IMPROVING TRANSFORM DOMAIN WYNER-ZIV VIDEO CODING PERFORMANCE *

Catarina Brites[1], João Ascenso[2], Fernando Pereira[3]

[1]catarina.brites@lx.it.pt, [2]joao.ascenso@lx.it.pt, [3]fp@lx.it.pt
[1,3] Instituto Superior Técnico – Instituto de Telecomunicações
[2]Instituto Superior de Engenharia de Lisboa – Instituto de Telecomunicações

ABSTRACT

Distributed video coding (DVC) is a new video coding paradigm based on two key Information Theory results: the Slepian-Wolf and Wyner-Ziv theorems. A particular case of DVC, the so-called Wyner-Ziv coding, deals with lossy source coding with side information at the decoder and enables a flexible allocation of complexity between the encoder and the decoder. This paper proposes an improved transform domain Wyner-Ziv video codec including: 1) the integer block-based transform defined in the H.264/MPEG-4 AVC standard, 2) a quantizer with a symmetrical interval around zero for AC coefficients, and a quantization step size adjusted to the transform coefficient bands dynamic range, and finally and 3) advanced frame interpolation for side information generation. The combination of these tools brings significant rate-distortion (RD) gains regarding the state-of-the-art results available in the literature.

1. INTRODUCTION

Nowadays, the digital video coding solutions available rely on the powerful hybrid block-based motion compensation/DCT transform (MC/DCT) architecture. All the ITU-T VCEG and ISO/IEC MPEG standards follow this approach, mostly targeting applications where the video content is encoded once and decoded multiple times, e.g. broadcasting or video streaming. In such applications, the video codec architecture is primarily driven by the one-to-many model of a single complex encoder and multiple light (cheap) decoders; typically the encoder is 5 to 10 times more complex than the decoder. The complexity burden of the encoder is mainly associated with the motion estimation and compensation tasks, which account for a major share of the coding gain in rate-distortion (RD) performance.

However, this architecture is being challenged by several emerging applications such as wireless video surveillance, multimedia sensor networks, wireless PC cameras and mobile camera phones. These applications have different requirements from those targeted by traditional video delivery systems. For example, in wireless video surveillance systems, low cost encoders are important since there is a high number of encoders and only one or few decoders.

Distributed video coding, a new video coding paradigm, fits well in these scenarios, since it enables to explore the video statistics, partially or totally, at the decoder only, relying on a low encoding complexity. From the Information Theory, the Slepian-Wolf theorem [1] states that it is possible to compress two statistically dependent signals, *X* and *Y*, in a distributed way (separate encoding, jointly decoding) using a rate similar to that used in a system where the signals are encoded and decoded together, i.e. like in traditional video coding schemes. The complement of Slepian-Wolf coding for lossy compression is Wyner-Ziv coding [2]. A particularly interesting case deals with the source coding of an *X* sequence considering that a dependent sequence *Y*, known as side information, is only available at the decoder. Wyner and Ziv showed that there is no increase in the transmission rate if the statistical dependency between *X* and *Y* is only explored at the decoder compared to the case where it is explored both at the decoder and the encoder (if *X* and *Y* are jointly Gaussian and a mean-square error distortion measure is considered).

The Wyner-Ziv coding paradigm may be applied in the pixel domain or in the transform domain. In [3], the IST-PDWZ (Pixel-Domain Wyner-Ziv) codec is presented; this codec uses a rather advanced side information solution at the decoder using motion compensated frame interpolation. The RD performance of this scheme can be further improved by using a transform coding tool with the same purpose as in traditional video coding, i.e. to exploit spatial correlation between neighboring sample values and to compact the block energy into as few transform coefficients as possible.

This paper is organized as follows: Section 2 presents a brief summary of the IST-PDWZ codec, which constitutes the starting point for this paper. This summary is necessary to introduce, in Section 3, the IST-Transform Domain Wyner-Ziv (IST-TDWZ) video codec which constitutes the novel element of this paper. In Section 4, several experimental results are presented to evaluate the IST-TDWZ RD performance regarding the best solutions in the literature. Conclusions and some future work topics are presented in Section 5.

