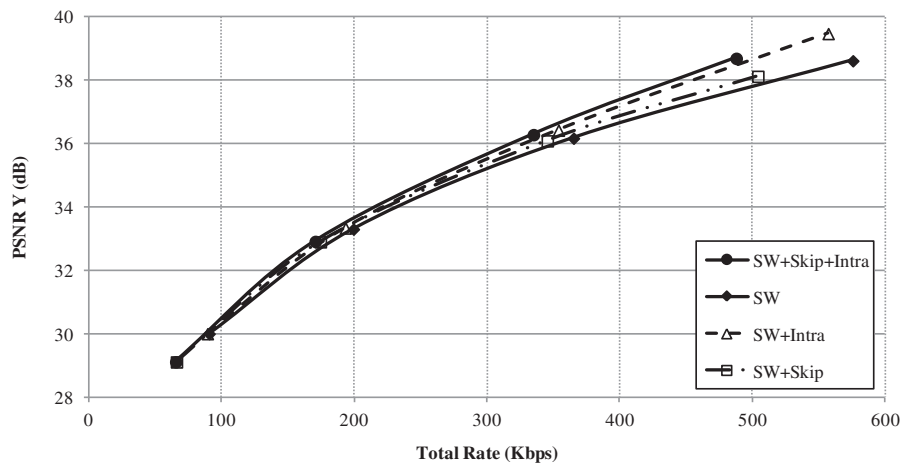


b

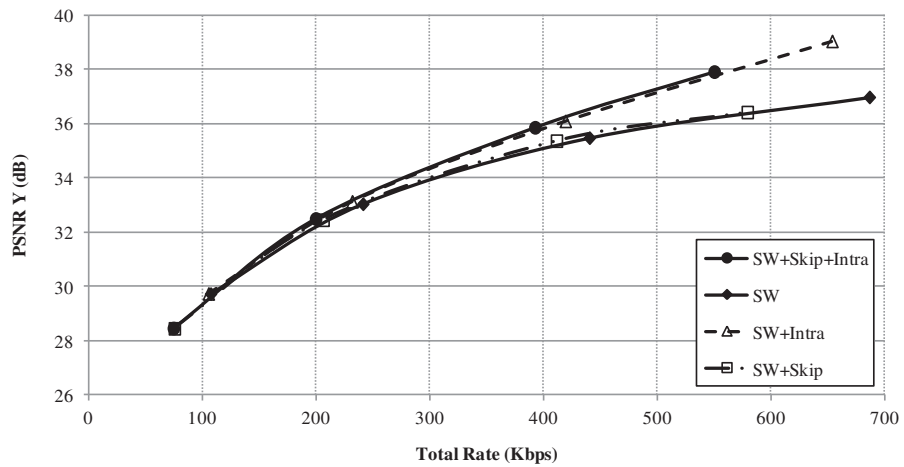


c

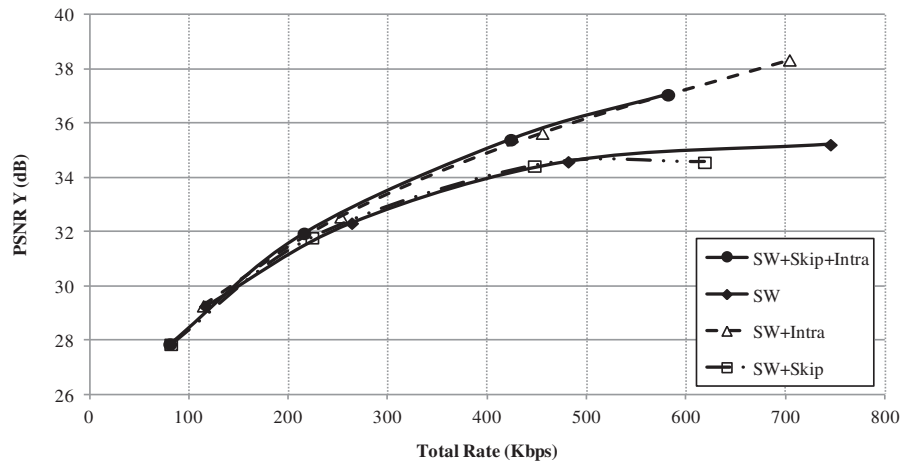
Figure 4 Compression results of the proposed feedback-free DVC system on Foreman in a GOP of (a) 2, (b) 4 and (c) 8. Four different versions in terms of available coding modes are considered.



a



b



c

Figure 5 Compression results on Soccer in a GOP of (a) 2, (b) 4 and (c). The results were obtained with the proposed feedback-free WZ coding architecture in four different settings regarding the available coding modes.

Table 1 Percentage of the total number of bit-planes in every test sequence assigned to the skip, intra or SW coding mode

	QM	GOP2			GOP4			GOP8		
		Skip (%)	Intra (%)	SW (%)	Skip (%)	Intra (%)	SW (%)	Skip (%)	Intra (%)	SW (%)
Foreman	1	84	3	13	79	4	17	74	5	21
	4	57	14	29	49	17	34	45	17	38
	7	28	22	50	25	24	51	24	24	52
	8	36	18	46	31	20	49	29	22	49
Soccer	1	55	9	36	48	10	42	45	10	45
	4	30	23	47	28	24	48	27	24	47
	7	16	30	54	15	31	54	15	53	32
	8	25	29	46	24	31	45	25	31	44

quality of the SI of the skipped frequency bands is sufficient as not to have a negative impact on the distortion at no rate cost. For a GOP size of 4, the quality of the SI drops which results in comparable performance whether SW coding is supplemented with the skip or intra coding mode. When the GOP size increases to 8, resulting in a declining SI quality, the intra coding mode - added as an alternative to SW coding - results in superior performance. Although the influence of adding coding modes is notable, the performance of the proposed encoder-driven rate control by itself, that is, the version with SW coding only, is, although lower, still respectable. The configuration with all three coding modes available delivers the best performance, taking advantage of the effects of both supplemental coding modes, namely, (1) an increase in quality of the reconstructed WZ frames due to the guaranteed successful decoding of intra-coded bit-planes at the expense of rate, which is compensated by (2) a reduction in rate by skipping bit-planes without notably compromising the distortion.

When all three coding modes are enabled, the statistics of the coding mode assignment for Foreman in Table 1 show that skip is dominant for the lower RD points. For instance, Table 1 reports that 84% of the bit-planes are actually skipped at the lowest RD point in a GOP of 2, while at the second RD point in a GOP of 2 the percentage of skipped bit-planes drops to 57%. In the higher rate region, that is, QM 7 and 8, SW coding dominates, accounting for around 50% of the bit-planes. The remaining bit-planes are more or less evenly distributed between the skip and intra modes. Regarding the distribution of the frequency bands affected by the skip mode, Table 2 reports a somewhat uniform distribution of the three frequency bands relevant to the lowest RD point all GOPs. At the highest rate point (QM8), however, it is clear that the majority of skipped bit-planes belong to high frequency bands. This is in line with traditional compression logic, where high-frequency information can often be discarded without causing an exaggerated increase in distortion.

