16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

LOW COMPLEXITY HYBRID RATE CONTROL FOR LOWER COMPLEXITY WYNER-ZIV VIDEO DECODING

José Diogo Areia, João Ascenso, Catarina Brites and Fernando Pereira

Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
phone: + (351) 218418460, fax: + (351) 2198418472, email: fp@lx.it.pt
web: http://www.img.lx.it.pt/

ABSTRACT[1]

Wyner-Ziv (WZ) video coding is an emerging paradigm which promises to lower the encoding complexity at the cost of additional decoding complexity regarding predictive coding solutions. However, the current decoding complexity increase is huge, very likely unacceptable for the relevant applications, asking for solutions which allow reducing this cost without negatively impacting the RD performance and the encoding complexity. This paper proposes hybrid rate control solutions which allow reaching the objectives above for Stanford-like WZ video codecs.

1. INTRODUCTION

Recently, the video coding community has been confronted with a growing interest in a new video coding paradigm, the so-called distributed video coding (DVC). The Slepian-Wolf and the Wyner-Ziv (WZ) theorems play a major role in this new coding paradigm. Those theorems state that it is possible to compress (under some statistical conditions) two statistically dependent signals in a distributed way (separate encoding, joint decoding) approaching the coding efficiency of conventional predictive coding schemes (joint encoding and decoding); the Slepian-Wolf and the Wyner-Ziv theorems regard lossless and lossy source coding, respectively. The new coding paradigm may avoid the computationally intensive motion compensated temporal prediction loop at the encoder, by shifting the exploitation of the temporal redundancy to the decoder. This is a significant advantage for a large range of emerging application scenarios, notably wireless video cameras, wireless low-power surveillance, video conferencing with mobile devices, and visual sensor networks.

Following these theorems, the practical design of Wyner-Ziv video codecs, a particular case of DVC, started around 2002, after important developments in channel coding technology. As of today, the most popular WZ video codec design in the literature is the so-called Stanford WZ architecture [1]; this codec works at the frame level and is characterized by a feedback channel based decoder rate control, see Section 2 for details. One of the basic objectives of this type of WZ codec is to reach low complexity encoding at the cost of an increased decoding complexity. For feedback channel based WZ video coding solutions, such as the Stanford WZ video codec architecture, the decoding complexity is largely associated to the Slepian-Wolf decoder, notably the high number of times the turbo decoder (a typical Slepian-Wolf decoder) has to be run following the decoder requests for more bits to ‘correct’ the side information estimation. As shown in [2], the decoding complexity associated to the WZ frames is much higher than the decoding complexity of predictive video codecs; this is not a desirable feature even if some increase in the decoding complexity may be acceptable to compensate the encoding complexity reduction. The usage of the feedback channel has implications beside the obvious need for that feedback channel to be available, notably: i) this coding architecture can only be used for real-time applications scenarios; ii) the application and the video codec must be able to accommodate the delay associated to the feedback channel; and iii) the usage of the feedback channel simplifies the rate control problem since the decoder, knowing the available side information, can easily regulate the necessary bitrate.

Considering the intimate relationship between the rate control, the feedback channel, the decoding complexity and the Rate-Distortion (RD) performance, it is essential to bear in mind three types of WZ video coding rate control:

1. **Decoder rate control (DRC)** – All (or most) rate control processing is made at the decoder which means there is a need for a return channel to bring rate control data from the decoder to the encoder; this is the most common rate control solution in the WZ video coding literature [1].
2. **Encoder rate control (ERC)** – All rate control processing is made at the encoder side (there is no need for any return channel for rate control purposes); see example in [3]. For complexity constrained encoders, the encoder rate control solutions must not be too complex; this type of solutions typically shows a lower RD performance than solutions with decoder rate control since the encoder has no access to the side information.
3. **Hybrid rate control (HRC)** – Rate control processing is made both at encoder and decoder; see example in [4]. While the encoder has to make a conservative estimation of the rate needed, the decoder has the task to complement this rate, if needed, using the feedback channel. While encoder rate overestimation is paid with losses in RD performance, encoder rate underestimation is paid with increased decoding complexity and delay.

