Anantrasirichai, N., Agrafiotis, D., & Bull, D. R. (2008). A concealment based approach to distributed video coding. In 15th IEEE International Conference on Image Processing, 2008 (ICIP 2008), San Diego. (pp. 2232 - 2235). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICIP.2008.4712234

Link to published version (if available):
10.1109/ICIP.2008.4712234

A CONCEALMENT BASED APPROACH TO DISTRIBUTED VIDEO CODING

Nantheera Anantrasirichai, Dimitris Agrafiotis, Dave Bull

University of Bristol, Woodland Road, Bristol, BS8 1UB, UK

ABSTRACT

This paper presents a concealment based approach to distributed video coding that uses hybrid Key/WZ frames via an FMO type interleaving of macroblocks. Our motivation stems from a previous work of ours that showed promising results relative to the more common approach of splitting the sequence in key and WZ frames. In this paper, we extend our previous scheme to the case of I-B-P frame structures and transform domain DVC. We additionally introduce a number of enhancements at the decoder including use of spatio-temporal concealment for generating the side information on a MB basis, mode selection for switching between the two concealment approaches and for deciding how the correlation noise is estimated, local (MB wise) correlation noise estimation and modified B frame quantisation. The results presented indicate considerable improvement (up to 30%) compared to corresponding frame extrapolation and frame interpolation schemes.

Index Terms— Data compression, DVC, Concealment

1. INTRODUCTION

The requirement of having very low complex video transmitters in many applications, e.g. low cost low power sensors, has motivated the video coding community to re-visit the information theory principles of Slepian-Wolf (SW) [1] and Wyner-Ziv (WZ) [2] and examine distributed video coding (DVC) systems [3]. The most common approach to DVC is that where the frames of a single source video are split into two categories, key frames and WZ frames [4][5][6]. Key frames are intra coded with a conventional encoder and are made available at the decoder. WZ frame coding, which may involve transformation, uses quantisation followed by channel coding applied in a bitplane by bitplane fashion, with the parity bits only being transmitted to the decoder. At the decoder, the key frames are used for creating an estimate of the WZ frames (side information). This side information (SI) is seen as the systematic part of the channel encoder's output as received at the decoder, i.e., the side information is seen as a noisy version of the original coded WZ frame. The received parity bits are used to correct the errors present in this noisy version of the WZ data. An excellent list of relevant papers can be found here [7].

In a previous work of ours [8] we proposed the use of Hybrid Key/Wyner-Ziv frames (KWZ) for better generation of the side information and more accurate estimation of the correlation noise. According to this approach the process of generating the side information is treated as a concealment task. By employing a macroblock pattern similar to the one specified by the dispersed flexible macroblock ordering (FMO) of H.264, we group the macroblocks of each frame into intra coded (key) and Wyner-Ziv groups. Temporal concealment is then used at the decoder for "concealing" (predicting) the missing WZ macroblocks using the information available from the already received 4-neighboring key MBs. The same key MBs are also used for estimating the correlation noise through motion estimation for the whole WZ group.

This approach was tested and found to offer performance advantages for the low delay (P frame) pixel domain DVC case relative to the more common approach of splitting the sequence in key and WZ frames, wherein SI generation involves concealing/recovering a whole missing frame; a process normally associated with poor quality results. Moreover in frame based approaches the estimation of the correlation noise relies on predicted and not actual received data.

In this paper, we extend our previous scheme to the case of I-B-P frame structures and transform domain DVC. We additionally introduce a number of enhancements at the decoder for improving the performance of our codec. The results presented indicate considerable performance improvements relative to frame based schemes. More specifically the contributions of this paper are the following:

- Extension of the Key-Wyner/Ziv framework to the IBP and transform domain DVC scenario
- Use of spatio-temporal concealment for generating the side information on a MB basis
- Mode selection for switching between the two concealment approaches and for deciding how the correlation noise is estimated
- Local (MB wise) correlation noise estimation
- Modified B frame quantisation

The rest of this paper is organized as follows: Section 2 describes the proposed scheme in four sub-sections. The GOP structure used is discussed in 2.1 the quantization of B-frames in 2.2, the side information generation in 2.3 and the correlation noise estimation in 2.4. Section 3 presents results and comparisons and Section 4 concludes this paper.