2. THE IST-PIXEL DOMAIN WYNER-ZIV (IST-PDWZ) VIDEO CODEC

The IST-PDWZ codec [3] is an improved PDWZ video coding solution which follows the same architecture as the one proposed by Aaron *et al.* in [4]. There are however some major differences regarding the coding solution proposed in [4], notably in the Slepian-Wolf codec and the frame interpolation module.

In a nutshell, the overall coding process is as follows: the video frames are organized into key frames and Wyner-Ziv frames, the odd and the even frames of the video sequence respectively; to start with, the key frames are assumed to be perfectly reconstructed at the decoder and the frames in between them are Wyner-Ziv encoded.

* The work presented was developed within VISNET, a Network of Excellence (http://www.visnet-noe.org), and DISCOVER, a Future Emerging Technology project (http://www.discoverdvc.org/) both funded by the European Commission.

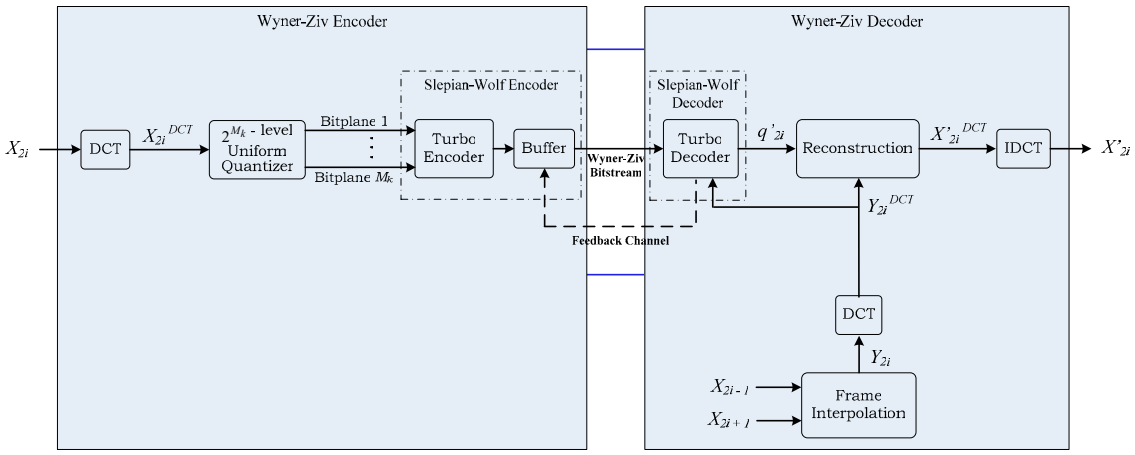

Figure 1 – Transform domain Wyner-Ziv codec architecture.

Each Wyner-Ziv frame of a video sequence, X_{2i}, is encoded sample by sample, i.e. pixel by pixel. The X_{2i} frame pixels are quantized using a 2^M-level uniform scalar quantizer, generating the quantized symbol stream. Over the resulting quantized symbol stream (constituted by all the quantized symbols of X_{2i} using 2^M levels) bitplane extraction is performed and each bitplane is then independently turbo encoded. The turbo encoder encloses two recursive systematic convolutional (RSC) encoders of rate ½ and a pseudo-random L-bit interleaver. Each RSC encoder outputs the parity stream and the systematic stream. After turbo encoding a bitplane, the systematic sequences are discarded and the parity sequences are stored in the buffer and transmitted in small amounts upon decoder request via the feedback channel. At the decoder, the frame interpolation module is used to generate the side information Y_{2i}, an estimate of the X_{2i} frame, based on two temporally adjacent key frames of X_{2i}; this estimate is then used by an iterative turbo decoder to obtain the decoded quantized symbol stream. The turbo decoder is constituted by two soft-input soft-output (SISO) decoders; each SISO decoder is implemented using the Maximum *A Posteriori* (MAP) algorithm to estimate the *a posteriori* probabilities of the data bits corresponding to the parity bits at the SISO decoders input. The iterative MAP algorithm used at the turbo decoder employs a Laplacian distribution to model the residual distribution between Y_{2i} and X_{2i}. The Laplacian distribution parameter is determined in an offline training phase, by constructing the residual histogram, between Y_{2i} and the X_{2i}. It is assumed the decoder has ideal error detection capabilities, i.e. the turbo decoder is able to measure in a perfect way the current bitplane error probability P_e. For example, if $P_e > 10^{-3}$, the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the bitplane turbo decoding task is considered successful. The side information is also used in the reconstruction module, together with the decoded quantized symbol stream, to help in the X_{2i} reconstruction task.