The influence of the skip and intra coding options becomes more prominent in case of Soccer 'a high-motion sequence' where SI generation is strained. Considering the system with SW coding only, Figure 5 shows that the RD performance experiences a saturation effect at the higher rate points, in particular for the larger GOPs. The encoder has difficulty to create an accurate approximation of the SI, and thereby accurate SW rate estimates. Adding only the skip mode does not increase the performance since for such a high-motion sequence, the quality of the SI at the decoder is lower. However, adding the intra coding mode drastically increases the compression performance and completely removes the performance saturation. Since intra coding is completely independent of the SI, it boosts

Table 2 Statistics of the skip mode selection in all considered test sequences, organized in a GOP of 2, 4 and 8

	QM	GOP2 (%)			GOP4 (%)			GOP8 (%)		
		Skip	Intra	SW	Skip	Intra	SW	Skip	Intra	SW
Foreman	1	34	38	na	34	40	na	33	42	na
		28	na	na	26	na	na	25	na	na
		na	na	na	na	na	na	na	na	na
		na	na	na	na	na	na	na	na	na
	8	2	3	4	13	2	3	14	2	3
		4	4	4	13	3	4	12	3	4
		4	6	2	6	3	6	2	7	3
		13	16	6	na	13	18	6	na	14
Soccer	1	31	50	na	28	54	na	27	57	na
		19	na	na	18	na	na	16	na	na
		na	na	na	na	na	na	na	na	na
		na	na	na	na	na	na	na	na	na
	8	0	0	2	18	0	2	18	0	2
		0	1	5	17	0	1	18	0	1
		2	4	5	9	2	5	8	2	4
		14	14	9	na	15	15	9	na	15

The results of two RD points are shown, namely the lowest rate point, corresponding to QM1, and the highest rate point, corresponding to QM8. When a frequency band is not coded, as prescribed, skip mode selection is not applicable (na), since the band is skipped by definition.

the performance when the SI quality is low. Under such conditions, replacing SW coding with intra coding decreases the distortion by reducing the number of bit-planes that fail to decode, in particular significant ones. When all modes are turned on, the system takes full advantage of the effect of the skip and intra modes.

As expected, due to the increased motion content in Soccer, the skip mode is selected less often compared to Foreman, in particular in the lower rate region (see Table 1). Towards the higher rates, the percentage of skipped bit-planes approaches the skip level of Foreman. According to Table 2, the distribution of the skip mode over the frequency bands is similar to Foreman. The most important observation on Soccer is the increased occurrence of the intra coding mode, which is imperative to maintain competitive compression performance for high-motion sequences as seen Figure 5.

To sum up, the skip mode targets rate reduction without invoking a degree of distortion that undermines the RD performance. This is particularly useful at low rates or for low- to medium-motion sequences. The intra coding mode targets bit-planes that are difficult or inefficient to code using SW principles. The SW coding mode holds the middle ground, offering good compression when the SI is of sufficient quality.

Since the entire mode decision process is encoder-driven, a rate overhead to signal the selected coding modes is present. As mentioned in Section 3, the MSI is compiled per frame and organized into a binary map, which is entropy coded using binary arithmetic coding. Per frequency band actually considered in the applied QM, the MSI indicates whether the band is skipped or not, while for every bit-plane in band that is not skipped the MSI signals whether intra or SW coding has been assigned. Table 3 provides insight in the mode signalling rate overhead (kbps) at all RD points and GOP sizes for Foreman and Soccer. As expected, the required rate to code the MSI is largest for QM8, since both the number of frequency bands and the number of bit-planes assigned

to each band is the largest. This setting corresponds to the largest number of mode decisions at the frequency level, in case of the skip mode, as well as the bit-plane level, in case of the intra or SW mode. Similarly, the rate overhead increases with increasing GOP size, due to the increased number of WZ frames for which mode decision is executed. Nevertheless, the MSI rate overhead remains rather small for all RD points and all GOP sizes, with the maximum rate reported in Table 3 equalling 1.29 kbps for the highest RD point in Soccer GOP of 8.

Rate points are determined by the quality of the intra frames (QP_{Intra}) and the employed QM.

4.3 The effect of side-information approximation at the encoder

To create an approximation of the SI, the encoder of the proposed WZ codec features a low-complexity SI generation technique. However, the proposed SI approximation method based on motion estimation is inevitably more complex than reference frame averaging. It is therefore necessary to examine the performance gain brought by the proposed technique at the cost of increased complexity.

The compression performance on Foreman and Soccer, shown in Figure 6a,b, respectively, illustrates the effect of both encoder-side SI approximation techniques on the performance of the proposed feedback-free DVC system. The performance when using the proposed low-complexity technique ‘marked ME’ and reference frame averaging ‘marked RA’ nearly coincides for Foreman, GOP2. Consequently, the proposed encoder-side SI estimation technique is not warranted in this case, since the limited gain in RD performance does not compensate for the increase in complexity. When the GOP size increases, however, the proposed technique clearly outperforms reference frame averaging. On Soccer, the proposed motion estimation-based method systematically asserts itself as the better of the two at all GOP sizes.

Overall, the proposed feedback-free system configured with the low-complexity OBMEC method, outperforms the system where the SI is approximated by reference frame averaging. However, the difference in compression performance is not overwhelming at all times. For example, when the motion activity is low and the GOP size is small, reference frame averaging seems a good option given the lower computational cost. In general, the choice of the SI approximation method should be evaluated in an application context to establish whether increased compression performance outweighs the added encoding complexity.

4.4 Performance comparisons against benchmark techniques

Compression results are presented on the Foreman, Carphone, Soccer and Football sequences, all at QCIF

Table 3 Rate overhead (kbps) to convey the selected coding modes from encoder to decoder at every considered rate point

	QP_{Intra}	QM	GOP2	GOP4	GOP8
Foreman	40	1	0.23	0.34	0.39
	34	4	0.40	0.61	0.72
	29	7	0.57	0.86	1.01
	25	8	0.62	0.96	1.14
Soccer	44	1	0.23	0.35	0.41
	36	4	0.45	0.68	0.80
	31	7	0.64	0.97	1.13
	25	8	0.73	1.10	1.29

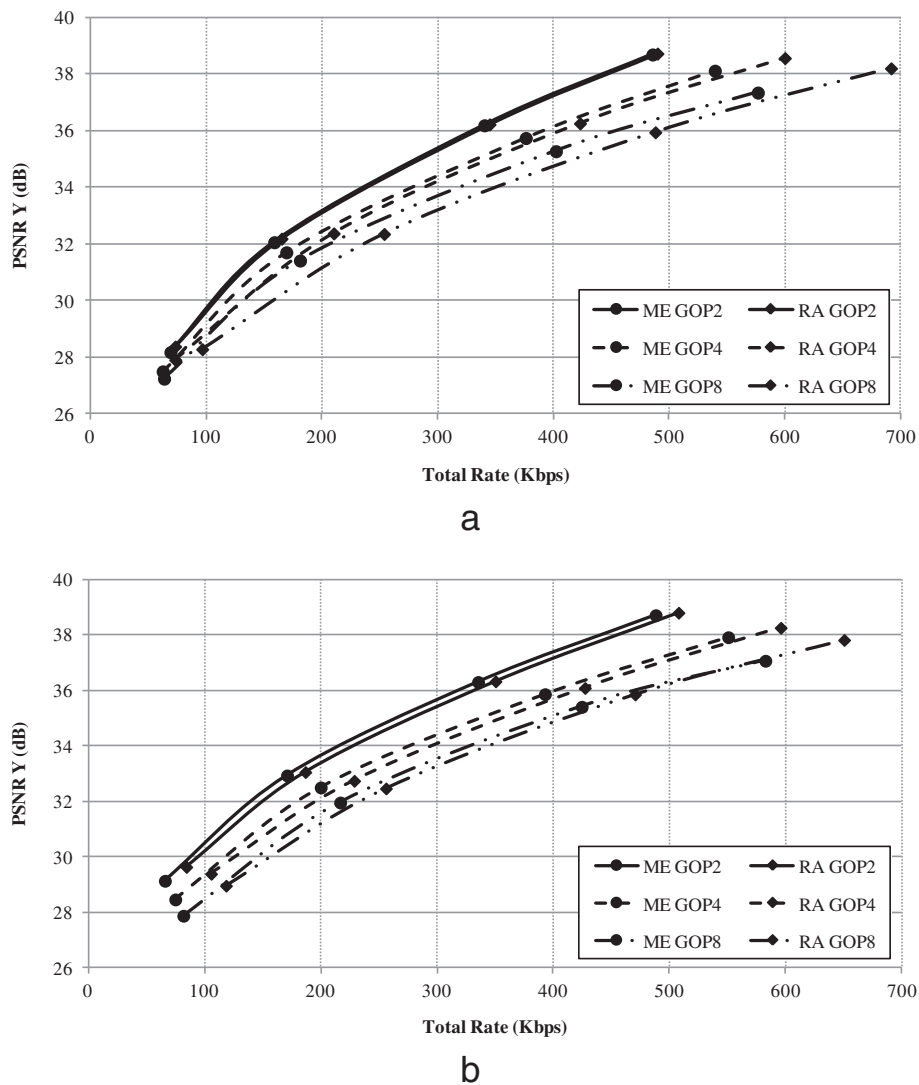


Figure 6 Comparative compression results on (a) Foreman and (b) Soccer for a GOP of 2, 4 and 8. The proposed feedback-free WZ architecture used either the proposed motion estimation-based (ME) low-complexity SI generation method either reference frame averaging (RA).