2. PROPOSED SCHEME

The first step in the proposed framework involves splitting of the current frame into Key and WZ groups of macroblocks, in a similar fashion to the dispersed FMO specified in H.264. Each Key group is encoded with H.264 in Intra mode. If transform-domain coding is applied, the WZ MBs are first transformed, using the same DCT like transform that is employed in H.264. The WZ MBs (pixels or coefficient values) are then quantised and bit-planes of the quantised symbols are extracted. These are then fed to the turbo encoder and parity bits are produced which are stored in a buffer. At the decoder, the SI is created using temporal and spatial error concealment (TEC/SEC) methods. The turbo decoder uses the parity bits and the (possibly transformed) SI to form the decoded bit-planes. If transform-domain coding is enabled, the reconstructed WZ group is then inverse-transformed. Finally, the decoded WZ group and the decoded Key group are merged and a de-blocking filter is applied to remove blocking artefacts occurring between WZ and Key MBs. The proposed scheme is described in detail below.

2.1. GOP structure

The proposed method employs a group-of-pictures (GOP) structure as the one illustrated in Figure 1 for a GOP length of 6. A fully intra coded Key frame is placed at the first frame of each GOP. The subsequently decoded P frames employ the two most recently decoded frames (Key and KWZ) as references for generating the SI. The two MB groups alternate from one P frame to another P frame so as to avoid creating potentially annoying regions of different subjective quality. The MB groups of the B frames alternate relative to the previous (in display order) reference frame**.**

2.2. B-Frame Quantisation

The SI generation for the WZ macroblocks of the B frames relies on bidirectional error concealment that utilises both forward and backward frames for predicting the "missing" MBs. As a result the SI for the WZ MBs of these frames is more reliable and closer to the transmitted data resulting in a lower bitrate for these frames. In order to further reduce this bitrate we increase the quantisation step size of these WZ MBs relative to the step size used for the WZ MBs of the P frames. As in the scalable extension of H.264 (SVC) [9] the associated PSNR fluctuation within the GOP should not appear subjectively annoying as long as the number of quantisation levels is chosen wisely. The relationship between the number of quantisation levels Q_B^n and Q_P^n for B and P frames respectively, for *n* bit-planes can be written as:

$$Q_B^n = \lceil a \cdot Q_P^n \rceil \qquad (1)$$

Based on experimental results, we have selected *a* so that $Q_B^n = 1/2(Q_P^n + Q_P^{n-1})$ i.e. $a = 0.75$. The resulting values of Q_B^n and Q_P^n for this value of *a* are shown in Table 2.1 for different *n*. Note that even though the number of bitplanes is not reduced (i.e. the number of parity bits produced at the encoder remains the same) the bitrate is decreased due to a smaller request for parity bits from the more significant bitplanes as shown in Table 2.2 (WZ data of the 60th frame of the *foreman* sequence in CIF format).

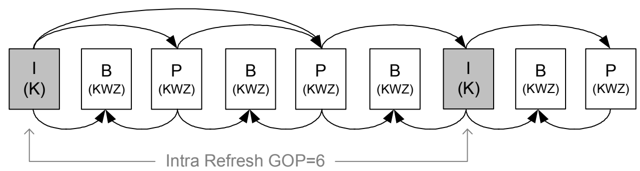

Figure 1: GOP structure of the proposed FMO DVC system.

Table 2.1 Quantisation level for P frame and B frame.