3. THE IST-TRANSFORM DOMAIN WYNER-ZIV (IST-TDWZ) VIDEO CODEC

The IST-TDWZ codec here presented uses as starting point the IST-PDWZ explained in the previous Section, i.e. reuses some of the pixel domain tools whenever this is adequate. Figure 1 illustrates IST-TDWZ solution whose architecture is similar to the one proposed by Aaron *et al.* in [5]. There are however significant differences between the IST-TDWZ solution proposed in this paper and the one proposed in [5], namely in the frame interpolation module, the Slepian-Wolf codec, the DCT and the quantizer.

3.1 IST-TDWZ Architecture

The overall coding architecture illustrated in Figure 1 works as follows: a video sequence is divided into Wyner-Ziv frames and key frames as in the IST-PDWZ solution. Over each Wyner-Ziv frame X_{2i}, it is applied a block-based discrete cosine transform (DCT). The DCT coefficients of the entire frame X_{2i} are then grouped together, according to the position occupied by each DCT coefficient within the 4×4 blocks, forming the DCT coefficients bands. After the transform coding operation, each DCT coefficients band b_k is uniformly quantized with 2^{M_k} levels (where the number of levels 2^{M_k} depends on the DCT coefficients band b_k). Over the resulting quantized symbol stream (associated to the DCT coefficients band b_k), bitplane extraction is performed. For a given band, the quantized symbols bits of the same significance (e.g. the most significant bit) are grouped together forming the corresponding bitplane array which is then independently turbo encoded.

The turbo coding procedure for the DCT coefficients band b_k starts with the most significant bitplane array, which corresponds to the most significant bits of the b_k band quantized symbols. The parity information generated by the turbo encoder for each bitplane is then stored in the buffer and sent in chunks upon decoder request through the feedback channel. The decoder performs frame interpolation using the previous and next temporally adjacent frames of X_{2i} to generate an estimate of frame X_{2i}, Y_{2i}. A block-based 4×4 DCT is then carried out over the Y_{2i} in order to obtain Y_{2i}^{DCT}, an estimate of X_{2i}^{DCT}. The residual statistics between correspondent coefficients in X_{2i}^{DCT} and Y_{2i}^{DCT} is assumed to be modelled by a Laplacian distribution; the Laplacian parameter is estimated offline for the entire sequence at the DCT band level, i.e. each DCT band has a Laplacian parameter associated. Once Y_{2i}^{DCT} and the residual statistics for a given DCT coefficients band b_k are known, the decoded quantized symbol stream q'_{2i} associated to the DCT band b_k can be obtained through an iterative turbo decoding procedure, as in IST-PDWZ solution. An ideal error detection capability is also assumed at the decoder to determine the current bitplane error probability of a given DCT band. After successfully

turbo decoding the most significant bitplane array of the b_k band, the turbo decoder proceeds in an analogous way to the remaining M_{k-1} bitplanes associated to that band. Once all the bitplane arrays of the DCT coefficients band b_k are successfully turbo decoded the turbo decoder starts decoding the b_{k+1} band. This procedure is repeated until all the DCT coefficients bands for which Wyner-Ziv bits are transmitted are turbo decoded.

