Extrapolating Side Information for Low-Delay Pixel-Domain Distributed Video Coding

Luís Natário[1], Catarina Brites[2], João Ascenso[3], Fernando Pereira[4]
[1,2,4]Instituto Superior Técnico – Instituto de Telecomunicações
[3]Instituto Superior de Engenharia de Lisboa – Instituto de Telecomunicações
{luis.natario; catarina.brites; joao.ascenso; fernando.pereira}@lx.it.pt

***Abstract*—Distributed Video Coding (DVC) is a new video coding approach based on the Wyner-Ziv theorem. Unlike most of the existing video codecs, each frame is encoded separately (either as a key-frame or a Wyner-Ziv frame) which results in a simpler and lighter encoder since complex operations like motion estimation are not performed. The previously decoded frames are used at the decoder to estimate the Wyner-Ziv frames – the frames are coded independently but jointly decoded. To have a low-delay codec, the side information frames (estimation of the Wyner-Ziv frames to be decoded) must be extrapolated from past frames. This paper proposes a robust extrapolation module to generate the side information based on motion field smoothening to provide improved performance in the context of a low-delay pixel-domain DVC codec.**

Keywords:* *distributed video coding, side information, motion extrapolation, low-delay

I. INTRODUCTION

Most of the existing coding schemes, namely the popular MPEG standards, are based in an architecture where the decoder is typically much more complex than the decoder mainly due to the computationally consuming operation of motion estimation done at the encoder. The Distributed Video Coding (DVC) approach based on the Wyner-Ziv (WZ) theorem (which is the extension of the Slepian-Wolf theorem for the lossy case with side information available at the decoder) allows reversing this scenario by shifting the motion estimation complexity from the encoder to the decoder enabling applications where the encoder's low complexity is a requirement. The Slepian-Wolf theorem states that is possible to compress in a distributed way (separate encoding and joint decoding) two statistically dependent signals at a rate similar to the rate obtained using a system where the signals are encoded and decoded jointly (as in the traditional video coding schemes).

In DVC schemes, each frame is encoded independently from previous and subsequent frames which results in a decrease of the typical encoding complexity. In order to have a low-delay codec, the frames must be decoded regardless of future frames, i.e. the side information must be created by extrapolation (as opposed to create the side information by interpolation using also future frames).

This paper proposes a side information extrapolation module that is able to generate accurate side information by employing an extrapolation model that uses overlapped motion estimation, motion field smoothening and spatial-interpolation for uncovered areas. This low-delay DVC architecture is particularly well suited for emerging applications where the encoder complexity must be as low as possible and low-delay is a 'must have' like in wireless low-power surveillance and mobile camera phones among others.

II. PIXEL-DOMAIN WYNER-ZIV CODEC ARCHITECTURE

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

The IST-Wyner-Ziv (IST-WZ) codec developed at IST [1] is based on the pixel-domain coding architecture proposed in [2]. The scheme, modified to support the low-delay extrapolation module, is depicted in Figure 1.

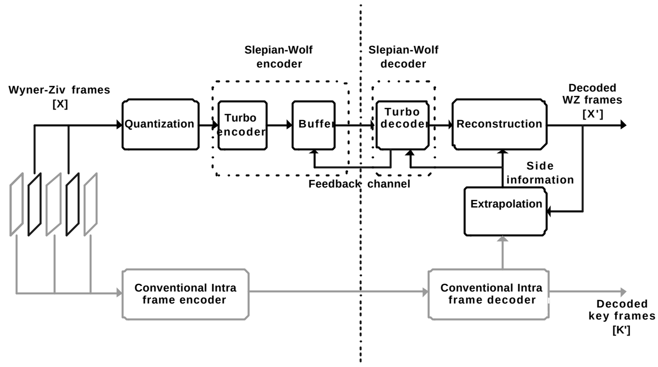

Fig. 1 – Wyner-Ziv codec architecture with side information extrapolation

In the IST-WZ codec, the frames encoded are of two types: key-frames and WZ-frames. The key-frames are intra coded using H.263+, for example. The WZ-frames, after being uniformly quantized, are encoded using a turbo-based Slepian-Wolf encoder.

The key-frames (and previously decoded WZ-frames, if decided) are used by the decoder to generate, by extrapolation, the side information that along with the WZ-bits received will be used to decode the WZ-frames. The Slepian-Wolf encoder generates sequences of parity bits for each bitplane output by the quantizer. These bits (which depend on the turbo encoder rate) are punctured and stored at the Slepian-Wolf encoder's buffer. Depending on the quality of the side information generated, more or less WZ-bits will be requested (via feedback channel) by the decoder to ensure that a given WZ-frame is successfully decoded with a given bit error probability. The side information generated must be as close to the original as possible to ensure that a minimum amount of bits is requested to decode a given WZ-frame.

III. EXTRAPLOLATING THE SIDE INFORMATION

Given a video sequence, one can predict a forthcoming frame at the decoder based on the past, i.e. using the previously decoded frames, by extrapolation. The advantage of using extrapolation (and not interpolation) for the generation of the side information is to enable a low-delay codec, since to decode a given frame no future frame is needed. The performance of this type of coding architecture (DVC) is fundamentally determined by the quality of the predicted frame (side information produced at the decoder) because, if the extrapolated side information frame is very similar to the WZ-frame being decoded, few coding errors have to be corrected and, therefore, few parity bits from the encoder's buffer are requested by the decoder, resulting in a low bitrate for the WZ-frames.