# of bit-planes (*n*)	Q_P^n	Q_B^n
4	16	12
3	8	6
2	4	3
1	2	2

Table 2.2 Bit spending for each decoded bit-plane of a B-Frame WZ group for different number of quantisation levels (BER 10^{-3})

Bit-plane	$Q_B^3 = Q_P^3$	$Q_B^3 = 0.75 Q_P^3$
1 (MSB)	12672	6336
2	22176	12672
3 (LSB)	31680	31680

2.3. Side Information Generation

The generation of the side information is equivalent to an error concealment process for missing macroblocks (WZ MBs) in the presence of all their 4-neighbours. We employ temporal and spatial error concealment methods (TEC/SEC) the application of which is controlled by a mode selection algorithm. A detailed description of the methods employed can be found in [10]. Below we briefly describe how they are applied in the context of P and B frame SI generation for each WZ macroblock.

The employed TEC method uses the external boundary matching error (EBME) of a WZ MB -defined as the sum of absolute differences between the multiple pixel boundary of MBs adjacent to the missing one (WZ) in the current frame and the same boundary of MBs adjacent to the replacement MB in the reference frame (see Figure 2) - in order to test possible motion vectors (MV_{WZn}) from a list that includes those of spatially and temporally adjacent MBs as well as the zero MV. Spatially adjacent MVs are generated through forward (and backward for the case of B frames) motion estimation for all 8x8 blocks of the Key MB group (MV_{Kn}) prior to the concealment process. The search range of the motion estimation is adjusted according to the distance between the current and the reference frame. The initial estimation of a WZ MV is further refined through overlapped block motion compensation (see Figure 2). In the case of B frames the replacement (SI) macroblock can also result from averaging a forward and backward replacement MB depending on the EBME.

The SEC module uses bordering Key pixels to conceal the missing WZ pixels of each WZ MB through bilinear interpolation or directional interpolation along detected edges. The type of interpolation used depends on the outcome of a decision algorithm that uses the directional entropy of neighbouring edges for choosing between the two interpolation approaches.

The mode selection algorithm examines the suitability of the TEC method for concealing each WZ MB, by evaluating the levels of motion compensated activity and spatial activity in the neighbourhood of that MB and switching to spatial concealment accordingly. Motion compensated temporal activity is measured as the mean squared error between the key MBs surrounding the missing one in the current frame and those surrounding the replacement MB in the reference frame. Spatial activity is measured as the variance of the surrounding key MBs in the current frame. More formally:

$$SA = E\,[\,(x-\mu)^2\,] \text{ and } TA = E\,[\,(x-x^*)^2\,] \tag{2}$$

where x are the pixels in the neighbourhood of the missing MB and x^* are the pixels in the neighbourhood of the replacement MB in the reference frame. SEC is employed if the spatial activity is smaller than the temporal activity and the latter is above a specific threshold (3 in this work). Otherwise TEC is used.

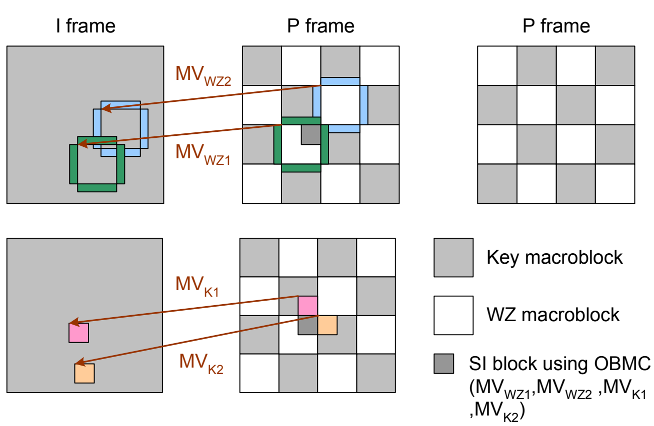

Figure 2: Example of the TEC process, as applied to a P frame.

2.4. Correlation Noise Estimation

We estimate the correlation noise on a macroblock basis using the 4-neighbouring Key macroblocks of each WZ MB. We model the noise as a Laplacian distribution with a specific variance that changes from MB to MB. The estimation of this variance is performed in one of two ways depending on the outcome of the concealment mode selection algorithm described in section 2.3.