After turbo decoding the M_k bitplanes associated to the DCT band b_k, the bitplanes are grouped together to form the decoded quantized symbol stream associated to the b_k band; this procedure is performed over all the DCT coefficients bands to which Wyner-Ziv bits are transmitted. Once all the decoded quantized symbol streams are obtained, it is possible to reconstruct the matrix of DCT coefficients, $X'_{2i}{}^{DCT}$. For some DCT coefficients bands, no Wyner-Ziv bits are transmitted; at the decoder, those DCT coefficients bands are replaced by the corresponding DCT bands of the side information, $Y_{2i}{}^{DCT}$. The remaining DCT bands are obtained using a reconstruction function which bounds the error between DCT coefficients of $X_{2i}{}^{DCT}$ and $X'_{2i}{}^{DCT}$ (also known as reconstruction distortion) to the quantizer coarseness. After all DCT coefficients bands are reconstructed, a block-based 4×4 Inverse Discrete Cosine Transform (IDCT) is performed and the reconstructed X_{2i} frame, X'_{2i}, is obtained. In the following, the novelties regarding the transform and quantization tasks in the IST-TDWZ solution are described in detail. In the next sections, the novel elements in the IST-TDWZ codec will be presented in detail.

3.1. The IST-TDWZ Transform

In the IST-TDWZ architecture, illustrated in Figure 1, the first stage towards encoding a Wyner-Ziv frame X_{2i} is transform coding (represented by the DCT module). The transform employed in the IST-TDWZ solution relies on the integer 4×4 block-based discrete cosine transform, as defined by the H.264/MPEG-4 AVC standard [6]; the DCT transform is applied to all 4×4 non-overlapping blocks of the X_{2i} frame, from left to right and top to bottom.

3.2. The IST-TDWZ Quantization

Quantization is the first step to encode the DCT coefficients band b_k, as depicted in Figure 1.

3.2.1 AC Coefficients: Quantization Approach

Two different quantization approaches are followed in the IST-TDWZ solution: 1) The DC coefficients band is quantized using a uniform scalar quantizer without a symmetric quantization interval around the zero amplitude; typically, the DC coefficients band is characterized by high amplitude positive values since each DC transform coefficient expresses the average energy of the corresponding 4×4 samples block. 2) The remaining AC coefficients bands are quantized using the uniform scalar quantizer with a symmetric quantization interval around zero in order to reduce the block artifacts effect. The AC coefficients are mainly concentrated around the zero amplitude. Small drifts between $Y_{2i}{}^{DCT}$ and $X_{2i}{}^{DCT}$ corresponding quantized symbols using a uniform quantizer without a symmetric quantization interval around the zero would result in errors to be corrected by the turbo decoder; if those errors are not corrected through the iterative turbo decoding operation, the annoying block artifact effect becomes visible in the decoded frame X'_{2i}. Thus, using a uniform scalar quantizer with a symmetric quantization interval around zero, low DCT coefficient values around zero are quantized under the same quantization interval index (independently of its signal) avoiding errors between $Y_{2i}{}^{DCT}$ and $X_{2i}{}^{DCT}$ corresponding quantized symbols and therefore reducing the annoying block artifact effect.

3.2.2 AC Coefficients: Varying Quantization Step Size

By letting the decoder know, for each X_{2i} frame, the dynamic range of each DCT coefficients band instead of using a fixed value it is possible to have a quantization interval width (step size) adjusted to the dynamic range of each band. The dynamic range of a given DCT band may be lower than a fixed selected value; since the same number of quantization levels is distributed over a shorter dynamic range, a smaller quantization interval width can be used. The smaller the quantization step size, the lower is the distortion at the decoder. In other words, the DCT coefficients can be more efficiently encoded varying the quantization step size according to the dynamic range of each DCT coefficients band. In the IST-TDWZ codec, the dynamic range for each DCT band of X_{2i} is transmitted frame by frame to the decoder and assumed to be error-free received. For 4×4 pixels block and 8-bit accuracy video data, the DC band dynamic range, is 1024; this value is maintained fixed for all the video frames. The quantization step size W for the AC bands b_k (k=2,…,16) is obtained from $W = 2|V_k|_{\max} / \left(2^{M_k} - 1\right)$ where $|V_k|_{\max}$ stands for the highest absolute value within b_k.