15Hz. All sequences are organized into GOPs of size 2, 4 and 8. The compression performance of the proposed feedback-free architecture with all three coding modes activated is benchmarked against the feedback-channel-based DISCOVER [13] codec, as well as H.264/AVC Intra [35]. In addition, compression results obtained with our previous hash-based DVC architecture from [7] are also included. The system generates SI by means of hash-based OBMEC, where the original WZ frames are downscaled and subsequently coded using H.264/AVC Intra coding at low bit-rate to create the hash information. Similar to DISCOVER [13], rate allocation is implemented as a request-and-decode approach via a feedback channel. Finally, the compression performance of the proposed feedback-free architecture is compared to the performance of the same system, including

all three coding modes, when rate allocation is implemented using the feedback mechanism as in DISCOVER [13]. In this context, it is important to emphasize that systems with decoder-driven feedback-based rate allocation enjoy the advantage in terms of RD performance, since the presence of a feedback channel not only guarantees decoding success but also ensures that decoding is achieved at a minimal rate. On the other hand, such systems are unable to support unidirectional application scenarios and link the encoding and decoding process, which may lead to excess delays.

A summary of the compression performance comparison of the proposed feedback-free architecture, expressed in Bjøntegaard [41] rate reduction (%) and peak signal-to-noise ratio (PSNR) improvement (dB), with respect to DISCOVER [13], our previous hash-based DVC system

from [7] and the proposed system with feedback, is given in Table 4.

Regarding the performance on Foreman, shown in Figure 7, the compression performance of the proposed feedback-channel-free architecture is roughly comparable to the performance of DISCOVER, H.264/AVC Intra and the hash-based system from [7] for a GOP of 2. Amongst these, the DVC systems slightly outperform H.264/AVC Intra at low rates but lag somewhat behind at the highest rates. However, the proposed DVC architecture using all three presented coding modes outperforms all other systems when rate-allocation is performed at the decoder via a feedback channel. At GOP sizes 4 and 8, the proposed feedback-free system still outperforms DISCOVER at low to medium rates, steadily losing ground towards the highest rate point. Table 4 reports a Bjøntegaard [41] rate gain of DISCOVER with respect to the proposed unidirectional system of 4.21% for a GOP of 2 which turns into a rate loss of 1.60% and 10.33% for GOP4 and GOP8, respectively. Our previous hash-based system with feedback [7] consistently outperforms both DISCOVER and the proposed feedback-free solution. When feedback is turned on in the proposed DVC architecture, the compression performance of the resulting system is boosted, regularly surpassing the performance of H.264/AVC Intra.

The results on Carphone, shown in Figure 8, are more or less similar to the results obtained on Foreman. The DVC systems have comparable performance at lower rates, superior to H.264/AVC Intra. In the higher rate region, H.264/AVC Intra becomes dominant. At the same time, DISCOVER outperforms the proposed feedback-free WZ architecture. Comparing these WZ architectures, Table 4 reports Bjøntegaard [41] rate gains

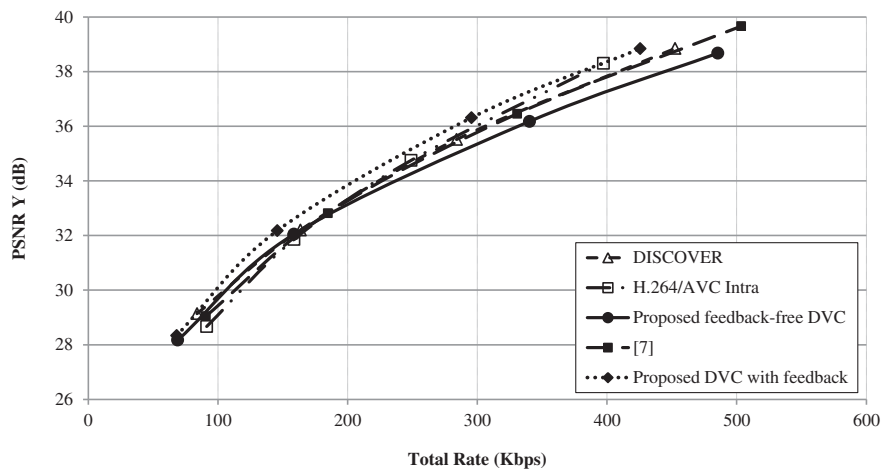
from 4.07% to 9.80% of DISCOVER compared to the proposed system. The two remaining feedback-based DVC architectures outperform both DISCOVER and the feedback-free solution, where the DVC system supporting the proposed mode decision process is consistently on top.

The complex motion content in Soccer is highly favourable to H.264/AVC Intra. Indeed, Figure 9 shows that the performance of H.264/AVC Intra is far better than the performance of the WZ codecs, which are hard-pressed estimating the motion at the decoder. However, the proposed feedback-free system outperforms the feedback-based DISCOVER codec at all GOP sizes and RD points, save the highest rate point where a slight performance loss is incurred. In general, DISCOVER is outperformed by the proposed system with Bjøntegaard [41] rate losses ranging from 9.80% in a GOP of 2 up to 15.94% in a GOP of 8. Similarly, Football exhibits a very high degree of motion and complex camera movements. Again, these conditions are to the advantage of H.264/AVC Intra, as supported by the results in Figure 10. In general, DISCOVER is outperformed by the proposed feedback-free WZ architecture in all GOP sizes, with Bjøntegaard [41] rate losses from 2.82% to 1.28% in a GOP of 2 and 8, respectively. As shown in Figure 5, the application of the intra mode safeguards the compression performance of the proposed codec when high-quality SI generation is strained. Regarding the performance of our previous hash-based system from [7] and the performance of the proposed system configured with feedback-based rate allocation, both systems significantly outperform DISCOVER and the proposed feedback-free architecture. This is expected, since hash-based architectures in combination with decoder-driven rate allocation via feedback are known to perform

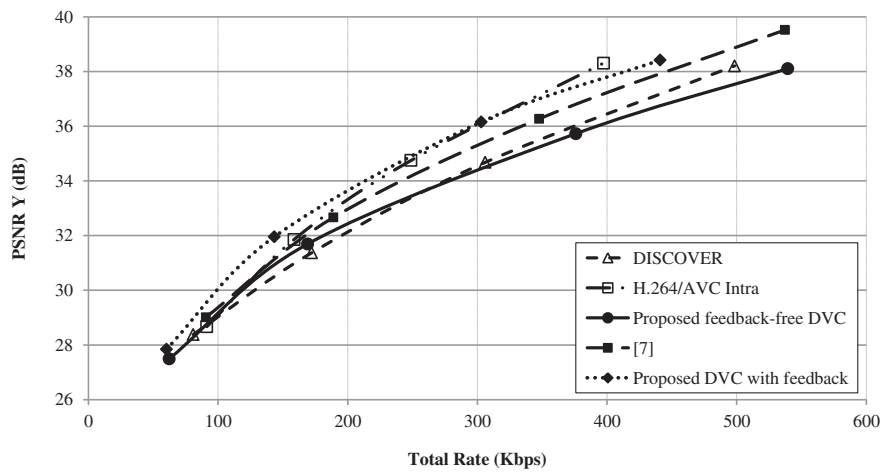
Table 4 Comparative compression performance of the proposed feedback-free architecture

