

# LOW COMPLEXITY INTRA MODE SELECTION FOR EFFICIENT DISTRIBUTED VIDEO CODING

João Ascenso<sup>1</sup>, Fernando Pereira<sup>2</sup>

<sup>1</sup>joao.ascenso@lx.it.pt, <sup>2</sup>fp@lx.it.pt

<sup>1</sup>Instituto Superior de Engenharia de Lisboa, <sup>2</sup>Instituto Superior Técnico, <sup>1,2</sup>Instituto de Telecomunicações

## ABSTRACT

Motion compensated frame interpolation (MCFI) is one of the most efficient solutions to generate side information (SI) in the context of distributed video coding. However, it creates SI with rather significant motion compensated errors for some frame regions while rather small for some other regions depending on the video content. In this paper, a low complexity Intra mode selection algorithm is proposed to select the most ‘critical’ blocks in the WZ frame and help the decoder with some reliable data for those blocks. For each block, the novel coding mode selection algorithm estimates the encoding rate for the Intra based and WZ coding modes and determines the best coding mode while maintaining a low encoder complexity. The proposed solution is evaluated in terms of rate-distortion performance with improvements up to 1.2 dB regarding a WZ coding mode only solution.

**Index Terms**—Distributed video coding, Wyner-Ziv coding, side information, coding mode selection, Intra coding

## 1. INTRODUCTION

Distributed video coding (DVC) is a coding paradigm exploiting the source correlation partially or totally at the decoder which theoretically can achieve no rate distortion (RD) performance loss when compared to the predictive coding approach [1], under certain conditions (Gaussian, memoryless sources and assuming a mean-squared error distortion measure). The video codecs exploiting this principle are called distributed video codecs; typically, they generate some side information (SI) at the decoder (a WZ frame estimate) by using motion estimation (ME) and compensation techniques. This typically leads to video codecs with two important characteristics: low encoding complexity and inbuilt robustness to channel losses due to the lack of a prediction loop.

Several Intra coding mode decision algorithms have been presented in the DVC literature in which the encoder selects the most critical blocks to be Intra coded. The Intra blocks are sent for low correlation regions between the SI and the corresponding WZ source where distributed coding is inefficient. In [2], the Intra mode selection depends on the sum of absolute differences (SAD) between each WZ block and the co-located block in the previous frame. In [3], both spatial and temporal block characteristics are considered by including the pixel variance of each block and the SAD (as defined in [2]), respectively. In [4], a DVC codec with three encoding modes is presented: Intra, Inter and WZ. At the encoder, a bitplane ME algorithm is performed and the ME residual error is used to select the coding mode for each block. Although [4] shows coding efficiency improvements, the DVC codec is now prone to error propagation and has significant higher encoding complexity. In all previous works, a WZ frame block is classified into one of two modes: WZ or Intra and the Intra blocks are used directly as the final decoded information since no SI is generated/used for them. In this paper, a novel approach is proposed where the Intra blocks are used as reference for the decoder to create SI and are further distributed coded; this coding

mode is called Intra/WZ since it is a mix of Intra and DVC coding. Thus, the Intra blocks in WZ frames are encoded with lower quality compared to the final decoded quality allowing some bitrate savings and are only sent for the regions where the motion compensated frame interpolation (MCFI) SI creation approach is estimated to fail. At the decoder, a higher SI quality is achieved for the Intra/WZ mode blocks when compared to the WZ coding mode blocks since Intra coded data is available to help the SI creation; higher SI quality leads to RD performance improvements. In this context, the main challenge to obtain a more efficient DVC codec, still with low encoding complexity, is the design of a mechanism for the encoder to select the blocks in the WZ frames to be Intra coded; this selection should be made without having available the MCFI SI and the decoded Intra coded blocks. In this paper, a novel block based coding mode selection algorithm is proposed which is able to efficiently determine for which blocks in the WZ frames it is worthwhile to send Intra coded data; the Intra coding mode must lead to significant WZ rate savings (considering the same target quality). This work also departs from [2-4] since it explicitly predicts the bitrate spent for each coding mode to obtain a more accurate mode selection.

This paper is organized as follows: in Section 2, a brief overview of the DVC codec is presented; in Section 3, the novel coding mode selection criterion is proposed and in Section 4 the RD performance is evaluated. Finally, in Section 5, some final remarks are drawn.