In this context, the main objective of this paper is to propose effective hybrid rate control solutions to reduce the very high WZ decoding complexity for Stanford-like architectures. Those solutions should fulfil two main requirements: i) the added encoding complexity must be negligible; and ii) the overall RD performance (negative) impact must be close to negligible.

---

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

These two requirements basically ask for a mechanism which is able to estimate, in a very simple way, the rate needed while avoiding i) overestimations to reduce the RD performance losses and ii) underestimations to maximize the decoding complexity gains. Naturally, there is a trade-off, since typically the higher the decoding complexity reduction (due to the reduction on the number of decoder requests), the higher is also the risk of overestimating the rate, which has a negative impact on the overall RD performance.

This paper is organized as follows: Section 2 briefly reviews the adopted Stanford-like WZ video codec; next, Section 3 presents the hybrid rate control (HRC) solutions proposed, and Section 4 addresses their performance evaluation. Finally, Section 5 concludes the paper.

2. WYNER-ZIV VIDEO CODEC

The Transform Domain Wyner-Ziv (TDWZ) video codec adopted and implemented by the authors is described in detail in [2]; this codec is based on the Stanford WZ video coding architecture presented in [1].

The TDWZ coding architecture illustrated in Figure 1 works as follows: a video sequence is divided into WZ frames and key frames. The key frames may be inserted periodically with a certain Group of Pictures (GOP) size or instead an adaptive GOP size selection process can be used to exploit the varying amount of temporal correlation along the video sequence. If no bitplane rate control is performed at the encoder as in [1] (i.e. the Bitplane Rate Allocation module in Fig.1 does not exist), a pure decoder rate control solution results, working in the following way:

1. The key frames are coded using a standard coding solution, e.g. H.264/AVC Intra.
2. The WZ frames are coded using a WZ coding approach; over each WZ frame, a 4×4 block-based Discrete Cosine Transform (DCT) is applied.
3. The DCT coefficients of the entire WZ frame are grouped together, according to the position occupied by each DCT coefficient within the 4×4 blocks, forming the DCT coefficients bands.
4. Each DCT band is uniformly quantized with a (varying) number of levels, setting the quality target.
5. Over the resulting quantization symbol stream, bitplane extraction is performed to form the bitplane arrays which are then independently turbo encoded. The encoder sends then to the decoder a fixed amount of parity bits for each bitplane using a certain puncturing period.
6. The decoder creates the so-called side information (SI) for each WZ frame, which should be a good estimate of the original WZ frame. The SI is created through a motion compensated frame interpolation process, using the previous and next decoded frames temporally closer to the WZ frame under coding.
7. A block-based 4×4 DCT is then carried out over the side information in order to obtain an estimate of WZ frame DCT coefficients.
8. The residual statistics between corresponding coefficients in the SI and the original WZ frame is assumed to be modelled by a Laplacian distribution whose parameter is online estimated at the decoder.
9. The decoded quantization symbol stream associated to each DCT band is obtained through an iterative turbo decoding procedure of each bitplane. Whenever the estimated bitplane error probability is higher than a predefined threshold, typically 10^{-3}, the decoder requests more parity bits from the encoder.
10. Once all decoded quantization symbol streams are obtained, the DCT coefficients are reconstructed using an optimal mean squared error estimate.
11. After all DCT coefficients bands are reconstructed, a block-based 4×4 Inverse Discrete Cosine Transform is performed and the decoded WZ frame is obtained.
12. Finally, to get the decoded video sequence, decoded key frames and WZ frames are conveniently mixed.