Different performances can be achieved by changing the M_k value for the DCT band b_k. In the IST-TDWZ codec performance evaluation 8 rate-distortion points were considered corresponding to the various 4×4 quantization matrixes depicted in Figure 2.

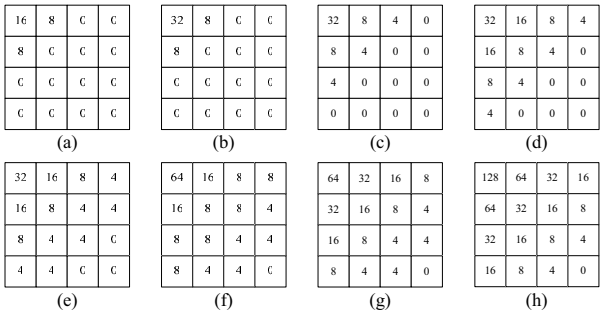

Figure 2 – Eight quantization matrices associated to different IST-TDWZ codec RD performances.

The first 7 matrices in Figure 2 are similar to the ones used in [5] while the matrix in Figure 2 (g) is proposed by the authors of this paper to evaluate the IST-TDWZ RD performance for higher bitrates. Within a 4×4 quantization matrix, the value at position k in Figure 2 indicates the number of quantization levels associated to the DCT coefficients band b_k; the value 0 means that no Wyner-Ziv bits are transmitted for the corresponding band.

3.3. The IST-TDWZ Slepian-Wolf Codec

The turbo encoder structure is similar to the one used in the IST-PDWZ codec: each rate ½ RSC encoder is represented by the generator $\left[1 \quad \frac{1+D+D^3+D^4}{1+D^3+D^4}\right]$. There are however some differences regarding the interleaver length L and the puncturing period P. In the IST-PDWZ solution, the length of the turbo encoder input is the frame size since in the pixel domain the frame is treated as a whole. In the IST-TDWZ solution, the size of each DCT coefficients band is given by the ratio between the frame size and

the number of different DCT coefficients bands. The parity bits produced by the turbo encoder are transmitted according to a pseudo-random puncturing pattern with the same structure as in the IST-PDWZ solution, using however a different puncturing period dynamic range. In the IST-TDWZ implementation, the puncturing period *P* ranges from 1 to 48 instead of 1 to 32, as in the IST-PDWZ solution; the higher dynamic range amplitude allows to obtain PSNR values for lower bitrates.

3.4. The IST-TDWZ Frame Interpolation

The IST-TDWZ codec uses the block-based frame interpolation framework proposed in [3] to generate accurate side information. In a nutshell, both key frames are first low pass filtered to improve the reliability of the motion vectors. Then a block matching algorithm is used to estimate the motion between the next and previous key frame. The bidirectional motion estimation module refines the motion vectors obtained in the previous step by using a bidirectional motion estimation technique; this technique selects a linear trajectory between the next and previous key frames passing at the center of the blocks in the interpolated frame. A spatial smoothing algorithm is then used to improve the accuracy of the motion vector field. Once the final motion vector field is obtained, the interpolated frame can be filled by simply using bidirectional motion compensation as defined in standard video coding schemes. This type of advanced frame interpolation solution has never been used before in the context of a Wyner-Ziv transform domain codec.