## 2. DVC CODEC ARCHITECTURE

The video codec adopted in this paper is a feedback channel based transform domain DVC codec [1]. Fig. 1 illustrates the architecture of the DVC encoder with two block-based coding modes: the usual WZ mode and the novel Intra/WZ mode. The transform domain DVC (TD-DVC) encoder includes a forward DCT transform, a uniform quantizer and a turbo encoder, the typical components of this coding architecture. At the encoder, the first step regards the definition of the group of pictures (GOP); each GOP has always one key frame while the remaining are WZ frames. The key frames are encoded using the H.264/AVC Intra coding scheme and for each WZ frame  $X_i$ , the encoder defines two frames,  $X_b$  (temporally in the past) and  $X_f$  (temporally in the future), which are used as references in the encoder mode selection algorithm and the SI generation technique. The novelty proposed here corresponds to the coding mode selection algorithm which classifies each block in the WZ frames in one of two coding modes: WZ or Intra/WZ. The coding mode decisions are concatenated in a binary decision map which is efficiently encoded with a Run Length Encoding (RLE) algorithm and Universal Length Variable Codes (UVLC). The Intra blocks (Intra/WZ mode) are encoded with H.264/AVC tools and have lower quality, this means higher quantization steps, when compared to the key frames blocks. All reconstructed Intra blocks (Intra/WZ mode) are subtracted from the corresponding blocks in the WZ frame  $X_i$  while for the remaining blocks (WZ mode) 128 (central luminance value) is subtracted to

$X_i$ ; this guarantees that the residual distribution is centered on the origin and has lower energy, improving the coding efficiency.

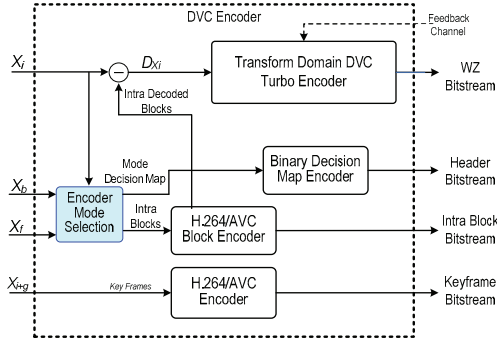


Figure 1 – Architecture of the DVC encoder.

Fig. 2 shows the proposed DVC decoder architecture with two SI modes, one for each coding mode (Intra/WZ and WZ). The TD-DVC decoder includes the turbo decoder, the reconstruction and the Inverse DCT modules. First, the motion compensated frame interpolation (MCFI) module generates an estimate of  $X_i$ , based on the decoded reference frames  $X'_b$  and  $X'_f$ . However, for this DVC decoder architecture, it is also necessary to create SI based on the Intra decoded blocks in WZ frames (Intra/WZ mode). Thus, the Intra blocks in the WZ frames and the binary decision map are decoded and used by the motion compensated quality enhancement (MCQE) module along with the MCFI interpolation. The MCQE is based on motion compensated temporal filtering, an averaging process applied across video frames to separate the quantization noise from the relevant content over time and is more efficient compared to MCFI, since it uses the Intra coded blocks for motion estimation (besides  $X_b$  and  $X_f$ ).

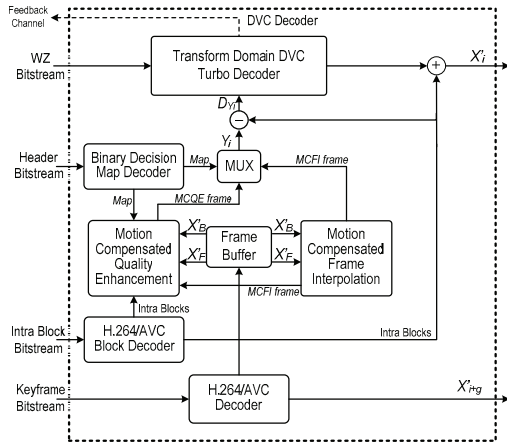


Figure 2 – Architecture of the DVC decoder.

The SI  $Y_i$  is constructed by multiplexing the MCFI estimated blocks (WZ mode) with the MCQE estimated SI blocks (Intra/WZ mode). Then, the corresponding residual frame  $D_{Y_i}$  is obtained by subtracting all decoded Intra blocks from the corresponding  $Y_i$  blocks; for the remaining blocks (WZ mode), 128 is subtracted instead. Then, all the blocks are decoded with side information in the TD-DVC decoder independently of their coding mode. To obtain the final decoded WZ frame the Intra decoded blocks are added to the TD-DVC decoder output; in case no Intra block data is received, 128 is added.