		GOP2		GOP4		GOP8	
		ΔR (%)	$\Delta PSNR$ (dB)	ΔR (%)	$\Delta PSNR$ (dB)	ΔR (%)	$\Delta PSNR$ (dB)
Foreman	DISCOVER	-4.21	0.21	1.60	-0.13	10.33	-0.56
	[7]	-1.80	0.13	-10.42	0.63	-18.36	1.10
	Proposed DVC with feedback	-11.55	0.66	-20.43	1.11	-27.94	1.54
Soccer	DISCOVER	9.80	-0.56	14.36	-0.84	15.94	-0.89
	[7]	-1.28	0.07	-10.66	0.55	-21.55	1.21
	Proposed DVC with feedback	-16.04	0.83	-23.82	1.30	-30.68	1.79
Football	DISCOVER	2.82	-0.18	1.94	-0.11	1.28	-0.10
	[7]	-7.27	0.44	-15.28	0.97	-21.39	1.34
	Proposed DVC with feedback	-13.36	0.81	-21.48	1.37	-27.45	1.74
Carphone	DISCOVER	-4.07	0.21	-9.80	0.47	-5.42	0.23
	[7]	-1.57	0.18	-8.49	0.58	-12.27	0.76
	Proposed DVC with feedback	-8.94	0.50	-14.28	0.72	-17.79	0.89

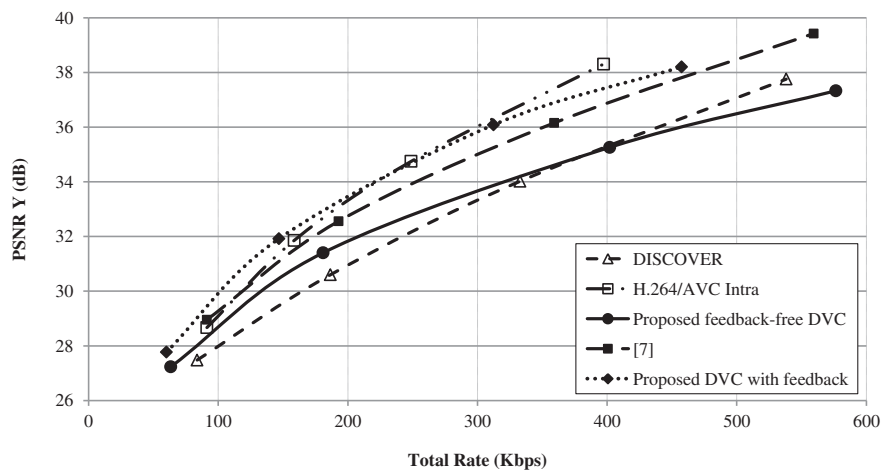
Performance is expressed in terms of Bjøntegaard deltas [41] of the proposed feedback-free DVC configured with all three coding modes, versus DISCOVER [13], the hash-based DVC system with feedback from [7] and the proposed DVC system configured with feedback-based decoder-driven rate allocation. Negative rate and positive PSNR delta values correspond to gains compared to the proposed feedback-free solution.



a

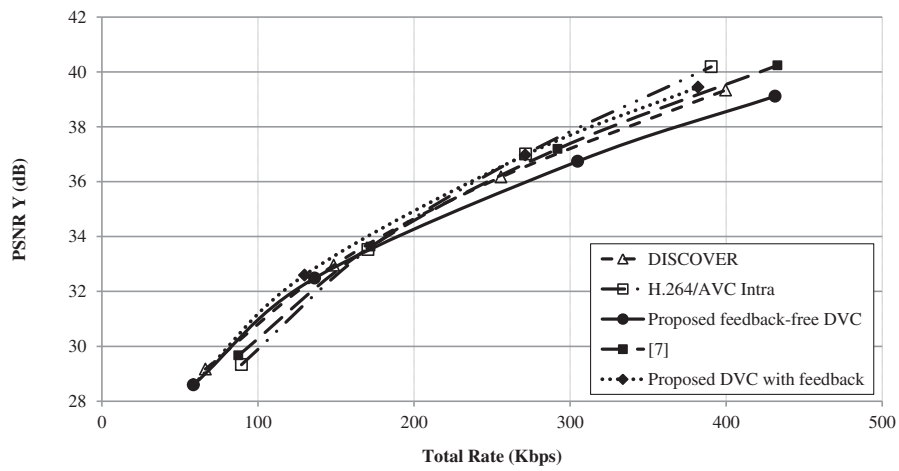


b

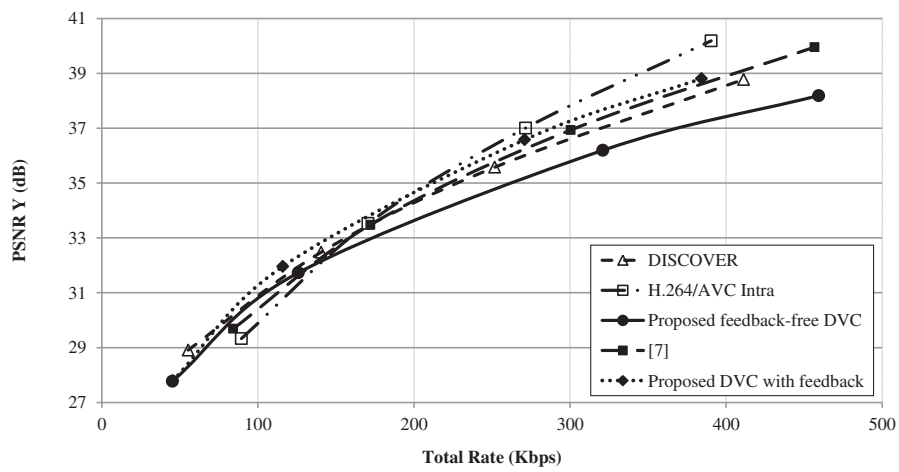


c

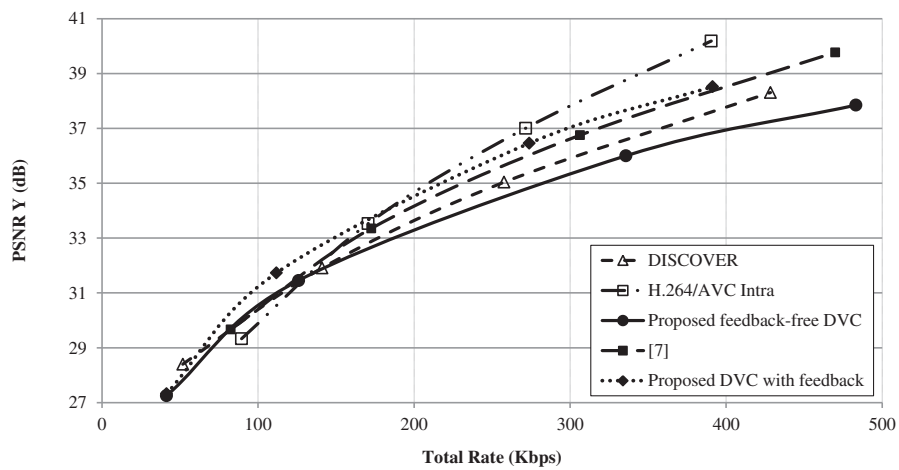
Figure 7 Compression results obtained on Foreman in a GOP of (a) 2, (b) 4 and (c) 8. The results include DISCOVER [13], H.264/AVC Intra [35], the proposed feedback-free DVC with all three modes, our previous hash-based DVC architecture from [7] and a version of the proposed DVC system using decoder-driven feedback-based rate allocation.