4. EXPERIMENTAL RESULTS

Figure 3 illustrates the IST-TDWZ RD performance for 101 frames of the *Foreman* and *Mother and Daughter* QCIF video sequences. In all the experiments, only the luminance data is considered for the RD performance evaluation. The Wyner-Ziv frame rate is 15 fps; the key frames are considered to be losslessly available at the decoder as done for the equivalent codecs in the literature. The dynamic range transmission corresponds to a maximum increase of 240 bits per frame (15 bands/frame times 16 bits/band due to packing purposes); no dynamic range value is sent for the DC band. Since the Wyner-Ziv codec performance is to be evaluated, the RD plots only contain the rate and the PSNR values for the Wyner-Ziv coded frames of a given video sequence. The IST-TDWZ RD performance is compared against H.263+ intraframe coding and H.263+ interframe coding with an I-B-I-B structure.

In the last case, only the rate and PSNR of the B frames is shown. From Figure 3, it is possible to observe that the usage of the transform coding tool provides coding improvements up to 1.8 dB when compared to the IST-PDWZ solution. From the results, it is also possible to conclude that the IST-TDWZ codec provides coding improvements up to 2.1 dB regarding the best equivalent solutions available in the literature [5].

5. CONCLUSIONS AND FUTURE WORK

In this paper, an improved TDWZ video coding solution, called IST-TDWZ, has been proposed. Experimental results show that exploiting the strong spatial correlation among neighboring pixels enables to approximate the Wyner-Ziv video coding efficiency to the interframe H.263+ performance, thus reducing the gap in quality between the two. Coding improvements up to 2.1 dB are achieved for the IST-TDWZ codec regarding the equivalent solutions available in the literature. The quantizer with varying dynamic range and the frame interpolation tools are responsible for the IST-TDWZ codec coding gains. As future work, it is planned to further enhance the RD performance of the codec by using techniques to explore the spatial correlation that exists between neighboring blocks, e.g. like the Intra modes of the H.264/MPEG-4 AVC.

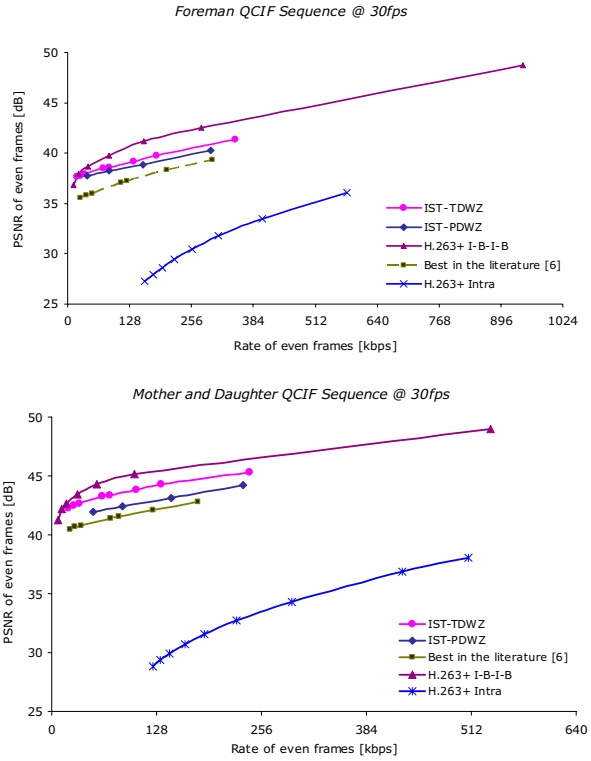

Figure 3 – TDWZ codec rate-distortion performance.

6. REFERENCES

[1] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", *IEEE Trans. on Inform. Theory*, Vol. 19, No. 4, July 1973.

[2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Trans. on Inform. Theory*, Vol. 22, No. 1, January 1976.

[3] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *5th EURASIP*, Slovak Republic, July 2005.

[4] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv Coding for Motion Video", Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, November 2002.

[5] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-Domain Wyner-Ziv Codec for Video", *VCIP*, San Jose, USA, January 2004.

[6] ISO/IEC International Standard 14496-10:2003, "Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding".