Embedding a Block-based Intra Mode in Frame-based Pixel Domain Wyner-Ziv Video Coding *

Alan Trapanese Marco Tagliasacchi Stefano Tubaro
Politecnico di Milano - Italy

João Ascenso Catarina Brites Fernando Pereira
Instituto Superior Técnico - Instituto de Telecomunicações, Lisbon - Portugal

Abstract

Distributed source coding principles have been recently applied to video coding in order to achieve a flexible distribution of the complexity burden between the encoder and the decoder. In this paper we elaborate on a pixel based Wyner-Ziv video codec that shifts all the complexity of the motion estimation phase to the decoder, thus achieving light encoding. We observe that the correlation noise statistics describing the relationship between the frame to be encoded and the side information available at the decoder is not spatially stationary. For this reason we introduce a mode decision scheme either at the encoder or at the decoder in such a way that when the estimated correlation is weak we opt for intra coding on a block-by-block basis. Moreover we discuss the effect of using a side information computed either from lossless or from quantized frames.

1 Introduction

Today's video coding architectures are based on the "downlink" broadcast model, where the video content is encoded once and decoded multiple times. All the ITU-T VCEG and ISO/IEC MPEG standards follow this approach relying on the hybrid block-based motion compensation/DCT transform (MC/DCT) architecture. 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. 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. Distributed video coding lays its foundations on distributed source coding principles stated by the Slepian-Wolf [1] and the Wyner-Ziv [2] theorems. Despite the theory has been well understood since the 70's, only recently practical video coding schemes have been presented targeting different application requirements ranging from low-encoding complexity, robustness to channel losses and scalability.

2 IST-PDWZ video codec architecture

The IST Pixel Domain Wyner-Ziv (IST-PDWZ) video codec we use in this paper [3] is based on the pixel domain Wyner-Ziv coding architecture proposed in [4]. However, there are major differences in the frame interpolation tools further discussed in [3]. This approach offers a pixel domain intra-frame encoder and inter-frame decoder with very low computational encoder complexity. When compared to traditional video coding, the proposed encoding scheme is less complex by several degrees of magnitude. Figure 1 illustrates the global architecture of the IST-PDWZ codec. In this architecture each even frame X_{2i} of the video sequence is called Wyner-Ziv frame and the two adjacent odd frames X_{2i-1} and X_{2i+1} are referred as key frames; in the literature [4] it is assumed that they are perfectly reconstructed (lossless) at the decoder. Each pixel in the Wyner-Ziv frame is uniformly quantized. Bitplane extraction is performed from the entire image and then each bitplane is fed into a turbo encoder. At the decoder, the motion-compensated frame interpolation module generates the side information, Y_{2i} [3], which will be used by the turbo decoder and reconstruction modules. The decoder operates in a bitplane by bitplane basis and starts by decoding the most significant bitplane and it only proceeds to the next bitplane after each bitplane is successfully turbo decoded (i.e. when most of

*The authors wish to acknowledge the support provided by the European Network of Excellence VISNET (http://www.visnetnoe.org)

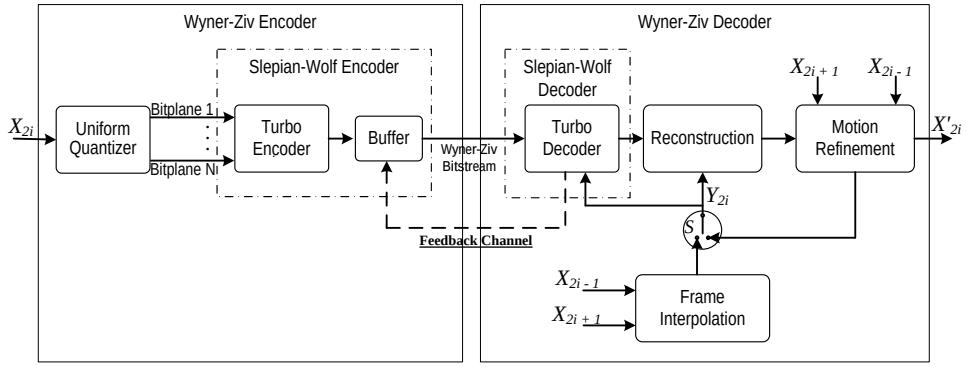

Figure 1: Block diagram of the IST-PDWZ codec

the errors are corrected).

