Low Complexity Hybrid Rate Control Schemes for Distributed Video Coding

Mohamed Haj Taieb, Jean-Yves Chouinard and Demin Wang

***Abstract*—Distributed video coding is a video paradigm where most of the computational complexity can be transfered from video encoders to the decoders. This allows for video sequences transmission involving inexpensive encoders and powerful central decoders. Unfortunately, due to typically numerous feedback requests and needed decoder run cycles, this often leads to unacceptably long decoding latencies. One approach to addressing the latency problem consists in estimating an initial number of parity bit chunks (INC) that are then sent at once to reduce the number of these decoders run cycles. The challenge is to properly estimate as accurately as possible the INC without neither underestimation nor overestimation. This paper proposes two INC estimation techniques based on the temporal correlation between successive Wyner-Ziv frames and on the correlation between the different bit-planes.**

***Index Terms*—Distributed video coding, hybrid rate control, feedback channel, rate estimation**

I. Introduction

DIGITAL video coding standards are evolving to achieve high compression performances using sophisticated and increasingly complex techniques for accurate motion estimation and motion compensation. These techniques are executed at the encoder, resulting in computationally consuming video encoding tasks. The decoder, on the other hand, can easily reconstruct a video sequence by exploiting the motion vectors computed at the encoder. This computational inbalance is well suited for common video transfer applications such as broadcasting and video streaming, where the encoder typically benefits from high computational means to compress the video sequence only once and then to send it to many computationally limited low cost devices.

However, with the emergence of wireless surveillance locally distributed cameras, cellular interactive video utilities, and many other applications involving several low cost video encoders at the expense of high complexity central decoder, traditional video encoding standards (e.g. H.264/AVC standard [1]) have been revised and the encoder-decoder task repartition has been reversed. Slepian and Wolf information-theoretic approach to lossless coding for correlated distributed sources [2] and its extension to lossy source coding with side information at the decoder, as introduced by Wyner and Ziv [3], constitute the theoretical framework for distributed source coding. This gave birth to a wide new field of applications, such as distributed video coding (DVC).

Manuscript received June 25, 2012; revised July 30, 2012. This work is supported by the Communications Research Centre Canada and the Natural Sciences and Engineering Research Council of Canada.

M. Haj Taieb and J.-Y. Chouinard are with the Department of Electrical and Computer Engineering, Laval University, Quebec, QC, G1V 0A8 Canada e-mail: mohamed.haj-taieb.1@ulaval.ca and jean-yves.chouinard@gel.ulaval.ca.

D. Wang is with Advanced Video Systems, Communications Research Centre Canada, Ottawa, ON, K2H 8S2 Canada e-mail: Demin.Wang@crc.ca.

Although the DVC paradigm have raised an important body of research developments to achieve competitive R-D performances, the inherently large decoding complexity remains unacceptable for most practical DVC applications. For turbo coding based DVC systems, unacceptly long delays are caused by the several required runs of turbo decoding using parity bit chunks gradually sent upon feedback requests. Therefore, limitation of the feedback channel is crucial for the design of low latency real-time DVC applications. LDPC based DVC coding schemes also require low computational complexity decoders.

The way the feedback channel is used by the encoder-decoder pair, highlights the trade-off between low latency and video sequence reproduction quality. On one hand, the feedback channel is useful to insure decoder rate control with a minimum forward rate, but at the price of several decoding loops. On the other hand, the encoder rate control without a feedback channel reduces drastically the system delay: the encoder needs to estimate the number of parity bits needed by the turbo decoder. If the estimated number of bits exceeds the minimum really needed, this increases the bit rate while if the number of parity bits is underestimated, the turbo decoding will not converge, leading potentially to visual artifacts in the reconstructed frames.

Between these two rate control schemes, a hybrid (trade-off) technique can be adopted where the encoder and the decoder cooperate to estimate the minimal rate using the feedback channel. In section 2, a review of the Discover DVC architecture is presented along with techniques for rate control. In section 3, a hybrid rate control technique based on the temporal correlation between the final number of parity bit chunks (FNC) is described. Another hybrid rate control technique based on the correlation at the bit-plane level is also proposed. A comparative study between the different estimators is presented.