### 3. INTRA/WZ VERSUS WZ MODE SELECTION

The SI quality has a major impact on the RD performance of a

DVC codec. To perfectly estimate the rate and distortion for the block based coding mode selection is necessary to have the SI generated for each mode (Intra/WZ and WZ). However, since the encoder computational capabilities are limited, the MCFI/MCQE framework cannot be replicated at the encoder to recreate the SI and, therefore, an efficient but low complexity coding mode selection criterion is needed. The proposed criterion exploits two rate models, both working at the block level: the first model estimates the Intra/WZ coding mode bitrate  $R^{IW}$  and the second model estimates the WZ coding mode rate  $R^{WZ}$ ; then, the coding mode with the minimum rate is chosen. It is here assumed that the distortion/quality achieved for both coding modes is similar since parity bits are sent for the same bands/bitplanes and the reconstruction function is the same.

#### 3.1. Intra/WZ coding mode rate estimation

The Intra/WZ coding mode  $R^{IW}$  has two terms: the Intra bitrate  $R^I$  and the WZ bitrate  $R^W$ . However, it can be assumed that  $R^I \gg R^W$  since the MCQE SI quality is typically much higher than MCFI SI quality and only the residual between the Intra blocks and the corresponding WZ block is encoded (less energy implies less rate). Thus,  $R^{IW} \cong R^I$  and it is only proposed a rate model for  $R^I$  that receives as input the DCT unquantized frame coefficients and predicts the rate spent by each Intra block.

To limit the encoding complexity, the model proposed here avoids the actual Intra coding process for the WZ frame blocks. The Intra coded blocks are encoded with efficient H.264/AVC Intra coding techniques, notably 4×4 DCT integer transform, 4×4 DC Intra prediction mode, uniform quantizer and the Context Adaptive Variable Length Codes (CA-VLC). Although these tools do not correspond to the highest H.264/AVC Intra coding efficiency (e.g. not all Intra prediction modes and CABAC are used), they are chosen in order to limit the encoder complexity. In [5], a popular RD model has been proposed: the so-called  $\rho$  domain RD model where  $\rho$  is the fraction of zeros among the DCT quantized coefficients in a block. It has been shown that the bitrate and  $\rho$  exhibit the linear relationship  $R = \theta(1 - \rho)$ , where  $\theta$  is the model parameter. Although this model works well to estimate the quantization parameter at frame level [5] it fails when applied to small block sizes such as the 4×4 DCT transform. Therefore, a new rate model is proposed here which extends the  $\rho$  domain model in [5] by including the DC coefficient value and the quantization step size  $Q$  for each block:

$$R^{IW} = \theta(1 - \rho) + \kappa \left( \frac{DC}{Q} \right) \quad (1)$$

where  $\theta$  and  $\kappa$  are the model parameters and  $Q$  can be expressed as a function of the quantization parameter  $QP$  in the following way [6]:

$$Q = 2^{\frac{QP-4}{6}}. \quad (2)$$

In (1),  $DC/Q$  is a measure of the energy of the quantized transform coefficients and provides a simple and efficient way to obtain a more accurate bitrate estimation for H.264/AVC Intra coding. The parameters  $\theta$  and  $\kappa$  can be calculated using a linear regression model with previous Intra encoded blocks data, since it is known how much bitrate was spent and the corresponding  $(1 - \rho)$  and  $(DC/Q)$  values. This process enables the adaptation of the rate model parameters to the spatial (more or less textured) and temporal (more or less motion) video sequence characteristics. Thus, considering  $n$  past encoded blocks,  $E_i = (DC/Q)$ ,  $Z_i = (1 - \rho)$  and  $R_i^{IW}$  the bitrate spent by encoding block  $i$  with H.264/AVC Intra tools, it comes:

$$\theta = \frac{\sum_{i=1}^n E_i^2 \sum_{i=1}^n Z_i R_i^{IW} - \sum_{i=1}^n Z_i E_i \sum_{i=1}^n R_i^{IW} E_i}{\sum_{i=1}^n E_i^2 \sum_{i=1}^n Z_i^2 - (\sum_{i=1}^n Z_i E_i)^2} \quad (3)$$

$$\kappa = \frac{\sum_{i=1}^n Z_i^2 \sum_{i=1}^n R_i^{IW} E_i - \sum_{i=1}^n Z_i E_i \sum_{i=1}^n Z_i R_i^{IW} E_i}{\sum_{i=1}^n E_i^2 \sum_{i=1}^n Z_i^2 - (\sum_{i=1}^n Z_i E_i)^2} \quad (4)$$