If TEC is selected, we employ the difference between the 4-neighboring key MBs and the corresponding motion compensated MBs in previously decoded frames, that were found to provide the best match during the motion estimation process, in order to estimate the Laplacian distribution parameter-α. In other words, after having performed motion estimation for the key MBs of the current frame as described in section 2.3 we take the difference between each key pixel and its best match (as indicated by ME) in one of the reference frames. The resulting distribution should follow closely that of the difference between the SI MB and the transmitted WZ MB as it employs actual received pixels in the vicinity of the processed MB, as opposed to frame based interpolated values (Figure 3). If SEC is applied for creating the SI MB then α is calculated using the variance of the difference between the 4-neighbouring Key MBs and this SI MB. As shown in the result section the performance of the codec is improved by this more accurate approach of estimating α.

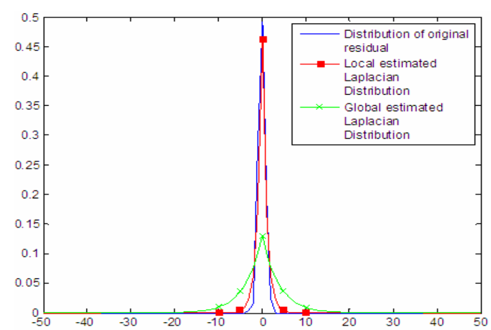

Figure 3: Correlation noise estimation example.

3. RESULTS AND DISCUSSION

The proposed codec was compared to the turbo-based pixel domain Wyner-Ziv coding scheme of [5] (frame interpolation) as well as that of [11] (frame extrapolation) by using two standard CIF video sequences; *foreman* and *hall*. Key frames (for the reference methods) and Key slice groups (for the proposed method) were intra-coded with H.264. Quantisation parameters of 40, 38, 36 and 32 were chosen so that the quality of the WZ data at quantisation levels of 2, 4, 8 and 16 respectively for P frames was close to the quality of the Key data. For transform-domain coding we employed the quantisation set introduced in [6]. A turbo encoder with two identical ½ rate constituent convolutional encoders was employed with a random puncturing period of 32 and a maximum number of iterations of 18. The acceptable bit error rate threshold was set to 10^{-3} for each bit-plane.

The rate-distortion performance for *foreman* and *hall* is shown in Figure 4 and Figure 5 respectively. The bitrates shown in the plots are for all coded frames, i.e. all hybrid

Key/WZ frames for the proposed method and all frames, including the key frames, for the reference methods. The plots show that the proposed pixel domain scheme offers a significant performance improvement at medium and higher bit rates and similar performance to the interpolation frame based schemes at low bit rates. Our method achieves around 30% reduction in bitrate with the high/irregular motion *foreman* sequence and a 10% reduction with *Hall*. The individual enhancements offered by the different elements of the codec (B frame quantisation, local alpha estimation) can also be seen in the two graphs. The performance benefits of the proposed scheme become more obvious for the case of transform domain DVC where our codec consistently outperforms the frame based scheme for both sequences and at all bit rates.

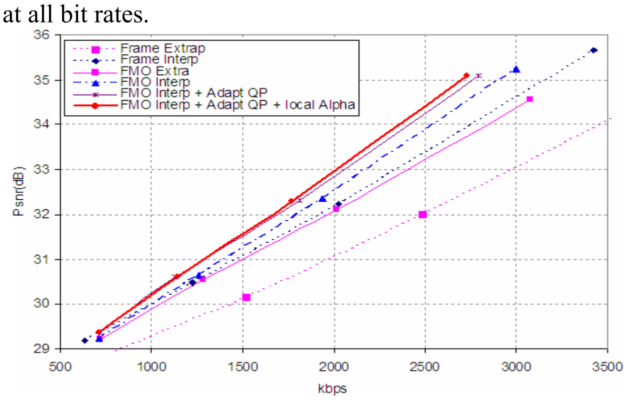

Figure 4: Pixel domain DVC performance (foreman @ 25fps).