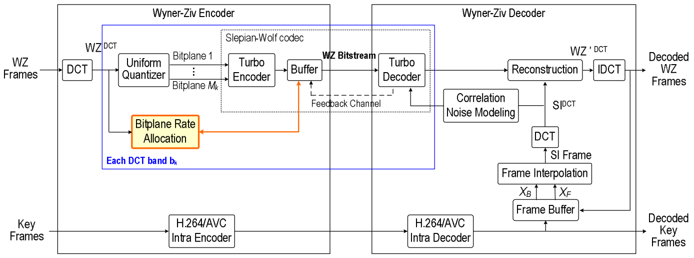

Figure 1 – Wyner-Ziv video codec architecture with low complexity hybrid rate control.

An extensive performance evaluation of the TDWZ video codec is presented in [2]. Figure 2 shows the TDWZ encoding and decoding complexity in comparison with the H.264/AVC Intra and H.264/AVC 'No Motion' standard solutions (predictions without using motion compensation) for Coast Guard QCIF@15 Hz.

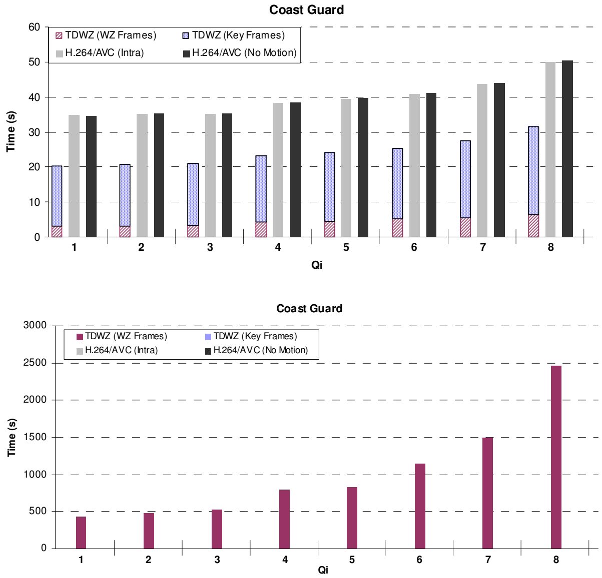

Figure 2 – TDWZ encoding and decoding complexity in comparison with H.264/AVC Intra and H.264/AVC No Motion for Coast Guard QCIF@15 Hz [2].

While it is clear that the WZ encoding complexity is lower than the alternative H.264/AVC standard solutions, the WZ decoding complexity is much higher; the H.264/AVC standard solutions decoding complexity is not visible in the chart because it is too small. In [2], it is shown that the iterative turbo decoding has always a share in the overall WZ decoding complexity higher than 80%, reaching values higher than 95% for the sequences with more complex motion such as Soccer. For these sequences, the side information quality is lower and, thus, a higher number of requests have to be made to 'correct' the poor side information. This evidence highlights the need to reduce the number of requests to significantly reduce the turbo decoding complexity. This paper proposes to reach this effect through low complexity hybrid rate control corresponding to the codec architecture in Fig. 1 with the Bitplane Rate Allocation module.

3. LOW COMPLEXITY HYBRID RATE CONTROL SOLUTIONS

This section presents the proposed hybrid rate control (HRC) algorithms which should ideally be able to fulfil the requirements defined in Section 1. In the following, the rate control acts over rate *chunks* where one chunk corresponds to a fixed amount of parity bits determined by the puncturing period; for each decoder rate request, a rate chunk is sent to the decoder.

Since the turbo codec works at the bitplane level, the proposed HRC solutions main challenge is to determine the *initial number of rate chunks* (INC) to be sent for each bitplane of each DCT band. Generally, the proposed HRC solutions are characterized by the two stages described in the following:

1. **Encoder Bitplane Rate Allocation** - The encoder estimates the initial number of rate chunks by bitplane and sends it to the decoder before any request is made. In this case, the decoding process, that before would always start with a single rate chunk, can now start from the encoder estimated number of chunks; this avoids wasting decoding resources by asking the turbo decoder to run for a rate whose probability of being sufficient is extremely low.
2. **Decoder Reaction** - If the encoder estimated rate at the bitplane level does not allow the turbo decoder to reach the predefined bitplane error probability (encoder rate underestimation case), the decoder uses the feedback channel to ask more parity bits from the encoder. Otherwise, the decoder runs only once for this bitplane; in this case, encoder rate overestimation may have happened.