II. Distributed video coding scheme

A. Discover DVC codec architecture

The architecture of the Discover WZ system based on turbo coding [4] is depicted in Figure 1. It is based on the Stanford WZ codec and includes several means for improving the rate distortion performance. The key frames are H.264/AVC intra encoded (intraframes) and transmitted to generate the side information to decode the Wyner-Ziv frames (interframes frames). The interframes are compressed using a 4x4 block discrete cosine transform (DCT). The DCT coefficients are fed to a uniform quantizer. The quantized coefficients are then fed to a turbo encoder consisting of the two constituent rate 1/2 recursive convolutional encoders (RSCs). Each of the two RSCs associates a parity bit to the quantized DCT-coefficients. To achieve compression of

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the transmitted data, the systematic bits are discarded since the decoder has already an interpolated version of the even frames. Moreover, the parity bits are stored in a buffer and are sent gradually, packet by packet, upon decoder feedback requests according to a periodic puncturing pattern. The feedback channel allows adapting the forward transmission rate to the changing virtual channel conditions. This also implies several turbo decoder runs. To alleviate the decoder computational hurdle, an initial number of parity bit packets is estimated by an hybrid encoder/decoder rate control mechanism [5]. These parity bit packets are sent once to the decoder and eventually subsequent packets will be sent if needed.

At the decoder, an interpolated version of the current WZ frame is produced using the already received neighboring key frames. The motion compensated temporal interpolation technique (MCTI) presented in [8], known as *bidirectional motion estimation with spatial smoothing* (BiMESS), was adopted for most DVC architectures. The BiMESS performances are improved using a hierarchical coarse-to-fine approach in bidirectional motion estimation [6] and sub-pixel precision for motion search [7]. The interpolated frame is then DCT transformed and the DCT coefficients represent the side information used to decode the WZ frames. The WZ DCT coefficients are modeled as the input of a virtual channel and the side information as its output. During the turbo decoding process, a Laplacian model is assumed for this virtual channel. The estimation of Laplacian distribution parameter α is based on an online correlation noise modeling technique at the coefficient/frame level: parameter α is estimated for each coefficient band of each frame [8].

The turbo decoder computes the systematic log-likelihood ratios. The systematic information is corrupted by a Laplacian noise whose parameter is, beforehand, online estimated (without using original data). Actually, there are no systematic bits and the side information is used instead. The received parity bits along with the side information are fed to the turbo decoder. After a number of iterations, the log-likelihood ratios are computed and then the bitplane value is deducted. To estimate the decoded bitplane error rate, without access to the original data, these log-likelihood ratios are used to compute a confidence score [5]. If this score exceeds 10^{-3}, then a parity bits request is sent back to the encoder. Otherwise, the decoding process is likely to be satisfactory. However, some errors can still persist even if the confidence score is below 10^{-3}. For this reason, an 8-bit long cyclic redundancy check (8-CRC) code is used to help detecting the remaining bitplane decoding errors. If the decoded bitplane CRC corresponds to the original data CRC, then the decoding process is considered successful, otherwise, more parity bits are requested. Using jointly the confidence score and the CRC code results in error detection performances as good as ideal error detection where the decoded bitplane is directly compared to the original bitplane [5].

After being decoded, these bitplanes are recombined to form the quantized symbols. These symbols and the side information are used to reconstruct the DCT coefficients. An optimal reconstruction function is proposed in [9] to minimize the mean squared error according to the Laplacian correlation model. For coefficients bands that have not been transmitted, the side information is directly considered in the reconstruction. Finally, an inverse 4×4 DCT is applied to the reconstructed frequency band to restore the WZ frame in the pixel domain.

B. Decoder rate control (DRC)

Decoder rate control was adopted for the first DVC implementation [10], because it resulted in the best rate-distortion performances. Excessive execution delays were experienced at the decoder: the technique did not estimate an initial number of parity bit chunks (INC) and involved sending these chunks until the decoder convergenced. There were no overestimation hence leading to the best performances. Two hybrid encoder/decoder rate control solutions to estimate the minimum rate R_{min} (or the INC) are presented in the following.