3 Adaptive block-based intra coding in WZ-frames

The turbo codec [4][3] assumes that the correlation statistics between the source to be encoded X_{2i} and its side information Y_{2i} is spatially stationary, thus the same Laplacian distribution is used to describe $X_{2i} - Y_{2i}$, neglecting any dependency on the spatial location and the frame index i. This simplification derives from the (false) assumption that the quality of the motion-interpolation estimate is constant in time and across the whole frame. In this paper we depart from this assumption by understanding that covered/uncovered regions, illumination changes and camera noise do affect the quality of the motion interpolation phase. Although more complex coding schemes might try to adapt the correlation statistics locally on a block-by-block basis, in this paper we start considering a simpler approach that consists of encoding in intra-frame mode those blocks whose correlation with the side information is weak.

Each WZ-frame is divided into non-overlapping 8×8 blocks and each block is either Wyner-Ziv or intra coded. For intra coded blocks we used a DCT-based approach similar to H.263+ intra mode. Wyner-Ziv blocks are encoded as explained in [3] where the only difference is the scanning order that skips intra encoded blocks. In the following we describe the criteria used to perform the mode decision either at the encoder or at the decoder. Then, for each frame we build a binary mode decision map, where each entry in the map indicates whether the block should be intra or Wyner-Ziv coded. In order to efficiently represent this information, we use a simple run-length encoding algorithm. For QCIF sequences at 30fps, we need to send 396 bits (one for each 8×8 block) for each Wyner-Ziv frame, adding up to approximately 5.8kbps. By run length encoding this information we can reduce the cost to 1-1.5kbps.

3.1 Intra mode decision at the encoder

First, we consider the case where the Wyner-Ziv vs intra decision is made at the encoder. As we are targeting a low encoding complexity scenario the encoder is constrained not to perform any motion search. Therefore we need to find a criterion that is easy to compute yet able to infer the quality of the motion-compensated side information available only at the decoder. We consider a simple strategy, similar to the one used in [5][6], by comparing each block $\mathbf{B}$ with the co-located block in the previous frame. In formula:

$$SAD_{\mathbf{B}} = \sum_{(x,y)\in\mathbf{B}} |X_{2i}(x,y) - X_{2i-1}(x,y)| \qquad (1)$$

If $SAD_{\mathbf{B}} > T^{enc}_{intra}$ the block is intra coded. With this modification the coding scheme is no more strictly speaking an intra-frame codec as it requires an extra buffer for the previous key frame.

The encoding process of WZ-frames proceeds as follows:

- for each block evaluate (1) and perform the mode decision (Wyner-Ziv or intra)
- encode and send the binary mode decision map
- encode the intra blocks as in H.263+ intra mode (DCT, scalar quantization, VLC entropy coding)
- extract the bitplanes to be Wyner-Ziv encoded concatenating the pixels of the WZ-blocks (each block is read in raster scan order before proceeding to the next block)
- run the turbo encoder for each bitplane and store the parity bits in a buffer

The decoding process turns out to be:

- build the side information Y_{2i} starting from the neighboring key frames X_{2i-1} and X_{2i+1}
- receive and decode the mode decision map
- receive and decode intra coded blocks