The number of DCT bands and the number of bitplanes for each DCT band to be coded is determined by the quantization matrices, labelled as Q_{index} with index = 1, …8, which set the target decoder quality [2].

3. 1 HRC Solution 1: FNC based HRC

The first hybrid rate control solution proposed works as follows:

1. For each WZ frame to be coded, there are three matrices available, FNC_{-1ij}, FNC_{-2ij} and FNC_{-3ij}, which contain the *final number of rate chunks* (FNC) sent for each bitplane of each DCT band for the last three WZ frames coded (-1, -2 and -3). The final number of rate chunks corresponds to the INC estimated by the encoder for those frames added to the additional number of rate chunks asked after by the decoder to reach the predefined bitplane error probability. For the first three frames, and whenever the necessary information is not available, e.g. for the second frame there is only one past frame, the FNC matrices are initialized with the data available for the previous matrix. For the first frame, the estimated initial number of rate chunks is simply 1.
2. For each DCT band and for each bitplane of the next WZ frame to be coded, the initial number of rate chunks is estimated as

$$INC_{ij} = \left\lfloor (1-k) \times med\,(FNC_{-1ij},\ FNC_{-2ij},\ FNC_{-3ij}) \right\rfloor$$

where INC_{ij} is the estimated initial number of chunks (INC) for the bitplane j of the DCT band i for the WZ frame to be coded, *med*(.) is the median operator, and k is a scale factor; the median operator was used due to its capability to remove noise (in this case, FNC outliers at the bitplane level). The floor function, $\lfloor . \rfloor$, is used in this equation since it is better to underestimate than to overestimate the number of rate chunks from a RD performance point of view.

The equation above tries to estimate INC_{ij} learning from the past FNC assuming some stationarity along time; however, a subtractive term is needed so that the encoder rate estimation algorithm may reduce the estimated initial number of chunks without getting saturated in unnecessary higher rate estimations which would happen without the subtractive term. The *k scale factor* imposes this necessary 'rate reduction effect' which has a (unpleasant) cost associated to the underestimations made even when the number of chunks is very stable and thus no rate reduction should happen. While increasing k would allow following more quickly strong rate decreases, it would also imply strong rate underestimations for other cases and, thus, a lower reduction of the decoding complexity. On the other hand, smaller k values may imply strong rate overestimation for same cases and, thus, a negative impact on the RD performance which should ideally be as close as possible to the RD performance corresponding to the pure decoder rate control case. As it will be shown in the next section, the value of k has been obtained after studying the *INC* evolution for each DCT band and bitplane.

This proposed encoder bitplane rate estimation process is very simple, implying that the added encoding complexity is negligible; this is an important factor in WZ video coding.

3.2 HRC Solution 2: Selective FNC based HRC

One of the drawbacks of HRC Solution 1 previously presented is that it always includes the *k scale factor* to avoid the *INC* saturation. However, this has a constant decoding complexity cost for rather simple situations such as those where *FNC* is constant along many frames. A possible way to overcome this effect is by considering, in the HRC solution, information about what happened in the previous frame for the same bitplane of the same DCT band, this means underestimation (decoder requests had to be made) or not (no decoder requests were made). So, HRC Solution 2 works as follows:

1. If underestimation happened for the same bitplane of the same DCT band in the previous frame (there were decoder requests), then it is assumed that rate saturation is not happening and thus

$$INC_{ij} = \left\lfloor med\,(FNC_{-1ij}, FNC_{-2ij}, FNC_{-3ij}) \right\rfloor$$

this means there is no need to include the k factor.

2. However, if no underestimation happened, it is assumed that overestimation may have happened and then *INC* is computed as for HRC Solution 1.

The next section will evaluate and compare the performance of the two proposed hybrid rate control algorithms.