a

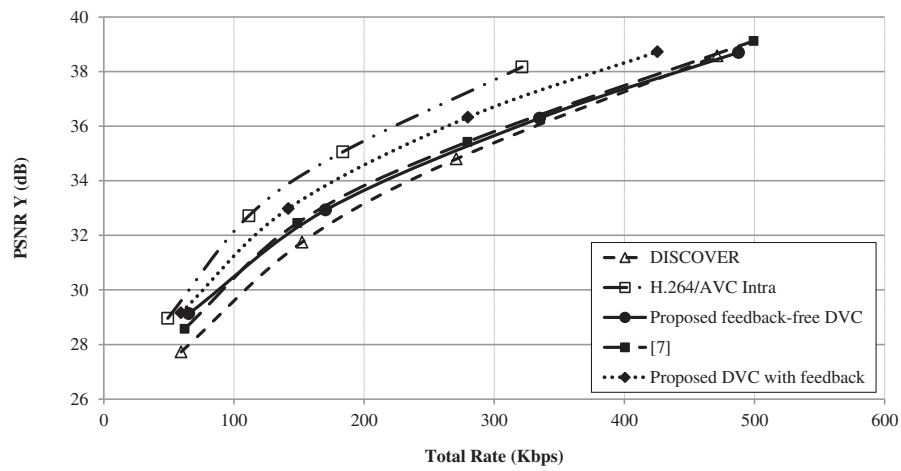


b

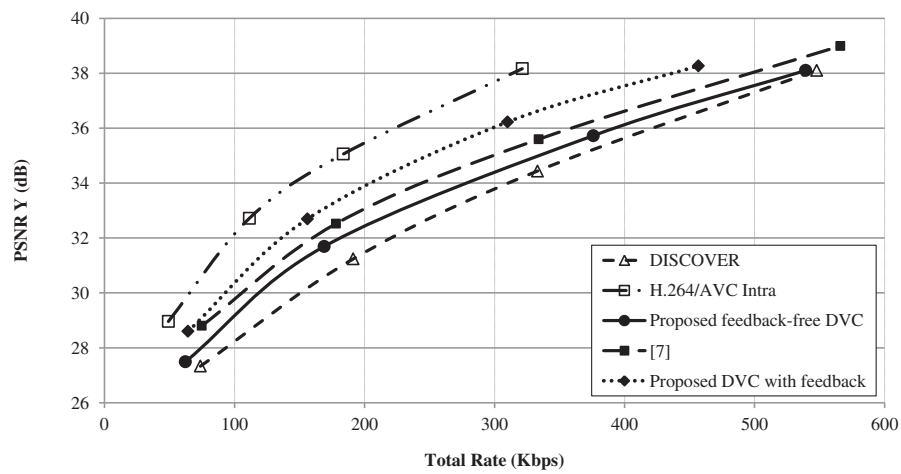


c

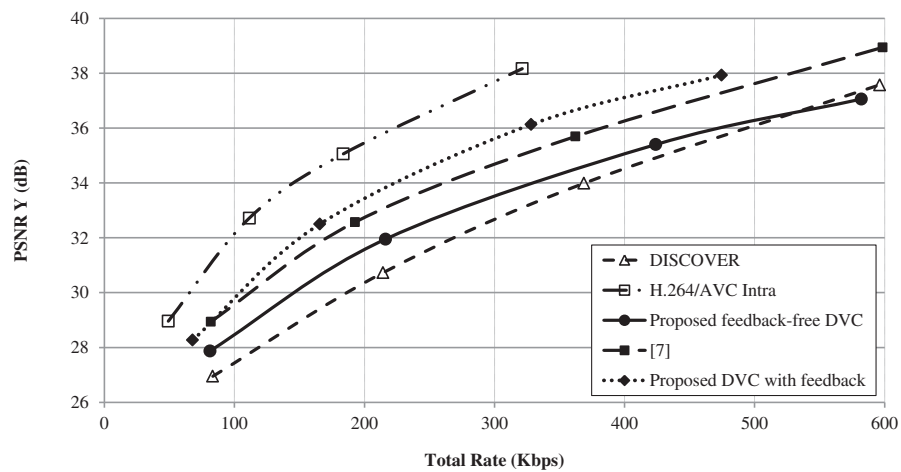
Figure 8 The RD behaviour on Carphone organized in a GOP of (a) 2, (b) 4 and (c) 8. The graph includes DISCOVER [13], H.264/AVC Intra [35], the proposed feedback-free DVC system, our previous hash-based WZ codec from [7] and the proposed DVC using feedback-channel-based rate allocation.



a



b



c

Figure 9 Compression performance comparison on Soccer in a GOP of (a) 2, (b) 4 and (c) 8. The presented systems are DISCOVER [13], H.264/AVC Intra [35], the proposed feedback-free DVC, the hash-based WZ video codec proposed in [7] and the proposed DVC system with all three coding modes using rate allocation with feedback.