C. Hybrid rate control based on the Slepian-Wolf theorem

The hybrid rate control technique in [5] aims at evaluating the minimum parity rate R_{min} for each bitplane of each band. The decoder estimates the correlation noise model parameter α and sends it back to the encoder. Knowing the original data and the model parameter, the encoder estimates at first the probability of crossover p_{co} and then the minimal rate R_{min} according to the Slepian-Wolf Theorem:

$$R_{min} = H\left(X|\,Y\right) = -p_{co}\log_2 p_{co} - \left(1-p_{co}\right)\log_2\left(1-p_{co}\right) \tag{1}$$

The crossover probability is estimated for each bitplane and corresponds to the probability that the bitplane x_{pb} is different from the estimated bitplane at the decoder $\hat{x}_{pb}$ using the side information y and the previously decoded bitplanes $\left(x_{pb-1},\ \ldots, x_2,\ x_1\right)$:

$$\hat{x}_{pb} = \arg\max_{i=0,1} \Pr\left(x_{pb} = i\,|y,\ x_{pb-1},\ \ldots, x_2,\ x_1\right) \tag{2}$$

where $\Pr\left(x_{pb} = i\,|y,\ x_{pb-1},\ \ldots, x_2,\ x_1\right)$ designates the *a posteriori* probability of event $x_{pb} = i$.

An example of $\hat{x}_{pb}$ calculation is depicted in figure 2. After determining $\hat{x}_{pb} = 1$, the crossover probability $\Pr\left(x_{pb} \neq \hat{x}_{pb}\right)$ is computed:

$$\Pr\left(x_2 \neq \hat{x}_2\right) = \frac{\int_{y-B_1}^{y-B_2} f(n)dn}{\int_{y-B_1}^{y-B_3} f(n)dn} = \frac{F\left(y-B_2\right) - F\left(y-B_1\right)}{F\left(y-B_3\right) - F\left(y-B_1\right)} \tag{3}$$

where $F(n)$ is the cumulative distribution function (CDF) of the Laplacian probability density function (PDF)[1]:

$$F(n) = 0.5\left(1 + \mathrm{sign}\left(n\right) - \mathrm{sign}\left(n\right)e^{-\alpha|n|}\right) \tag{4}$$

The crossover probability p_{co} is computed at the encoder which has no knowledge about side information y: thus the next step consists in integrating over all possible values of y. Finally the average over the original WZ DCT coefficients is taken:

$$p_{co} = \frac{1}{N}\sum_{x\in WZ}\left[\int_{V_{\min}}^{V_{\max}} \Pr\left(x_{pb} \neq \hat{x}_{pb}\right)\frac{\alpha}{2}e^{-\alpha|y-x|}dy\right] \tag{5}$$

[1]The use of the CDF avoids the need to perform integration.

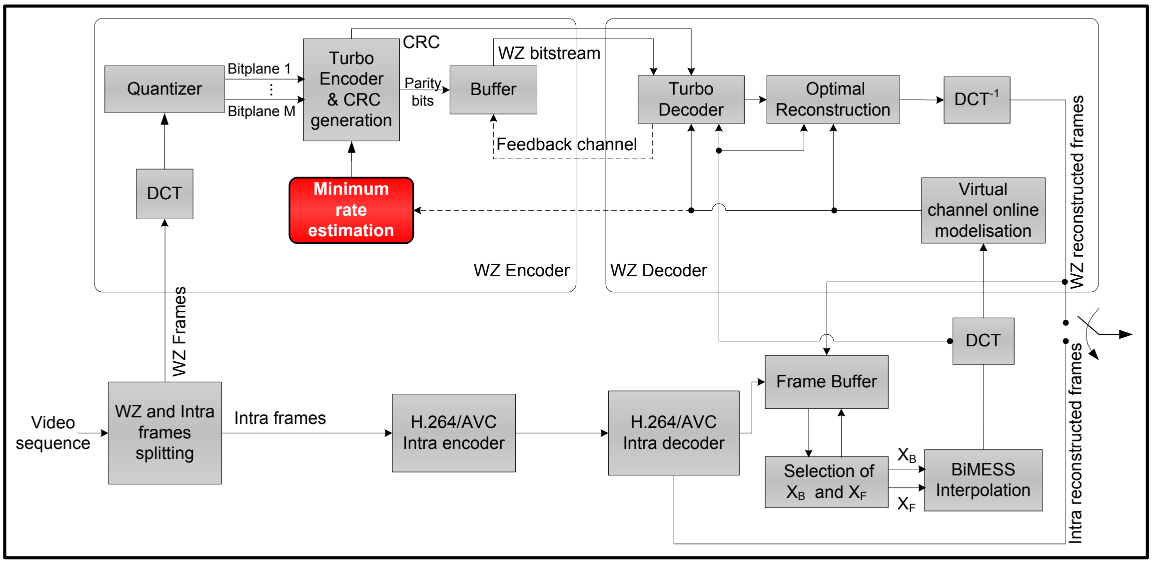

Fig. 1. Transform domain Wyner-Ziv video codec (Discover).

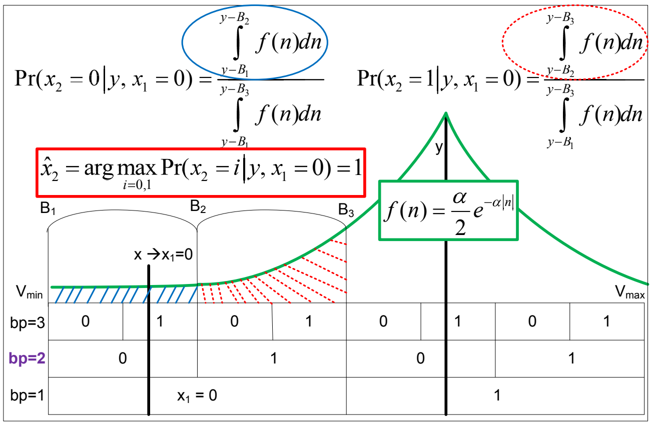

Fig. 2. *Computation of $\hat{x}_{pb=2}$ as a function of conditional probabilities for a given side information y and the previously decoded bitplane $x_{pb=1}$.*

where N is bitplane length. Thereby, the crossover probability computation p_{co} requires averaging N relatively complex integrals. This involves considerable computations at the encoder (supposed to be light in the DVC paradigm). For each DCT band, the p_{co} computation is thus given by :

$$p_{co} = \frac{1}{N} \sum_{x \in WZ} \left[\int_{V_{\min}}^{V_{\max}} \frac{F(y - B_2) - F(y - B_1)}{F(y - B_3) - F(y - B_1)} \frac{\alpha}{2} e^{-\alpha|y - x|} dy \right] \tag{6}$$

D. Low complexity hybrid rate control

The previous technique incurs some additional encoder complexity to estimate the minimum rate. In [11], a low complexity hybrid rate control technique is proposed. For each bit-plane j, of band i, the initial number of parity bits chunuks (INC) is estimated using the final number of parity bits chunks (FNC) sent for the same bit-planes in the previous 3 WZ frames:

$$\begin{aligned} INC(i,j) = &\operatorname{floor}[(1-k) \times \\ &\operatorname{median}(FNC_{-1}(i,j),\ FNC_{-2}(i,j),\ FNC_{-3}(i,j))] \end{aligned} \tag{7}$$

where k is a scale factor such as $k = 0.1$ for the first five DCT bands ($i = 1, \ldots, 5$) and $k = 0.05$ for the remaining bands. The term $(1 - k)$ prevent from estimation rate saturation.

III. Proposed Algorithms for Initial Number of Chunks (INC) Estimation

In this section, two minimum rate estimation techniques are proposed based on the temporal evolution of the final number of parity bits chunks (FNC) for each bit-plane of each DCT band.

A. Estimation algorithm based on temporal correlation (TC)

This algorithm exploits the FNC's temporal stationarity to perform a two-step estimation of the minimum rate. More specifically, consider the estimation of the INC for the sixth bit-plane of the first DCT band of the WZ frame number $t = 36$ as shown in figure 3. This figure displays the temporal evolution of the FNC determined by decoder rate control. The first step consists in computing the INC using the three previous FNC values:

$$\begin{aligned} INC_{t=36}^{(\text{step 1})}(\text{band} = 1, \text{bp} = 6) &= a \times FNC_{35}(1,6) \\ &+ a^2 \times FNC_{34}(1,6) + a^3 \times FNC_{33}(1,6) \end{aligned} \tag{8}$$

where a is a scale factor such that $a + a^2 + a^3 = 1 \Rightarrow a = 0.54$ if there is under-estimation at the previous frame ($t = 35$) and such that $a + a^2 + a^3 = 0.8 \Rightarrow a = 0.47$ if there is over-estimation to avoid saturation.

The first step estimation calculates a weighted average between the previous FNCs values. However, when there is a peak or a trough, this estimation is not close enough to the actual value. This (a peak or a trough) can be detected by observing the previous bit-plane. For instance, it can be observed from the bit-plane $bp = 5$, that there is a peak, and an offset can be computed between the first step estimated value and the actual value. This offset is expected to be found again for the bit-plane $bp = 6$. Thus the first step estimation can be adjusted as follows:

$$INC_{t=36}^{(\text{step 2})}(1,6) = INC_{t=36}^{(\text{step 1})}(1,6) + \operatorname{offset}(bp = 5) \tag{9}$$

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

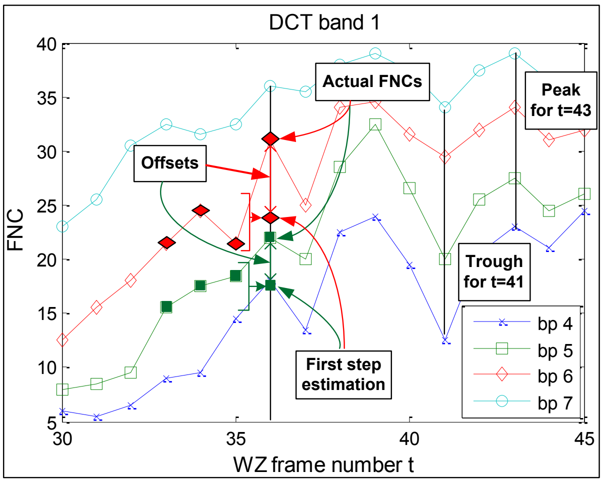

Fig. 3. *FNC temporal variation with peaks and troughs occuring for the different bit-planes. The offset between the first step estimation for bp=5 can be used to adjust the estimation for bp=6.*

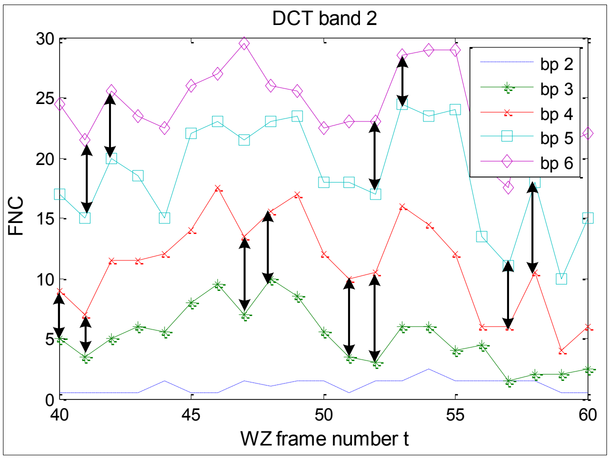

Fig. 4. *FNC offset between two successive bit-planes.*

B. Estimation algorithm based on bit-plane correlation (BP)

It is noticed from figure 4 that the offset between the FNCs of two successive bit-planes is almost the same at times t and $t+1$:

$$FNC_t(i,j) - FNC_t(i,j-1) \approx FNC_{t+1}(i,j) - FNC_{t+1}(i,j-1) \quad (10)$$

This observation is used to compute an INC estimation for a bit-plane bp based on the FNC of the previous bit-plane $bp-1$. For instance, the estimation of the INC for the second DCT band sixth bit-plane for the WZ frame $t=53$, is:

$$FNC_{53}(2,6) = \begin{cases} FNC_{53}(2,5) + a \times [FNC_{52}(2,6) - FNC_{52}(2,5)], \\ \quad \text{if } FNC_{52}(2,6) \leq INC_{52}(2,6) \\ \quad \text{or } FNC_{53}(2,5) \leq INC_{53}(2,5) \\ \text{otherwise :} \\ FNC_{53}(2,5) + [FNC_{52}(2,6) - FNC_{52}(2,5)], \end{cases} \quad (11)$$

where a is a scale factor, set to 0.5, to prevent from saturation if there is an overestimation at the sixth bit-plane of WZ frame number $t=52$ or an overestimation at the fifth bit-plane of WZ frame number $t=53$.