4. PERFORMANCE EVALUATION

This section presents the test conditions and performance results for the proposed hybrid rate control solutions.

4.1 Test Conditions

To evaluate the performance of the proposed hybrid rate control solutions, four video sequences have been used, notably Coast Guard, Foreman, Hall Monitor and Soccer. All these sequences have 150 frames, except Hall Monitor that has 165 frames; all sequences were coded with QCIF spatial resolution, a frame rate of 15 Hz, and a GOP size of 2 frames.

For these experiments, eight rate-distortion points have been considered as defined in [2], Q_{index} with index = 1, …8; while Q_1 corresponds to the lowest quality and rate, Q_8 corresponds to the highest quality and rate. Within each 4×4 quantization matrix, the value for each DCT band indicates the number of quantization levels associated to that DCT band; the value 0 means that no WZ bits are transmitted for the DCT band.

To analyse the decoding complexity, two different measures were used: i) the number of rate chunks needed for each bitplane of each DCT band; and ii) the decoding runtime. To obtain the decoding runtime, each sequence was encoded and decoded separately. The computer used has an Intel Core 2 Duo CPU 6400 @ 2.13GHz with 2 GB of RAM.

4.2 Number of Requests per DCT Band

Figure 3 shows, for HRC Solution 1, the average (along time) number of chunks per DCT band for Q_4 and Q_8, with the *k* value mentioned below, for the Hall Monitor sequence, for three situations: i) DRC number of chunks (ideal numbers for HRC); ii) INC (with HRC); and iii) FNC (with HRC). The chart shows that, on average, there is typically initial rate underestimation for this rather stable sequence due to the *k* factor. While there is no negative impact on the RD performance, the decoding complexity reduction may not be as high as expected due to the *k* factor. For other sequences (not shown due to space limitations), notably high motion sequences like Soccer, there is some rate overestimation and, thus, a small negative impact on the RD performance since more bits are used for the same final quality.

To ensure that the number of rate overestimations is not too high, the *k* scale factor has been adjusted in order to obtain the best possible underestimation versus overestimation trade-off for all sequences. Since the sequences have different types of motion, the results vary and a compromise had to be found. The trade-off found was setting *k* equal to 0.1 for the first five DCT bands and to 0.05 for the remaining DCT bands. Notice that the first DCT bands use a higher number of rate chunks per bitplane because the quantization steps are lower and, thus, may need a stronger 'reduction effect' in the *INC* estimation.

4.3 RD and Decoding Complexity Performances

In this section, the RD and decoding complexity performances are studied for all tested sequences; Figure 4 and Figure 5 show the corresponding charts for the Hall Monitor (low motion) and Soccer (high motion) sequences for HRC Solution 1.

In general, the results obtained show that the decoding complexity can be significantly reduced without significantly affecting the RD Performance, especially for sequences with low motion. For high motion sequences, there is a small RD performance reduction which is, however, largely compensated by the significant decoding complexity reduction. This RD performance reduction and decoding complexity reduction trade-off may be moved by choosing a different *k* value; a higher *k* value would reduce the decoding complexity gains but would also reduce the RD performance losses.

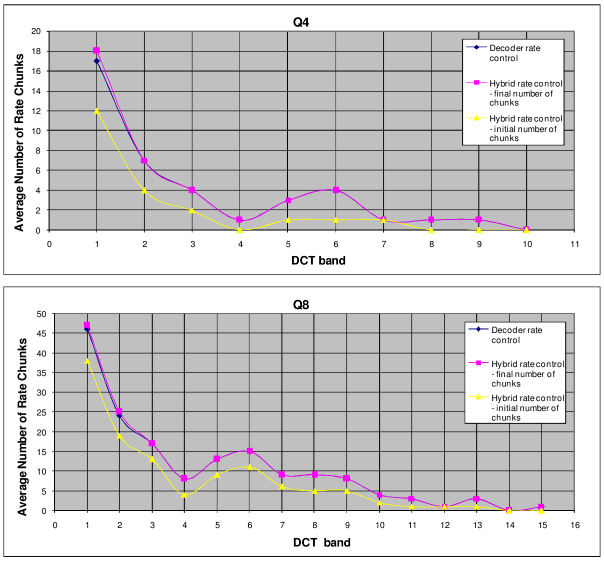

Figure 3 – Average number of chunks per DCT band for the Hall Monitor sequence (HRC Solution 1).