Although (3) and (4) may look rather complex, in fact only sums and multiplications are performed and most results can be reused more than once; so little extra complexity is introduced by this regression method. In conclusion, the algorithm to estimate the Intra/WZ coding mode bitrate cost is the following:

1. For the first WZ frame, perform H.264/AVC Intra coding for all blocks, collect the corresponding parameters  $E_i$ ,  $Z_i$  and  $R_i$  and calculate the parameters  $\theta$  and  $\kappa$  using (3) and (4). Although all blocks are Intra coded to accurately know their bitrate, the coding mode selection will select the most advantageous coding mode, i.e.  $\min \{R_i^{IW}, R_i^{WZ}\}$ .
2. For the next WZ frames, the bitrate  $R_i^{IW}$  for a block  $i$  is calculated with (1) and (2) and the coding mode selection determines the blocks to be Intra/WZ coded.
3. When the number of Intra/WZ coded blocks in one or more past WZ frames is more than 50% of the total number of blocks in a frame (e.g. in QCIF this corresponds to  $1584/2 = 792$  blocks),  $\theta$  and  $\kappa$  are recalculated with (3) and (4). This avoids the continuous updating of the model parameters after coding each WZ frame since in some cases there is not enough data to obtain reliable parameters with the linear regression of (3) and (4).

### 3.2. WZ coding mode rate estimation

To perform the coding mode selection, it was developed a model to estimate the WZ rate spent for each block in the WZ frames for which the MCFI SI is used at the decoder. Consider a  $4 \times 4$  residual block  $N = X - Y$  in the pixel domain with  $X$  corresponding to the block in the original frame and  $Y$  the corresponding block in the SI. Then, a  $4 \times 4$  DCT integer transform is applied to  $N$  obtaining  $R = C_f N C_f^T \otimes E_f$ , where  $C_f$  is the H.264/AVC forward transform,  $E_f$  is the post scaling factor matrix and  $\otimes$  is the point to point multiplication. The statistical correlation between  $X$  and  $Y$  DCT coefficients is often modeled by a Laplacian distribution model [1]. When the SAD criterion is used to measure the distortion  $D$ , the Laplacian distribution source RD function is [7]:

$$R(D) = \begin{cases} \ln\left(\frac{1}{\alpha D}\right) & 0 < D < 1/\alpha \\ 0 & D \geq 1/\alpha \end{cases} \quad (5)$$

where  $\alpha$  is the Laplacian distribution parameter. Assuming a uniform distribution over the quantizer bins, the distortion  $D$  can be written as:

$$D = \sum_{k=0}^{N-1} \int_{u_k}^{u_{k+1}} |u - v_k| \frac{1}{\Delta_k} du = \frac{\Delta}{4} \quad (6)$$

where  $u_k$  and  $u_{k+1}$  are the limits of the quantizer bin  $k$ ,  $N$  is the number of quantizer bins and  $\Delta$  is the quantization step. The Laplacian distribution parameter  $\alpha(u, v)$  can be estimated for each DCT coefficient  $(u, v)$  as:

$$\alpha(u, v) = \sqrt{\frac{2}{\sigma_R^2(u, v)}} \quad (7)$$

where  $\sigma_R^2(u, v)$  is the variance of the residual DCT coefficient;  $\sigma_R^2(u, v)$  can be written as in [8]:

$$\sigma_R^2(u, v) = \sigma_N^2(u, v) K(u, v) \quad (8)$$

$$K(u, v) = [C_f R_x C_f^T]_{u,u} [C_f R_x C_f^T]_{v,v} E_f^2(u, v) \quad (9)$$

where  $\sigma_N^2$  is the variance of the residual block (pixel domain),  $R_x$  corresponds to the correlation matrix (as defined in [8]) and  $K$  is a matrix which defines the linear relationship between  $\sigma_R^2(u, v)$  and  $\sigma_N^2(u, v)$ . Moreover, it was shown in [8] that  $\sigma_N(u, v) \approx \sqrt{2} MAD$ ; thus, substituting (8) in (7) it comes:

$$\alpha(u, v) = \sqrt{\frac{1}{K(u, v) MAD_i^2}} \quad (10)$$

where  $MAD_i$  represents the mean absolute difference (MAD) between the SI  $Y$  and the source  $X$  for a luminance block  $i$ :

$$MAD_i = \frac{1}{A^2} \sum_{(i,j) \in B_i} |X(i, j) - Y(i, j)| \quad (11)$$

where  $A$  corresponds to the size of the block and  $B_i$  is a block in  $X$ . Using (6) to estimate the distortion, (10) to estimate the  $\alpha$  Laplacian distribution parameter and the rate model in (5) the novel WZ rate model comes:

$$R_i^{WZ}(u, v) = \begin{cases} \ln\left(\gamma \frac{4\sqrt{K(u, v)} MAD_i}{\Delta}\right) & 0 < \Delta < 4/\alpha \\ 0 & \Delta \geq 4/\alpha \end{cases} \quad (12)$$