IV. Simulations and Discussion

In this section, the proposed estimators (EST: TC and BP) are compared with the estimator of Kubasov *et al.* [5] and the estimator of Areia *et al.* [11]. Three QCIF video sequences at 15 frames per second are considered for the simulation tests: `Foreman`, `Soccer` and `Coastguard`. These sequences are downloaded from the Discover website [4]. All the 149 frames of the sequences are considered, corresponding to 74 WZ frames. The frame size is 144 × 176 =25344 pixels, leading to bit-planes length of 25344/16=1584 bits for each DCT 4×4 component. The puncturing period length is 48 which results in a chunk size of a $(1584/48) \times 2 = 33 \times 2 = 66$ parity bits sent at each feedback request. This corresponds to 33 parity bits for each of the two RSC encoders. The estimated initial number of chunks (INC) involves sending INC ×66 parity bits at once.

A. Estimators comparison criteria

To compare between the estimators' performances, two points have to be considered:

1) Overestimation: it engenders rate increase and the average, over the N WZ frames, number of chunks sent in excess is given by:

$$\text{Excess} = \frac{\sum_{n=1}^{N} \max\left(\left[INC^{EST}(n) - FNC^{DRC}(n)\right], 0\right)}{N} \quad (12)$$

2) Underestimation: when the INC is below the FNC the decoder will ask gradually for more parity bits chunks. For each feedback request, the turbo decoding will be launched again, thus causing delays. The average number of feedback requests over the N WZ frames is given by:

$$\text{Request} = \frac{\sum_{n=1}^{N} \max\left(\left[FNC^{DRC}(n) - INC^{EST}(n)\right], 0\right)}{N} \quad (13)$$

To assess the accuracy of the estimator as a whole, taking into account the overestimation and underestimation, the average, over the N WZ frames, absolute difference (estimation error) between the INC and the FNC obtained with decoder rate control is evaluated as:

$$\text{Difference} = \frac{\sum_{n=1}^{N} \left|FNC^{DRC}(n) - INC^{EST}(n)\right|}{N} \quad (14)$$

B. Comparison of the minimum rate estimation algorithms

Prior to comparing the estimators performances from a global point of view, figure 5 shows the temporal evolution of the TC estimated INC as well as of the FNC^{DRC} for some selected bit-planes. FNC^{DRC} indicates the target number of chunks to be estimated. It can be seen from this figure that the proposed temporal (TC) solution gives more accurate estimates than the algorithm of Areia *et. al* [11]. The TC estimator is able to follow more flexibly the rapid variation of the FNC^{DRC} since it adjusts the estimation according to the previously decoded bit-plane.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

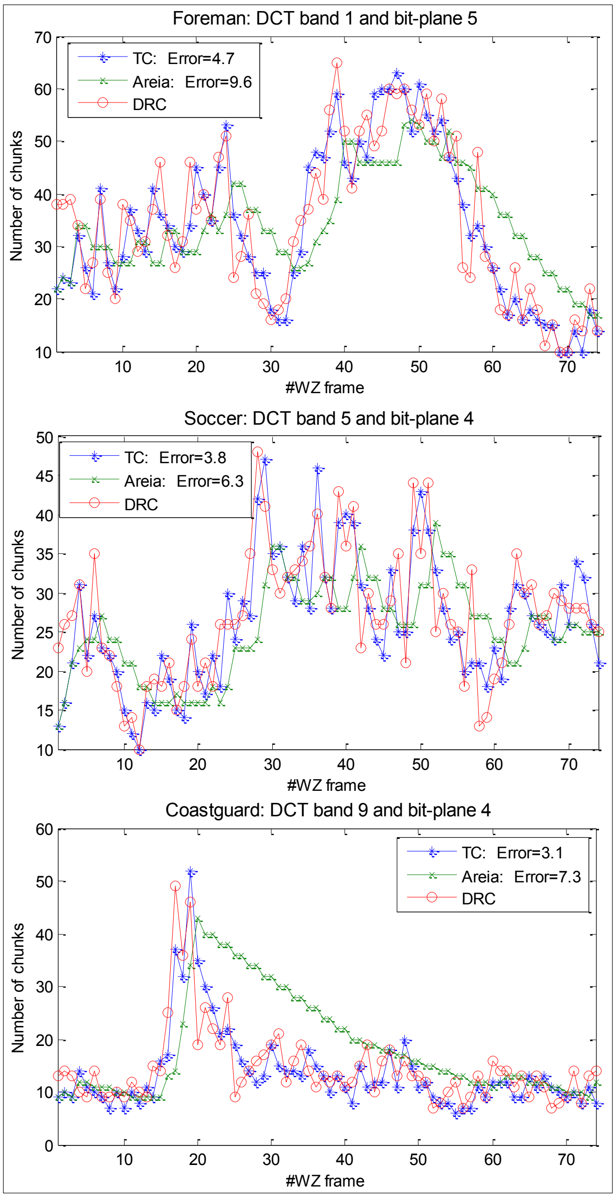

Fig. 5. *Temporal evolution of the estimated INC and the FNC^{DRC} for some bit-planes of the quantization index $Qi = 8$.*

In figure 6, a comparison of the different estimators is presented through the three previously cited criteria for the video sequences `Foreman`, `Soccer` and `Coastguard`. Eight quantization matrix indices (Qi) are considered. The estimator absolute difference criterion stipulated that the proposed solutions (especially TC) are more accurate and give closer estimates to the number of chunks required for the decoding convergence. The proposed rate estimation solutions are more accurate and the rate-distortion curves displays a reasonable rate increase caused by over-estimation. The estimator performances are then summarized in tables I and II. The first table depicts the percentage of the decoder complexity reduction compared to the DRC solution. The second table presents the percentage of the rate increase compared to the DRC solution. The proposed estimators can reduce significatly the decoder latencies (an average reduction of 87.5 % for the TC solution) without a severe impact on the rate-distorsion performances (only 8.93% rate increase).

TABLE I
Decoder complexity reduction percentage relative to decoder rate control.

	Kubasov	Areia	BP	TC
`Foreman`	53.22%	84.59%	82.33%	87.07%
`Soccer`	55.68%	86.74%	84.76%	89%
`Coastguard`	64.28%	84.75%	80.36%	86.43%
Average	57.72%	85.36%	82.48%	87.5%

TABLE II
Rate increase percentage caused by over-estimation compared to decoder rate control.

	Kubasov	Areia	BP	TC
`Foreman`	0.65%	17.09%	9.79%	10.46%
`Soccer`	0.53%	8.51%	5.87%	7.31%
`Coastguard`	1.04%	15.24%	10.87%	9.03%
Average	0.74%	13.61%	8.84%	8.93%

V. Conclusion and Future Work

In this paper new techniques for low complexity rate control are proposed. These solutions are inspired from the temporal behavior of the FNC which displays not only a temporal stationnarity between successive WZ frames but also a correlation between successive bit-planes. More precise estimation allowing lower decoding delays are obtained thanks to these techniques at the expense however of a slight rate increase. These techniques depend strongly on the hypotheses of temporal correlation and the structure of the FNC. If in some instances these hypotheses are not verified, then the estimation can be severely compromised. As future investigation, the rate estimation can be helped by a down-sampled version of the WZ frame sent and decoded firstly. In light of the obtained FNC, the INC for the remaining WZ frame can be estimated.

References

[1] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. WileyInterscience, 2003.
[2] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory 19 (4)*, pp. 471–480, July 1973.
[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory, vol. IT-22, no. 1*, pp. 1–10, January 1976.
[4] T. D. codec evaluation, "[Online]. Available: http://www.discoverdvc.org/."
[5] C. G. D. Kubasov, K. Lajnef, "A hybrid encoder/decoder rate control for Wyner-Ziv video coding with a feedback channel," in *Proceedings of Int. workshop on Multimedia Signal Processing*, Chania, Crete, Greece. October 2007, pp. 251–254.
[6] F. P. J. Ascenso, C. Brites, "Content adaptive Wyner-Ziv video coding driven by motion activity," in *IEEE International Conference on Image Processing*, Atalanta, USA. October 2006.
[7] S. Klomp, Y. Vatis and J. Ostermann, "Side Information Interpolation with Sub-pel Motion Compensation for Wyner-Ziv Decoder," *International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, Setubal, Portugal. August 2006.
[8] C. Brites and F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner Ziv Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 53, no. 2, September 2008.
[9] D. Kubasov, J. Nayak and C. Guillemot, "Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information," in *Proc. of MMSP, IEEE International Workshop on Multimedia Signal Processing*, October 2007.
[10] B. Girod, A. Aaron, S. Rane, and D. R. Monedero, "Distributed video coding," in *Proceedings IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, January 2005., pp. 71–83. Invited paper.
[11] J. D. Areia, J. Ascenso, C. Brites and F. Pereira, "Low Complexity Hybrid Rate Control for Lower Complexity Wyner-Ziv Video Decoding," in *16th European Signal Processing Conference (EUSIPCO)*, August 2008.