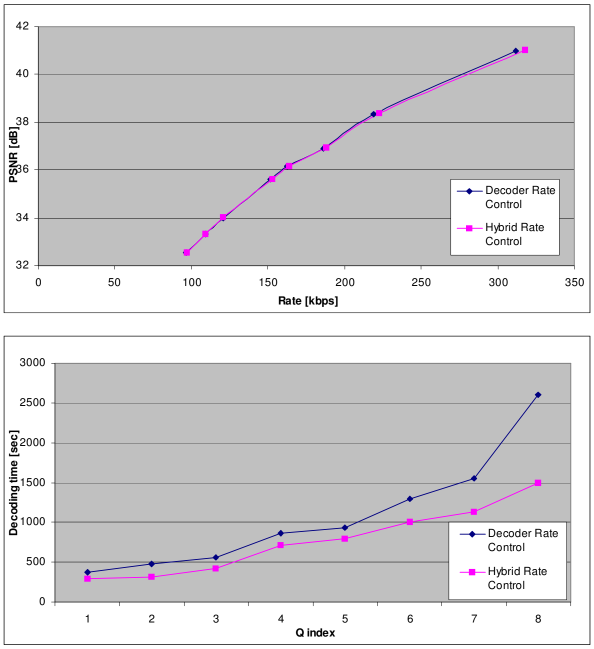

Figure 4 – RD performance and decoding complexity for the Hall Monitor sequence (HRC Solution 1).

For HRC Solution 2, the decoding complexity is even lower for the low motion sequences, like Hall Monitor, without RD performance reduction. However, for high motion sequences like Soccer, the increased decoding complexity reduction comes at an even higher loss in RD performance since the removal of the

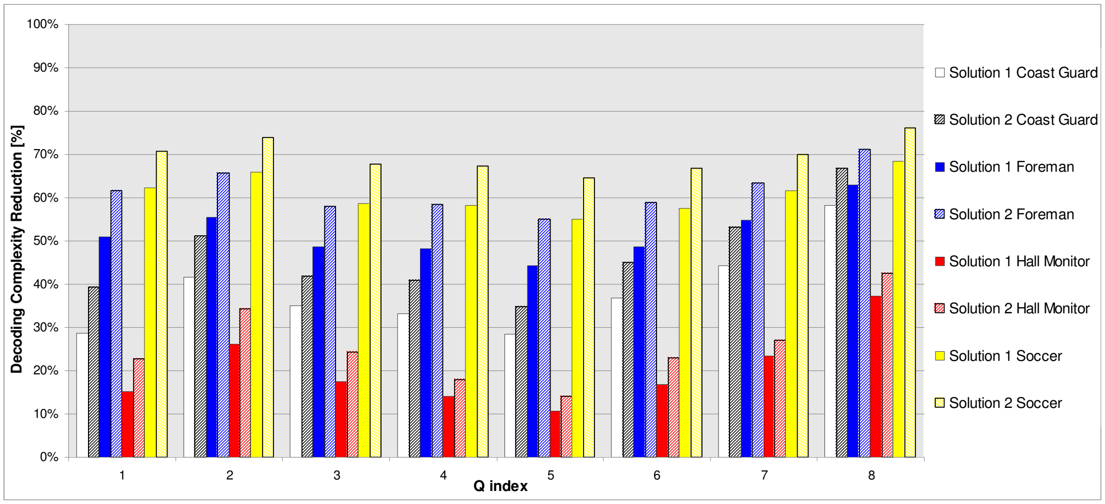

Figure 6 – Decoding complexity reduction (%) for all tested sequences and RD points.

k factor creates much overestimation and, thus, more rate is unnecessarily spent for the same quality.