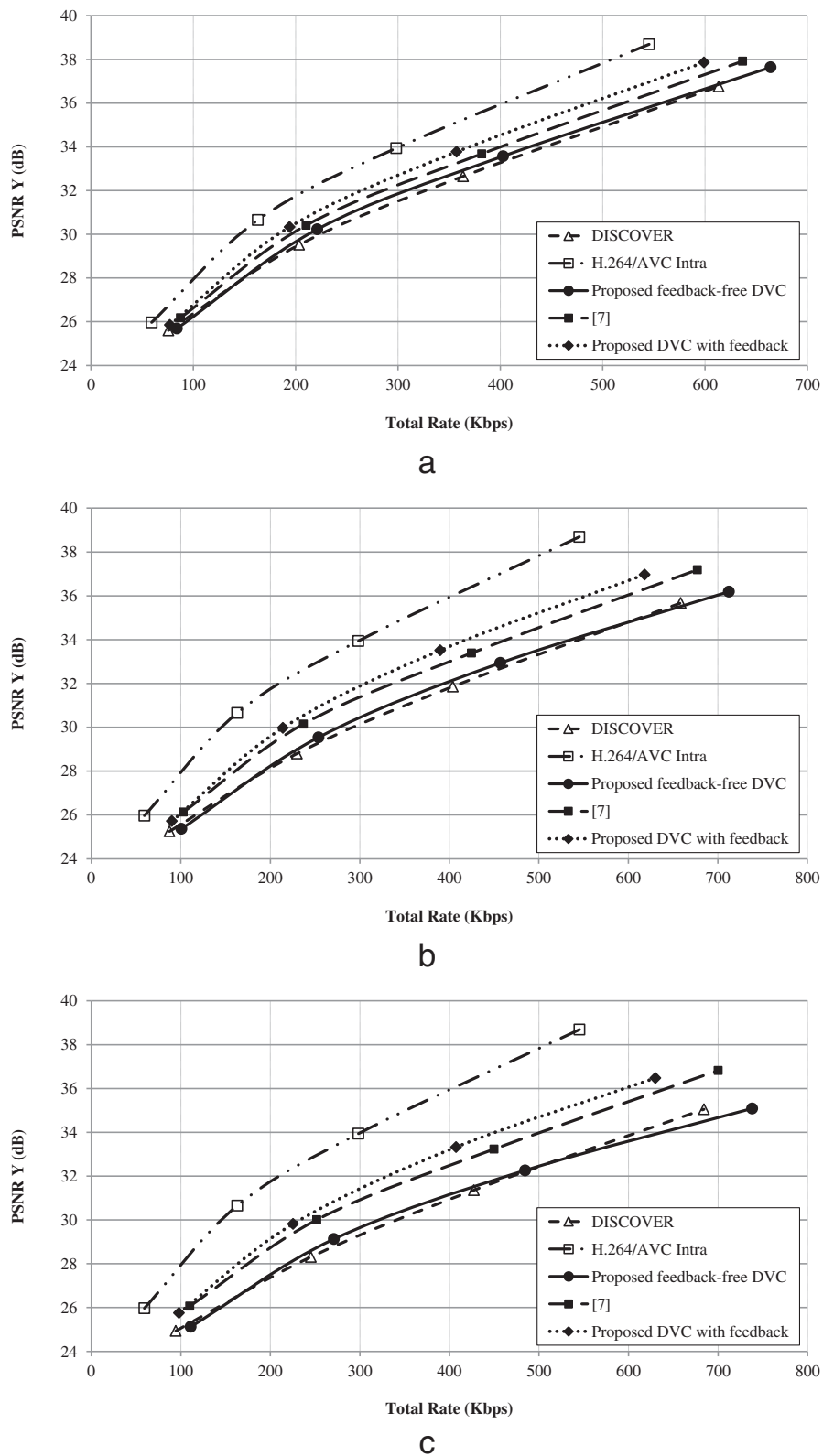


Figure 10 The relative RD performance on Football in a GOP of (a) 2, (b) 4 and (c) 8. The RD performance is shown for DISCOVER [13], H.264/AVC Intra [35], the proposed feedback-free WZ codec, our previous hash-based DVC system from [7] and the proposed DVC system with feedback.

particularly well compared to alternative DVC systems when the motion content is complex.

It is noteworthy that the reported average compression improvements over DISCOVER constitute an important achievement seeing the fact that the proposed system does not employ a feedback channel. Albeit rate losses versus DISCOVER are observed, they are not dramatic.

4.5 Complexity assessment

Since low encoding complexity is the prime motivation for DVC, an evaluation of the complexity characteristics of the proposed feedback-channel-free architecture is unavoidable. To this end, the methodology used in [13,42] is adhered to; that is, the encoding and decoding complexity are evaluated using execution time measurements under regulated conditions^b.

In addition to a regular TDWZ coding part, that is, block-based DCT, quantization and SW encoding, the proposed feedback-free DVC encoder also performs (1) coarse SI approximation, (2) bit-plane entropy estimation and (3) mode decision. Moreover, the MSI, as well as the intra-coded bit-planes, undergo (4) binary arithmetic coding. On the other hand, bit-planes belonging to skipped frequency bands are not coded at all, which has a beneficial impact on the encoding complexity. All things considered, the encoding complexity of proposed feedback-free system is expected to be higher than a comparable architecture with feedback.

On the other hand, the pursued strategy is expected to have a beneficial effect on the decoding complexity. A decoder-driven rate control scheme is based on repeated ‘request-and-decode’ operations until decoding proves successful. Although decoder-driven rate control guarantees decoding at a minimal SW rate, the series of decoding attempts severely increases the decoding complexity. In contrast, encoder-driven rate control assigns a fixed ‘take-it-or-leave-it’ amount of non-systematic information and only one decoding attempt is made. Hence, the decoding complexity of the proposed feedback-free system is expected to be significantly lower than that of the feedback-based systems.

4.5.1 Encoding complexity

The execution times (s) for encoding the entire Foreman and Soccer sequences with H.264/AVC Intra and the proposed feedback-free WZ codec are compared in Table 5. Results are presented for a GOP of 2 and 8, at every considered RD point, determined by the quantization parameter (QP) for the key frames and the QM used for the WZ frames.

The total execution time C_{FBB} of the proposed encoder is split into separate components. The part required to code the key frames is denoted by C_{Key} . The complexity contribution C_{WZ} encompasses the block-based DCT, quantization, as well as the rate control algorithm, the mode decision process and the intra coding of the MSI. Additionally, the result of the mode decision process is

Table 5 Encoding execution time (s) for encoding the entire Foreman and Soccer sequences

		AVC Intra		Proposed feedback-free DVC								
		QP										
		Intra	C_{Intra} (s)	QM	C_{Key} (s)	C_{WZ} (s)	C_{SI} (s)	C_{Hash} (s)	C_{FBB} (s)	$\frac{C_{\text{FBB}}}{C_{\text{Intra}}}$	$\frac{C_{\text{FBB+RA}}}{C_{\text{Intra}}}$	
Foreman	GOP2	40	14.22	1	7.01	0.95	0.99	0.70	9.65	0.68	0.61	
		34	16.54	4	7.83	1.56	0.97	0.07	10.43	0.63	0.57	
		29	18.97	7	8.95	2.41	0.95	0.07	12.37	0.65	0.60	
		25	21.92	8	10.45	2.94	0.94	0.07	14.40	0.66	0.61	
	GOP8	40	14.22	1	1.71	1.65	1.57	0.12	5.04	0.35	0.25	
		34	16.54	4	2.02	2.90	1.63	0.12	6.66	0.40	0.31	
		29	18.97	7	2.29	4.05	1.57	0.11	8.02	0.42	0.34	
		25	21.92	8	2.62	4.99	1.56	0.12	9.29	0.42	0.35	
Soccer	GOP2	44	13.05	1	6.79	0.89	0.92	0.07	8.66	0.66	0.59	
		36	14.49	4	7.35	1.65	0.95	0.07	10.02	0.69	0.63	
		31	16.12	7	7.82	2.38	0.91	0.07	11.17	0.69	0.64	
		25	20.72	8	9.79	2.90	0.91	0.07	13.66	0.66	0.62	
	GOP8	44	13.05	1	1.62	1.57	1.61	0.11	4.91	0.38	0.25	
		36	14.49	4	1.95	2.75	1.58	0.11	6.38	0.44	0.33	
		31	16.12	7	2.03	4.07	1.55	0.11	7.76	0.48	0.39	
		25	20.72	8	2.52	5.05	1.56	0.12	9.25	0.45	0.37	

Comparison between H.264/AVC Intra (C_{Intra}), the proposed feedback-free WZ codec (C_{FBB}) and the proposed feedback-free WZ codec using reference frame averaging ($C_{\text{FBB+RA}}$).

also covered by C_{WZ} , that is, SW or intra coding of the bit-planes depending on the selected mode. The component C_{SI} corresponds to the time required to generate the coarse approximation of the SI at the encoder, while C_{Hash} represents the time to form and code the hash frames.

Table 5 learns that the encoding complexity C_{FBF} of the proposed feedback-free system is still significantly lower than C_{Intra} , the encoding complexity of H.264/AVC Intra. Specifically, the complexity ratio C_{FBF}/C_{Intra} fluctuates around 65% to 69% for all RD points of Foreman and Soccer in a GOP of 2. The larger the GOP size, the larger the complexity reduction of DVC over conventional coding solutions, as more frames are coded using the WZ principle. Table 5 shows that the ratio C_{FBF}/C_{Intra} ranges from 42 down to 35% for Foreman and from 48% to 38% for Soccer in a GOP of 8.

To illustrate the consequence in terms of added complexity of the proposed hash-based motion estimation approach to estimate the SI at the encoder, the encoding time ratio of the proposed feedback-free system using reference frame averaging and H.264/AVC Intra ($C_{FBF + RA}/C_{Intra}$) is presented as well. As expected, Table 5 shows that the encoding complexity when the SI is approximated by the average of the corresponding reference frames is lower than the proposed method, including motion estimation. The reduction in relative encoding complexity of reference frame averaging with respect to H.264/AVC Intra ranges from 4% to 13%.

4.5.2 Impact of encoder-driven rate control and mode decision on the decoding complexity

As explained above, the complexity associated with SW decoding is drastically reduced for the proposed system due to the reduction in soft decoding runs. Additionally, the proposed feedback-free architecture includes a second source of SW decoding complexity reduction, namely, the presence of skip and intra coding modes. The proposed system implements SW coding by means

of LDPCA [38] channel codes, which uses the message passing algorithm for decoding and is vastly more complex than binary arithmetic decoding. Skipped bands are simply replaced by the corresponding bands in the SI, further reducing the complexity of the decoder.