In (12), the model parameter  $\gamma$  is included to compensate for any mismatch in the assumptions made in the construction of the rate model. To estimate the rate spent by a block using the WZ mode it is only necessary to calculate  $R_i^{WZ} = \sum_{(u,v) \in B_i} R_i(u, v)$ , i.e. to sum the WZ bitrate for the various block coefficients which are WZ encoded using the quantization matrix  $Q$ ;  $Q$  defines for which bands WZ bits are sent and thus it is used to calculate the corresponding  $\Delta$ . To obtain  $MAD_i$ , it is necessary to create an estimate of the SI  $Y$  at the encoder but with lower complexity (thus reducing the estimated SI quality). In this case, the fast motion compensated interpolation described in [9] is used, which has a very low complexity, roughly 3–4 times greater than the average of the two collocated blocks in the backward and forward reference frames. Since the SI at the encoder is quite different (lower quality) from the SI at the decoder, some rate overestimation can occur; the model parameter  $\gamma$  is used to compensate for this lack in the accuracy of the  $Y$  estimation at the encoder (it was experimentally found that  $\gamma = 2/5$  leads to good results).

### 4. RD PERFORMANCE

In order to stimulate comparisons, the test conditions defined by the DISCOVER project [10], which also implemented a state-of-the-art DVC codec, are used here. Due to space limitations, the Foreman and Soccer sequences at QCIF@15Hz were selected since they have very different motion and texture characteristics; results for GOP size 2 and 4 are shown. The accuracy of the coding mode selection algorithm is evaluated; Fig. 3 shows the frame level bitrate evolution for the Foreman sequence with GOP size 2, for the proposed Intra and WZ rate models, respectively. For each chart in Fig. 3, two pair of curves are drawn associated to the RD points with the lowest ( $QI = 1$ ) and highest ( $QI = 8$ ) average bitrate. For each RD point, two curves are plotted: 1) the actual frame bitrate spent to achieve successful decoding, and 2) the estimated bitrate for all the blocks in the WZ frame with the corresponding Intra (Fig. 3 – top) or WZ (Fig. 3 – bottom) model. As shown, the Intra

rate model can predict very accurately the Intra bitrate for low and high bitrate regimes. As expected, the WZ bitrate model has a higher error when compared to the Intra model due to several factors, e.g. simple SI estimation at the encoder, Laplacian model failure, etc. Still the estimated rate can follow rather accurately the changes in the actual bitrate, just with some rate underestimation at the encoder, this leads to a rather conservative Intra block selection, i.e. only when they effectively provide improvements.

The RD performance of the proposed DVC codec (with the architecture of Figs. 1 and 2) with the novel coding mode selection (described in Section 3) is compared with the performance of the same DVC codec without Intra/WZ coding mode; for benchmarking, also the H.264/AVC Intra and DISCOVER DVC codec performance are shown. As shown in Fig. 4 for GOP 4, the DVC codec with the proposed coding mode selection outperforms the DVC codec without it for all sequences and RD points. As expected, the higher the motion content, the higher is the RD performance gain. As the GOP size increases, the quality of the MCFI SI tends to decrease and thus the Intra/WZ mode selection is more effective, obtaining higher compression gains. Thus, the highest performance shown is obtained for the Soccer sequence, where up to 1.2dB gains in RD performance can be observed. The DVC codec with the proposed mode selection also outperforms the DISCOVER codec for all sequences and GOP sizes and the H.264/AVC Intra coding gap is also consistently reduced.