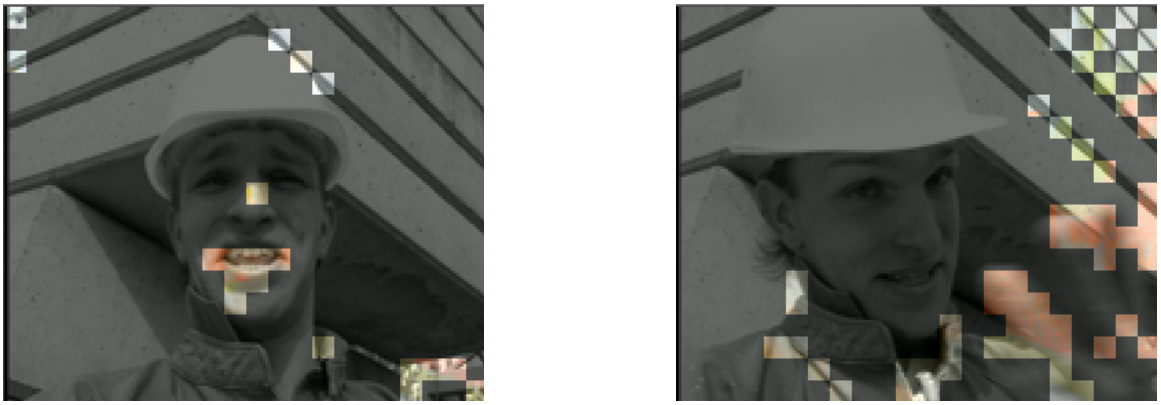

Figure 2: Examples of mode decision map (at the decoder). Greyed areas refer to Wyner-Ziv blocks. *Foreman* sequence. Top: frame 157, when Foreman opens the mouth. Bottom: frame 255, when the hand disappears.

- for each WZ-block, for each bitplane, starting from the most significant one, correct the side information Y_{2i} to reconstruct X_{2i} by requesting parity bits and running the turbo decoding until the error probability is below a predefined threshold

This approach tends to be quite conservative in the mode decision, opting for intra coding also when the correlation with the side information is good (i.e. due to camera panning that cannot be captured without motion estimation). In order to overcome this limitation we propose to shift the mode decision at the decoder side, as detailed in the next section.

3.2 Intra mode decision at the decoder

At the decoder it is possible to better estimate the correlation statistics in sequences characterized by significant motion, as we are free to perform motion estimation. The motion interpolation algorithm adopted in our scheme [3] assigns to each block of the WZ-frame a *direct* motion vector, i.e. the same motion vector is applied with reversed signs to indicate the predictor blocks in the previous and next key frames. The decoder does not know the frame to be decoded, nonetheless it can infer the quality of the motion interpolation by looking at the consistency between the forward and backward predictors:

$$SAD_{\mathbf{B}} = \sum_{(x,y)\in\mathbf{B}} |X_{2i-1}(x+dx, y+dy) + - X_{2i+1}(x-dx, y-dy)|, \quad (2)$$

where the motion vector with components (dx, dy) is applied to block $\mathbf{B}$. If $SAD_{\mathbf{B}} > T^{dec}_{intra}$ then the block is intra coded. Since the mode decision is computed at the decoder, the feedback channel is used to communicate the mode decision map to the encoder. Encoding and decoding are intertwined, as the encoder cannot start processing the frame X_{2i} until the decoder has computed the side information and sent the mode decisions back via feedback channel. For this reason this scheme can be adopted only in real time applications when encoding and decoding take place synchronously. The encoding/decoding process is the following:

- encoder: encode key frames X_{2i-1} and X_{2i+1}
- decoder: decode key frames X_{2i-1} and X_{2i+1}
- decoder: build the side information Y_{2i}
- decoder: perform mode decision based on X_{2i-1} and X_{2i+1} according to (2)
- decoder: encode and send the mode decision map through the feedback channel
- encoder: decode the mode decision map and perform Wyner-Ziv/intra encoding accordingly
- decoder: receive and decode the requested intra coded blocks
- decoder: for the WZ-blocks, for each bitplane, starting from the most significant one, correct the side information Y_{2i} to reconstruct X_{2i} by requesting parity bits and running the turbo decoding until the error probability is below a pre-defined threshold

4 Experimental results

We carried out extensive experimental results on the *Foreman* and *Coastguard* sequences in order to evaluate the effect of the mode decision. The first experiment applies the mode decision to the scheme that uses lossless keyframes as reported in the literature [4]. In this case we observed no improvements or even loss of coding efficiency with respect to the case that encodes the whole frame in Wyner-