To quantify the aforementioned complexity reduction, three different WZ architectures are compared. The first system is the proposed feedback-free DVC system, including all modes. The second system is identical to the first, save for the encoder-driven rate control, which has been replaced by a decoder-driven scheme using feedback. A comparison between the two focuses on the effect of feedback-channel suppression on the SW decoding complexity. The third system performs decoder-driven rate control but the skip and intra modes are stripped. The last system provides insight in the effect of the coding modes on the decoding complexity.

For all three WZ codecs, the execution time for decoding all bit-planes of the WZ frames of Foreman and Soccer in a GOP of 2 and 8 at every considered RD point is summarized in Table 6. As expected, the feedback-free version operates a little under 10% of the complexity of the version with feedback and including all coding modes, and even under 5% of the complexity of the system with feedback and SW coding only; this both for Foreman and Soccer, GOP2 and GOP8.

5 Conclusions

This work presented a novel encoder-driven rate control solution for unidirectional DVC. The proposed scheme first creates an encoder-side approximation of the SI available to the decoder. As an alternative to reference frame averaging, a novel low-complexity SI estimation technique is proposed to imitate the hash-based SI generation technique used at the decoder. Next, the encoder determines the theoretical lower bound to represent the transformed and quantized original WZ frames, given the estimated SI. To this end, a SID correlation channel is estimated between the original frame and the SI, from

Table 6 Bit-plane decoding execution time (s) for Foreman and Soccer, GOP of 2 and 8

	RD	GOP2					GOP8				
		C_{FBF}	$C_{FB + M}$	C_{FB}	$\frac{C_{FBF}}{C_{FB+M}}$	$\frac{C_{FBF}}{C_{FB}}$	C_{FBF}	$C_{FB + M}$	C_{FB}	$\frac{C_{FBF}}{C_{FB+M}}$	$\frac{C_{FBF}}{C_{FB}}$
Foreman	1	7.02	93.50	335.16	0.08	0.02	18.85	293.15	644.16	0.06	0.03
	2	31.26	303.53	746.78	0.10	0.04	66.74	765.75	1724.90	0.09	0.04
	3	68.42	808.87	1440.34	0.08	0.05	116.46	1591.80	2650.03	0.07	0.04
	4	78.56	1061.49	2021.75	0.07	0.04	134.00	2093.32	3743.83	0.06	0.04
Soccer	1	8.90	229.00	386.40	0.04	0.02	21.36	537.63	831.98	0.04	0.03
	2	28.45	515.77	973.47	0.06	0.03	65.46	1130.35	1982.65	0.06	0.03
	3	41.87	922.29	1595.61	0.05	0.03	89.28	1810.61	3109.63	0.05	0.03
	4	47.21	1022.06	2221.39	0.05	0.02	99.96	1966.63	4206.63	0.05	0.02

The presented systems include (1) the proposed feedback-free DVC architecture with all three coding modes available (C_{FBF}), (2) the identical system with mode decision but with feedback-based rate control (C_{FB+M}) and (3) the matching architecture with feedback and only the SW coding mode (C_{FB}).

which the conditional entropy is derived for every bit-plane. To increase the performance, the proposed system features multiple coding modes. The original WZ frames, together with the coarse SI estimate and the derived correlation model forms the basis of the mode selection process. At a frequency band level, bands for which the quality of the SI is expected to be high are *skipped*. At a bit-plane level, bit-planes for which the SI is believed to be of low quality, *intra* coding is applied. The remaining bit-planes are coded using *SW* coding principles. In this context, the final *SW* rate is adjusted using a novel formula, as to limit the number of bit-planes that fail to decode. At the decoder, an efficient SI refinement strategy is exploited. At every SI refinement stage, a *SW* coded bit-plane that failed to decode properly at any previous level is decoded anew, given the updated SI of higher quality that is available at the current refinement level. Experimental results illustrated the effect of the proposed SI approximation technique, the proposed selection of distinct coding modes and mode decision process. From an RD perspective, the proposed feedback-free WZ architecture clearly outperforms the feedback-based benchmark DISCOVER codec, on high-motion sequences. When the motion content is medium or low, the proposed system outperforms DISCOVER at low rates but is outperformed at medium, and in particular high rates, albeit not in crushing fashion. Finally, the encoder complexity of the proposed system is still significantly lower than the complexity associated with H.264/AVC Intra only coding. On the other hand, the decoding complexity that corresponds to the bit-plane decoding part is significantly reduced by the encoder-drive rate control scheme with a 'take-it-or-leave-it' approach instead of request-and-decode rounds in feedback-based decoder-driven rate control, as well as by the inclusion of the skip and intra coding modes.

Endnotes

^aThis paper has been presented in part in the Proceedings of SPIE, 2012 [12]. ^bThe execution time tests used the executables of the JM implementation of H.264/AVC and the proposed system with were conducted under the same hardware and software conditions. The employed hardware was a personal computer with Intel® Core™ i7 at 2.2 GHz and 16 GB of RAM. The executables were obtained using the Visual Studio C++ v8.0 compiler in release mode and run under the Windows 7 operating system.

Abbreviations

AC: Alternating current; AVC: Advanced video coding; CCE: Correlation channel estimation; DC: Direct current; DCT: Discrete cosine transform; DISCOVER: Distributed coding for video services; DVC: Distributed video coding; GOP: Group of pictures; LDPCA: Low-density parity-check accumulate; LLR: Log likelihood ratio; MCI: Motion-compensated interpolation; ME: Motion estimation; MSE: Mean square error; MSI: Mode signalling information; OBME: Overlapped block motion estimation;

PDF: Probability density function; PRISM: Power-efficient robust high-compression syndrome-based multimedia coding; PSNR: Peak signal-to-noise ratio; QCIF: Quarter common intermediate format; QM: Quantization matrix; QP: Quantization parameter; RA: Reference frame averaging; RD: Rate-distortion; SAD: Sum of the absolute differences; SI: Side-information; SIRL: Side-information refinement level; SW: Slepian-Wolf; TDWZ: Transform-domain Wyner-Ziv; WZ: Wyner-Ziv.

Competing interests

Parts of the research presented in this article have been filled by IBBT under patent applications EP07120604.9 (T Clerckx, A Munteanu, Motion estimation and compensation process and device, November 2007) and PCT/EP2011/071296 (N Deligiannis, A Munteanu, J Barbarien, Method and device for correlation channel estimation, November 2011).

Acknowledgements

This work is supported by the Fund for Scientific Research-Flanders (projects G004712N and G014610N) and the iMinds Institute (ISBO Project Smartcam and the ICON project 'Little Sister: low cost monitoring for care and retail').