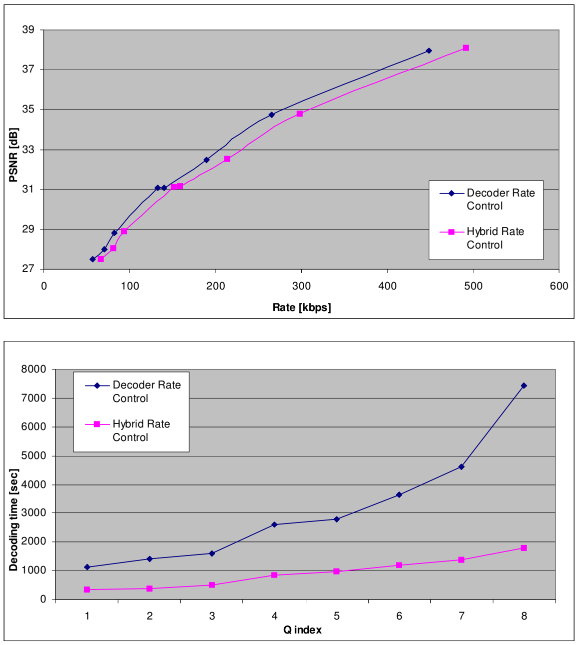

Figure 5 – RD performance and decoding complexity for the Soccer sequence (HRC Solution 1).

Figure 6 shows the decoding complexity (DC) reduction in percentage when moving from DRC to HRC; this means (DC_{DRC} – DC_{HRC} / DC_{DRC})×100 % for all tested sequences and RD points, for the HRC Solutions 1 and 2. Figure 6 shows that decoding complexity reductions up to around 70 % may be reached for the more complex sequences (those more in need since they have higher decoding complexities), almost independently of the RD point. The complexity reductions are lower for the more quiet sequences but reductions are still interesting, notably considering the extremely low added encoding complexity of the novel HRC algorithm.

Figure 6 also shows that the decoding complexity reduction is always higher for HRC Solution 2. However, as said before, while for the low motion sequences there is no negative RD performance impact, for the high motion sequences, the RD performance loss starts to be relevant. This fact may, finally, lead to two alternative HRC solutions, notably: i) **HRC Solution 1** is selected since it is the one always bringing decoding complexity reductions without relevant RD performance losses; or ii) a **HRC Combined Solution** is defined where HRC Solution 1 is used for high motion frames and HRC Solution 2 is used for low motion frames; the HRC Combined Solution would require adding an encoder frame analysis module to classify the frames as low or high motion, further increasing the encoding complexity to further decrease the decoding complexity.

5. FINAL REMARKS

This paper proposes HRC solutions for WZ video coding with the objective to reduce the huge decoding complexity at the cost of additional negligible encoding complexity. Although, applications may be ready to pay with additional decoding complexity the WZ reduction in encoding complexity, this trade-off must be reasonable which may not happen without the adoption of tools as those proposed here.

REFERENCES

[1] B. Girod, A. Aaron, S. Rane and D. Rebollo Monedero, “Distributed Video Coding”, Proceedings of the IEEE, vol. 93, nº 1, pp. 71 - 83, Jan. 2005.

[2] C. Brites, J. Ascenso, J. Pedro, F. Pereira, “Evaluating a Feedback Channel based Transform Domain Wyner-Ziv Video Codec”, accepted for publication in Signal Processing: Image Communication.

[3] C. Brites, F. Pereira, “Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding”, Int. Conf. on Image Processing (ICIP’2007), San Antonio, TX, USA, Sep. 2007.

[4] D. Kubasov and C. Guillemot, “A Hybrid Encoder/Decoder Rate Control for Wyner-Ziv Video Coding with a Feedback Channel”, IEEE International Workshop on Multimedia Signal Processing, Crete, Greece, Oct. 2007.