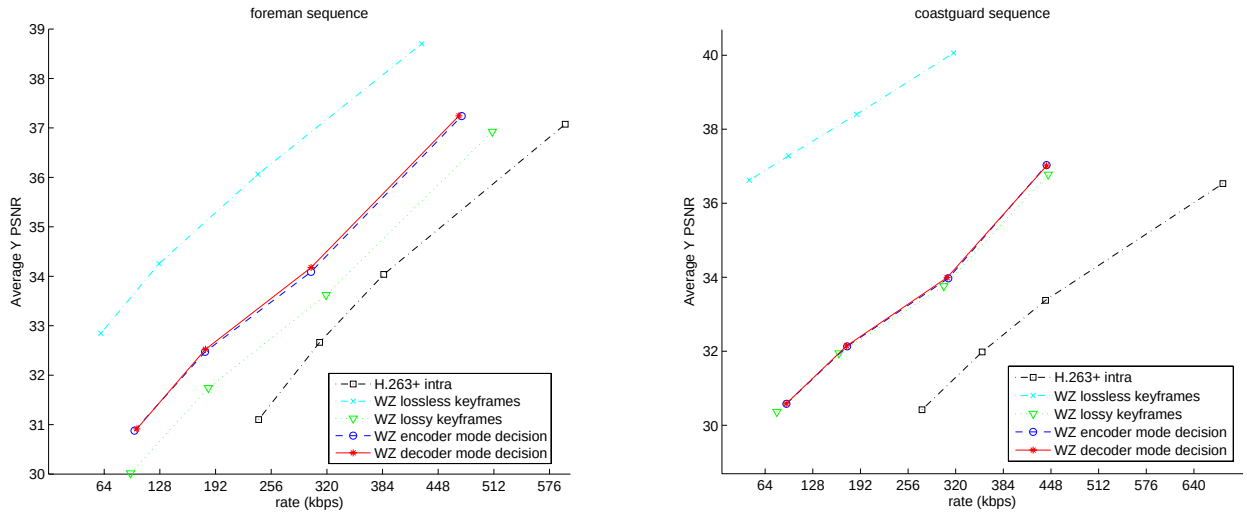

Figure 3: Left: *Foreman* sequence. QCIF@30fps. 400 frames. Right *Coastguard* sequence. QCIF@30fps. 300 frames

Ziv mode. The reason is that the quality of the side information estimated from lossless key frames is so good that intra coding part of the frame turns out to be inconvenient in rate-distortion sense. Nevertheless the assumption of perfectly reconstructed (lossless) key frames is unpractical in real world applications: first, the reconstructed sequence exhibits quality fluctuations, especially at low bitrates; second, the bit budget required to lossless encode the key frames can be much larger than for the Wyner-Ziv frames.

By lossy encoding the key frames, the coding efficiency drops significantly, as shown in Figure 3. For the lossy key frames we choose a quantization step size that gives approximately the same quality as the Wyner-Ziv frames when using a given number of bitplanes. We use a QP ($qstep = 2 \cdot QP$ [1]) equal to 13, 10, 8, 5 for each of the decoding test point reported in Figure 3. In this scenario the mode decision scheme seems to give a coding efficiency gain, as much as 1dB on average for *Foreman* and 0.3dB for *Coastguard*. Some of the frames (between frames 260 and 320) take more advantage of the mode decision gaining up to 3dB. The mode decision at the decoder is only slightly better than at the encoder. We set the quantization step size of the intra coded blocks to be the same as for the key frames. Figure 2 shows an example of a mode decision map. We can observe that the mode decision is consistent to what could be expected, as parts of the images that cannot be properly motion interpolated from the neighboring key frames are intra coded.

[1] as in H.263+ standard

5 Conclusions

In this paper we describe a intra mode decision scheme for frame-based pixel domain Wyner-Ziv video coding. We show that it is possible to achieve slight coding efficiency gains when the motion-compensated interpolation fails. Our future work will integrate the intra mode decision into a DCT-based Wyner-Ziv codec.

References

[1] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, January 1976.

[3] J. Ascenso, C. Brites, and F. Pereira, "Interpolation with spatial motion smoothing for pixel domain distributed video coding," in *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, (Slovak Republic), July 2005.

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proceedings of the 36th Asilomar Conference on Signals, Systems, and Computers*, vol. 1, (Pacific Grove, CA), pp. 240–244, October 2002.

[5] R. Puri and K. Ramchandran, "PRISM: A New Robust Video Coding Architecture based on Distributed Compression Principles," in *Allerton Conference on Communication, Control and Computing*, (Urbana-Champaign, IL), October 2002.

[6] R. Puri and K. Ramchandran, "PRISM: A video coding architecture based on distributed compression principles," Tech. Rep. No. UCB/ERL M03/6, ERL, UC Berkeley, March 2003.