Received: 15 April 2013 Accepted: 25 September 2013

Published: 17 October 2013

References

1. D Slepian, JK Wolf, Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory* **19**, 471–480 (1973)
2. AD Wyner, J Ziv, The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **22**, 1–10 (1976)
3. Z Xiong, A Liveris, S Cheng, Distributed source coding for sensor networks. *IEEE Signal Process. Mag.* **21**, 80–94 (2004)
4. R Puri, A Majumdar, K Ramchandran, PRISM: a video coding paradigm with motion estimation at the decoder. *IEEE Trans. Image Process.* **16**, 2436–2448 (2007)
5. B Girod, A Aaron, S Rane, D. Rebollo-Monedero, Distributed video coding. *Proc. IEEE* **93**, 71–83 (2005)
6. F Pereira, L Torres, C Guillemot, T Ebrahimi, R Leonardi, S Klomp, Distributed video coding: selecting the most promising application scenarios. *Signal Process. Image Commun.* **23**, 339–352 (2008)
7. N Deligiannis, F Verbist, A Iossifides, J Slowack, R Van de Walle, P Schelkens, A Munteanu, *Wyner-Ziv video coding for wireless lightweight multimedia applications*. *EURASIP Journal on Wireless Communications and Networking*, Special Issue on Recent Advances in Mobile Lightweight Wireless Systems, 2012
8. N Deligiannis, F Verbist, J Barbarien, J Slowack, R Van de Walle, P Schelkens, A Munteanu, Distributed coding of endoscopic video, in *IEEE International Conference on Image Processing (ICIP)*, (Brussels, 11–14 September 2011)
9. C Brites, J Ascenso, JQ Pedro, F Pereira, Evaluating a feedback channel based transform domain Wyner-Ziv video codec. *Signal Process. Image Commun.* **23**, 269–297 (2008)
10. L Stankovic, V Stankovic, S Wang, S Cheng, Correlation estimation with particle-based belief propagation for distributed video coding, in *IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, (Prague, 22–27 May 2011), pp. 1505–1508
11. N Deligiannis, J Barbarien, M Jacobs, A Munteanu, A Skodras, P Schelkens, Side-information dependent correlation channel estimation in hash-based distributed video coding. *IEEE Trans. Image Process.* **21**, 1934–1949 (2012)
12. F Verbist, N Deligiannis, SM Satti, A Munteanu, P Schelkens, Iterative Wyner-Ziv decoding and successive side-information refinement in feedback channel-free hash-based distributed video coding, in *Proceedings of SPIE 8499*, Applications of Digital Image Processing XXXV, 849900, (San Diego, CA, 2012)
13. X Artigas, J Ascenso, M Dalai, S Klomp, D Kubasov, M Quaret, The DISCOVER codec: architecture, techniques and evaluation, in *Picture Coding Symposium (PCS)* (Lisboa, 07–09 November 2007)
14. C Brites, F Pereira, Encoder rate control for transform domain Wyner-Ziv coding, in *IEEE International Conference on Image Processing (ICIP)*. (San Antonio, TX, 16–19 September 2007), pp. 5–8
15. C Brites, F Pereira, Probability updating for decoder and encoder rate control turbo based Wyner-Ziv video coding, in *IEEE International Conference on Image Processing (ICIP)*. (Hong Kong, 26–29 September 2010), pp. 3737–3740
16. M Morbee, J Prades-Nebot, A Pizurica, W Phillips, Rate allocation for pixel-domain distributed video coding without feedback channel, in *IEEE*

- International Conference on Acoustics Speech, and Signal Processing (ICASSP)*, (15–20 April 2007), pp. 521–524
17. ABB Adikari, WAC Fernando, WARJ Weerakkody, W-Z Iterative, decoding for unidirectional distributed video coding. *IEE Electronics Letters* **43**, 93–95 (2007)
 18. WARJ Weerakkody, WAC Fernando, ABB Adikari, Unidirectional distributed video coding for low cost video encoding. *IEEE Trans. Consum. Electron* **53**, 788–795 (2007)
 19. JL Martinez, G Fernandez-Escribano, H Kalva, WARJ Weerakkody, WAC Fernando, A Garrido, Feedback free DVC architecture using machine learning, in *IEEE International Conference on Image Processing (ICIP)*, (San Diego, CA, 12–15 October 2008), pp. 1140–1143
 20. C Yaacoub, J Farah, B Pesquet-Popescu, Feedback channel suppression in distributed video coding with adaptive rate allocation and quantization for multiuser applications. *EURASIP Journal on Wireless Communications and Networking* **2008**, 1–13 (2008)
 21. X Artigas, L Torres, Improved signal reconstruction and return channel suppression in distributed video coding systems, in *47th International Symposium ELMAR*, (Zadar, 08–10 June 2005), pp. 53–56
 22. S Cheng, Z Xiong, Successive refinement for the Wyner-Ziv problem and layered code design. *IEEE Trans. Signal Process.* **53**, 3269–3281 (2005)
 23. C Brites, F Pereira, An efficient encoder rate control solution for transform domain Wyner-Ziv video coding. *IEEE Transactions on Circuits and Systems for Video Technology* **21**, 1278–1292 (2011)
 24. D Chen, D Varodayan, M Flierl, B Girod, Wyner-Ziv coding of multiview images with unsupervised learning of disparity and Gray code, in *IEEE International Conference on Image Processing (ICIP)*, (San Diego, CA, 12–15 October 2008), pp. 1112–1115
 25. D Kubasov, J Nayak, C Guillemot, Optimal reconstruction in Wyner-Ziv video coding with multiple side information, in *IEEE Multimedia Signal Processing Workshop (MMSp)*, (Chania, 01–03 October 2007), pp. 251–254
 26. C Berrou, A Glavieux, P Thitimajshima, Near shannon limit error-correcting coding and decoding: Turbo codes, in *IEEE International Conference on Communications (ICC)*, (Geneva, 23–26 May 1993), pp. 1064–1070
 27. W-J Chien, LJ Karam, BLAST-DVC, bitplane selective distributed video coding. *Multimedia Tools and Applications* **48**, 437–456 (2010)
 28. S Mys, J Slowack, J Škorupa, N Deligiannis, P Lambert, A Munteanu, R Van de Walle, Decoder-driven mode decision in a block-based distributed video codec. *Multimedia Tools and Applications* **58**, 239–266 (2012)
 29. T Do, HJ Shim, B Jeon, Motion linearity based skip decision for Wyner-Ziv coding, in *IEEE International Conference on Computer Science and Information Technology*, (Beijing, 08–11 August 2009), pp. 410–413
 30. J Ascenso, F Pereira, Low complexity intra mode selection for efficient distributed video coding, in *IEEE International Conference on Multimedia and Expo (ICME)*, (Cancun, 28 June–03 July 2009), pp. 101–104
 31. S Mys, J Slowack, J Škorupa, P Lambert, R Van de Walle, Introducing skip mode in distributed video coding. *Signal Process. Image Commun.* **24**, 200–213 (2009)
 32. N Deligiannis, A Munteanu, T Clerckx, J Cornelis, P Schelkens, On the side-information dependency of the temporal correlation in Wyner-Ziv video coding, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Taipei, 19–24 April 2009), pp. 709–712
 33. Y Steinberg, N Merhav, On successive refinement for the Wyner-Ziv problem. *IEEE Trans. Inf. Theory* **50**, 1636–1654 (2004)
 34. N Deligiannis, F Verbist, J Slowack, R Van, P de Walle, AM Schelkens, Joint successive correlation estimation and side information refinement in distributed video coding, in *20th European Signal Processing Conference (EUSIPCO)*, (Bucharest, 27–31 August 2012), pp. 569–573
 35. T Wiegand, GJ Sullivan, G Bjontegaard, A Luthra, Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology* **13**, 560–576 (2003)
 36. A Aaron, S Rane, E Setton, B Girod, Transform-domain Wyner-Ziv codec for video, in *SPIE Visual Communications and Image Processing Conference, VCIP*, (San Jose, CA, 20–22 January 2004), pp. 520–528
 37. C Brites, J Ascenso, F Pereira, Improving transform domain Wyner-Ziv video coding performance, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Toulouse, 14–19 May 2006)
 38. D Varodayan, A Aaron, B Girod, Rate-adaptive codes for distributed source coding. *Signal Process.* **86**, 3123–3130 (2006)
 39. H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions for Circuits and Systems for Video Technology* **17**, 1103–1120 (2007)
 40. J Škorupa, J Slowack, S Mys, P Lambert, R Van de Walle, C Grecos, Stopping criteria for turbo coding in a Wyner-Ziv video codec, in *Picture Coding Symposium (PCS)*, (Chicago, IL, 06–08 May 2009), pp. 1–4, 6–8
 41. G Bjontegaard, *Calculation of average PSNR differences between RD-curves* (ITU-T Video Coding Experts Group (VCEG), Austin, TX, 2001). Document VCEG-M33, April 2001
 42. 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, 17–19 December 2007), pp. 801–815

doi:10.1186/1687-6180-2013-156

Cite this article as: Verbist et al.: Encoder-driven rate control and mode decision for distributed video coding. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:156.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com