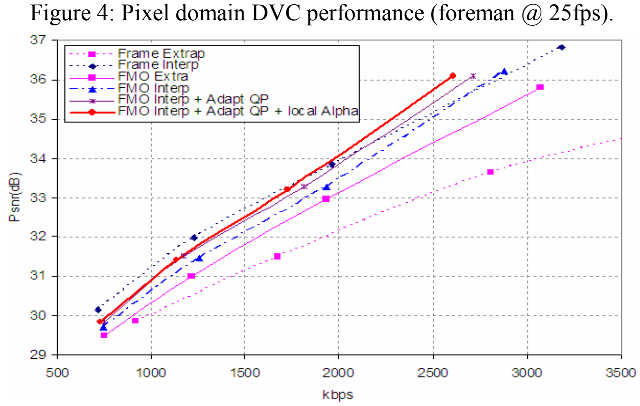

Figure 5: Pixel domain DVC performance (Hall @ 25fps).

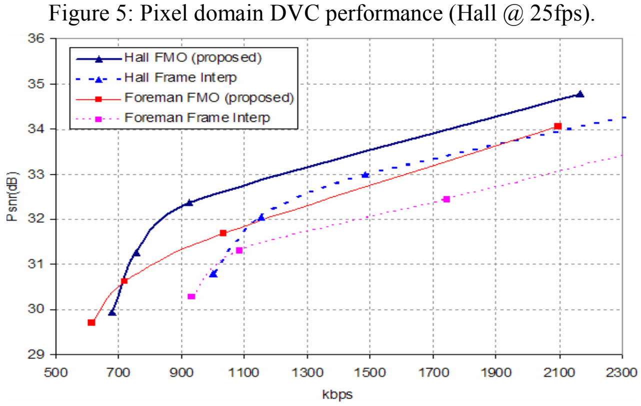

Figure 6: Transform domain DVC performance at 25fps.

4. CONCLUSIONS

We have presented a concealment based approach to distributed video coding that uses hybrid Key/WZ frames via an FMO type interleaving of macroblocks. Our approach allows better SI generation and more accurate correlation noise estimation, both performed at the MB level. Spatio-temporal concealment with mode selection is used for predicting the WZ MBs, with the mode selection also controlling the local alpha estimation. Results for both I-P-P and I-B-P GOP structures with and without use of a transform, show considerable improvement compared to corresponding frame extrapolation and frame interpolation schemes respectively (up to 30% reduction in bitrate – 1.2dB PSNR improvement). Part of this improvement is also due to the proposed quantisation of B-frames.

ACKNOWLEDGMENT

This work was done as part of the EPSRC funded grant EP/E065538.

REFERENCES

[1]. D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," IEEE Transactions on Information Theory, vol. 19, no. 4, 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, no. 1, January 1976.

[3]. B. Girod, A. Aaron, S. Rane, D. Rebello-Mondero, "Distributed Video Coding," Proc. of the IEEE, vol. 93, no. 1, Jan. 2005.

[4]. A. Aaron, R. Zang, and B. Girod, "Wyner-Ziv Coding of Motion Video," in ASILOMAR Conf. on Signals and Systems, Nov. 2002.

[5]. J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," in EURASIP, Video/Image Processing and Multimedia Communications, June 2005.

[6]. C. Brites, J. Ascenso, F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance", in Proc ICASSP, May 14-19, 2006

[7]. The DISCOVER project, http://www.discoverdvc.org/

[8]. D. Agrafiotis, P. Ferré, D. R. Bull, "Hybrid key/Wyner-Ziv frames with flexible macroblock ordering for improved low delay distributed video coding", VCIP 2007, 2007, California, USA.

[9]. H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.9, pp.1103-1120, Sept. 2007

[10]. D. Agrafiotis, D. R. Bull, N. Canagarajah, "Enhanced error concealment with mode selection", IEEE Trans on Cir. Systs Video Tech., vol. 16, no. 8, pp. 960-973, August 2006.

[11]. L. Natario, C. Brites, J. Ascenco and F. Pereira "Extrapolating Side Information for Low Delay Pixel Domain Distributed Video Coding", Int. Workshop on Very Low Bitrate Video Coding, Sept. 2005