Several methods can be thought to extrapolate a side information frame. The simplest approach is to use the previous decoded frame. However, and since motion is generally present in video sequences, an extrapolation based on the motion observed in the previously decoded frames fits better the purpose of producing an extrapolated frame similar to the WZ-frame being decoded. The motion estimation must be done carefully in order to ensure that only true motion is captured and that a reliable extrapolated frame is produced. Since the motion observed in the previously decoded frames is projected to the WZ-frame time slot, the motion estimation targets capturing the blocks that minimize some distortion measure and are associated to the true motion. The side information extrapolation module proposed in this paper is composed of several blocks (as depicted in Figure 2):

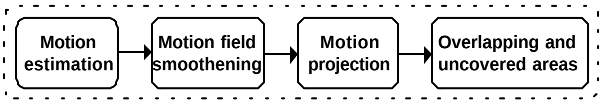

Fig. 2 – Side information extrapolation module

- **Motion estimation** – Motion vectors are estimated for overlapped 8×8 pixel blocks using the two previously decoded frames (Figure 3). The block overlapping (superposition of the blocks used to perform motion estimation) is used to reduce the block artifacts in the side information frame caused by block motion-compensation.

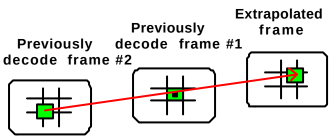

Fig. 3 – Motion projection

- **Motion field smoothening** – For each block, a new motion vector is calculated by averaging all neighboring motion vectors. This leads to a smoothened motion vector field where true motion is captured and a better side information frame obtained.
- **Motion projection** – The pixels from the last decoded frame (or other) are projected to the next time instant using the motion field obtained above assuming that the motion is linear and that, therefore, the warping of frame *i-2* into frame *i-1* will linearly continue from frame *i-1* to frame *i* (Figure 3).
- **Overlapping and uncovered areas** – Whenever one pixel is estimated by more than one pixel in the previous frames, an average between the values is taken as the prediction for that position. Whenever no pixel in the previous frames is assigned as a prediction for a given frame, it is predicted by local spatial interpolation from three neighbors (up, left and up-left), scanning the frame from top to bottom and left to right.

IV. TESTS AND RESULTS

The tests performed envisioned evaluating, first of all, the extrapolation module with the tools proposed versus similar systems present in the literature [3]. The results obtained for the sequence *Foreman* using the first 100 frames (using a frame structure *I-WZ-I-WZ* and accounting only the WZ-frames bitrate) show in Figure 4 that the system proposed here is performing slightly better than the best comparable results from the literature [3] (for the single sequence for which comparable results are available). Figure 4 also shows that using the proposed extrapolation module either than the interpolation module proposed in [1] results in a significantly poorer RD performance. When the side information is extrapolated instead of being interpolated between two key-frames a worse prediction of the side information is obtained – this is a direct consequence of the low-delay constraint.

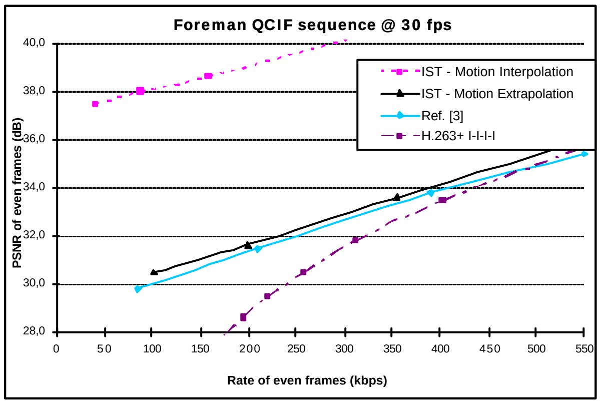

Fig. 4 – RD performance comparison for Foreman

Figure 5 presents the RD performance using the extrapolation module proposed, for different GOP-sizes (QCIF), for the rather still surveillance like sequence *Galleon* (src20) from the VQEG test sequence set. The usage of fewer key-frames results in a progressive decrease of the quality since that the side information generated by projecting a decoded WZ-frame does not offer the same quality of projecting a decoded intra frame. Nevertheless, the RD results obtained for all GOP sizes tested clearly outperform the RD results obtained using a H.263+ intra encoder.

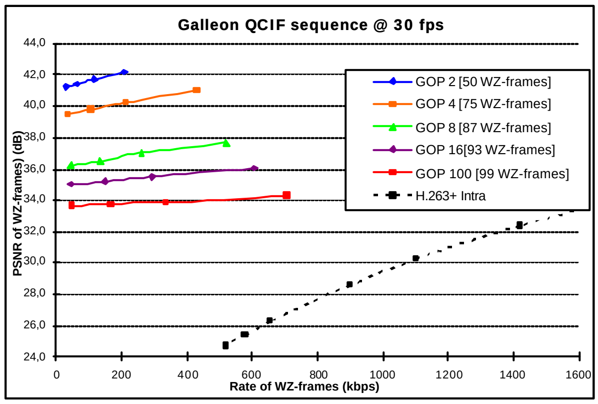

Fig. 5 – RD performance for Galleon using different GOP sizes

V. CONCLUSIONS

The proposed Wyner-Ziv codec solution shows that it is possible to have a low-delay codec with a simple and lightweight encoder providing substantially better quality than traditional intra encoders. This extrapolation-based DVC codec is especially well suited for stable sequences (e.g. surveillance sequences) where low-delay and low encoder complexity are strong demands.

REFERENCES

[1] J. Ascenso, C. Brites and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", 5th EURASIP Conf. on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, July 2005

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

[3] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-Domain Wyner-Ziv Codec for Video", Proc. Visual Communications and Image Processing, pp. 520-528, San Jose, USA, January 2004