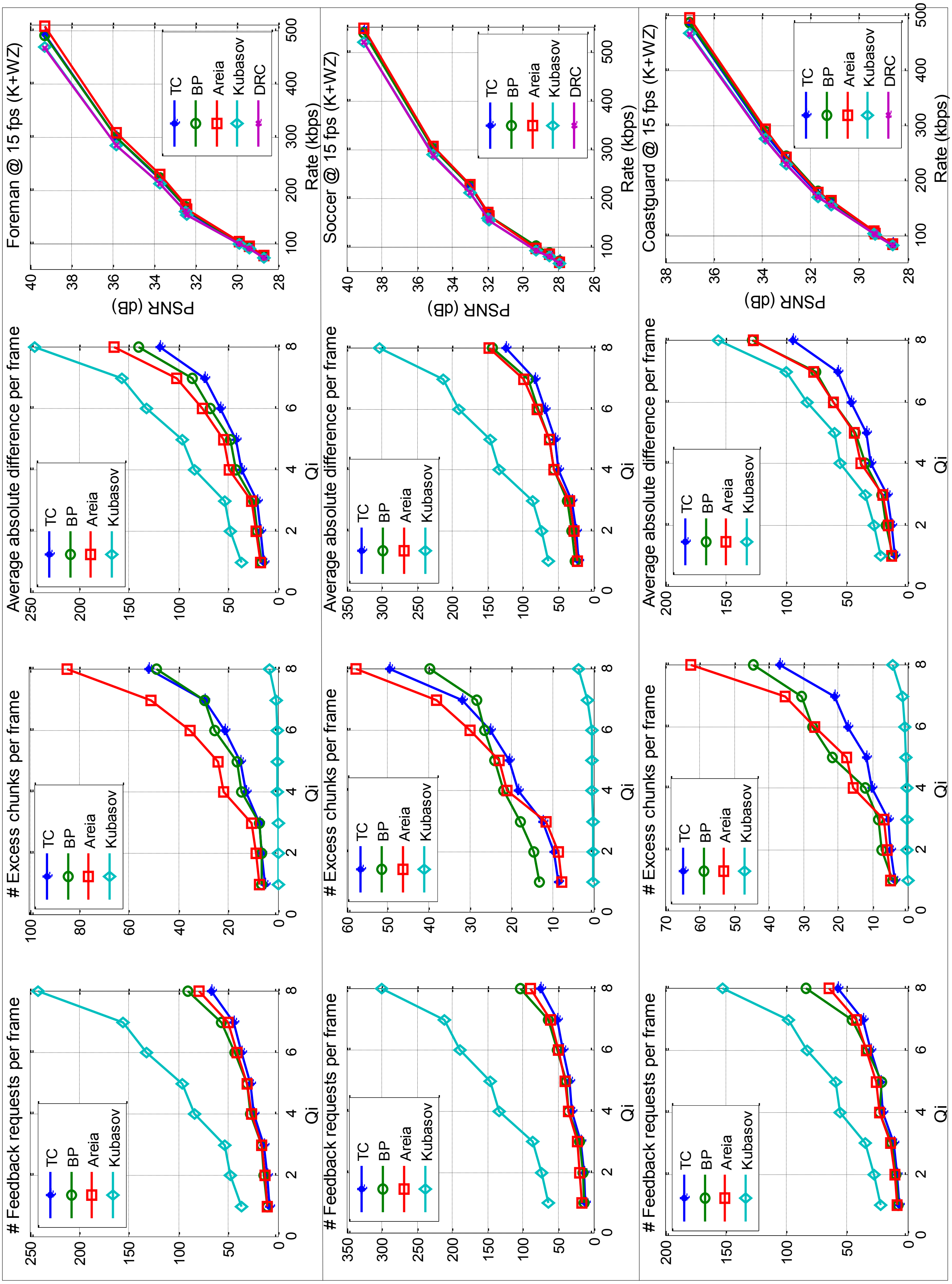

Fig. 6. *Comparison between the different R_{min} estimators performances.*

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)