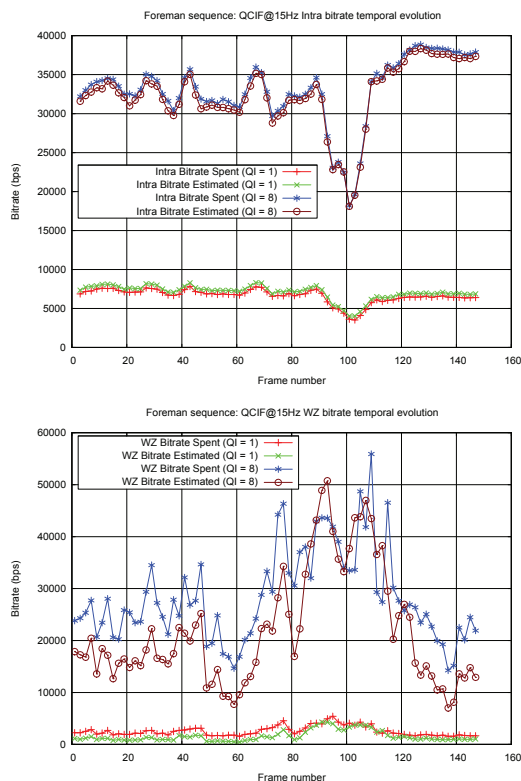


Figure 3 – Intra (top) and WZ (bottom) bitrate temporal evolution.

## 5. CONCLUSIONS

This paper proposes a novel block level coding mode selection algorithm, at the encoder, in the context of a turbo-based transform domain DVC codec. The experimental results show RD gains up to 1.2 dB for critical test conditions where typically available DVC codecs have poorer RD performance: complex video sequences and

longer GOP sizes. As future work, the inclusion of a distortion model for each coding mode could lead to more accurate decisions.

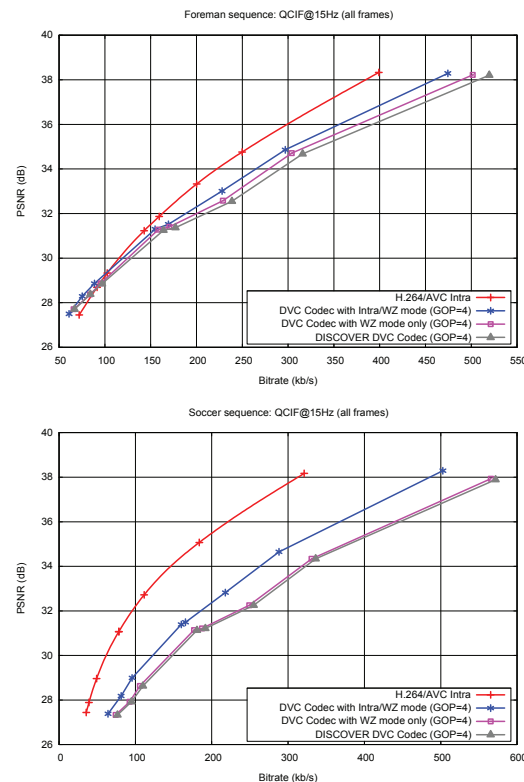


Figure 4 – RD performance for Foreman (top) and Soccer (bottom).

## 6. REFERENCES

- [1] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, J. Ostermann, "Distributed Monoview and Multiview Video Coding", *IEEE Signal Processing Magazine*, vol. 24, n° 5, pp. 67 – 76, September 2007.
- [2] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Embedding a Block-based Intra Mode in Frame-based Pixel Domain Wyner-Ziv Video Coding", *International Workshop on Very Low Bitrate Video Coding*, Sardinia, Italy, September 2005.
- [3] M. Tagliasacchi, A. Trapanese, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Intra Mode Decision Based on Spatio-Temporal Cues In Pixel Domain Wyner-Ziv Video Coding", *IEEE ICASSP*, Toulouse, France, May 2006.
- [4] T. Clerckx, A. Munteanu, J. Cornelis, and P. Schelkens, "Distributed Video Coding with Shared Encoder/Decoder Complexity", *IEEE ICIP*, San Antonio, TX, USA, September 2007.
- [5] Z. He, S. K. Mitra, "A Linear Source Model and a Unified Rate Control Algorithm for DCT Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, n° 11, November 2002.
- [6] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, n° 7, July 2003.
- [7] T. Berger, "Rate Distortion Theory: A Mathematical Basis for Data Compression", Prentice-Hall, 1971, pp. 94-98.
- [8] I.-M. Pao, M.-T. Sun, "Modeling DCT Coefficients for Fast Video Encoding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, n° 4, June 1999.
- [9] C. Brites, F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding", *IEEE ICIP*, San Antonio, TX, USA, September 2007.
- [10] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER Codec: Architecture, Techniques and Evaluation", *Picture Coding Symposium*, Lisbon, Portugal, November 2007.