ENHANCED SPATIALLY INTERLEAVED DVC USING DIVERSITY AND SELECTIVE FEEDBACK

Nantheera Anantrasirichai, Dimitris Agrafiotis, Dave Bull

Department of Electrical & Electronic Engineering,
University of Bristol, Bristol, BS8 1UB, UK

ABSTRACT

Systems with cheap/simple/power efficient encoders but complex decoders make applications such as low cost, low power remote sensors practical. Bandwidth considerations however are still an issue and compression efficiency has to remain high. In this paper, we present a distributed video codec (DVC) that we are developing with the aim of achieving such a low power paradigm at the cost of only a small compression performance deficit relative to the current state of the art, H.264. The proposed system employs spatial interleaving of KEY and Wyner-Ziv data which allows efficient side information (SI) generation through block-based error concealment, a Gray code that increases the accuracy of bit probability estimation, and a diversity scheme that produces more reliable results by exploiting multiple SI generated data. Simulation results show an improvement of the proposed scheme over H.264 intra coding of up to 1.5 dB. We additionally propose two mechanisms for selective parity bit feedback requests that can further reduce the WZ bitrate by up to 15%.

***Index Terms*—** DVC, Video coding, Multi-hypothesis

1. INTRODUCTION

Distributed video coding (DVC) is currently seen as a promising approach to low power, low complexity encoding which, in theory at least [1,2], could offer similar performance to that of typical hybrid video codecs. A common DVC scenario involves splitting of the video frames in two categories, KEY frames and Wyner-Ziv (WZ) frames [3,4], which undergo conventional (usually intra)coding and Wyner-Ziv coding respectively. The WZ data, which may undergo transformation, are quantised and fed to a channel coder in a bit-plane by bit-plane fashion. At the decoder, the decoded KEY frames are used for creating an estimate of the WZ frames, the side information (SI). This SI is seen as the systematic part of the channel coder output, i.e. a noisy version of the original WZ frame. The received parity bits are used to correct the errors present in the SI. Clearly the number of bits spent for the WZ data mainly depends on how accurate the SI and the estimation of statistical behaviour of the virtual channel are.

In this paper, we employ our previously proposed hybrid KEY/WZ frame approach whereby spatial interleaving of KEY and WZ blocks takes place using a chessboard pattern similar to the FMO dispersed pattern of H.264 [5].We refer to our codec as spatially interleaved DVC (SPI-DVC). Generating the SI in such a hybrid frame allows the use of neighbouring intra coded data through application of an error concealment scheme. The proposed technique addresses problem encountered with non-linear motion trajectories which frame-based methods cannot easily handle [6,7]. The KEY/WZ frames also allow for more precise noise estimation, performed at the block level.

We introduce the use of a Gray code in our SPI-DVC codec, first introduced to DVC in [8].The Gray code can improve the system performance since it has a hamming distance of one which provides higher error resilience than the natural binary code. It also offers a higher accuracy of bit probability estimation. Here, we apply an XOR (exclusive OR) operator directly on the quantisation symbols. This allows modification in both pixel- and transform-domains.

To increase the accuracy of the statistics used in the turbo decoding, we follow a diversity technique, similar to those employed in wireless communication systems. Diversity techniques exploit multiple received signals (e.g. through the use of multiple antennas) to combat problems caused by channel fading. In the DVC scenario, this process translates into a multi-hypothesis SI framework, wherein multiple SI data available at the decoder are used to determine the conditional probability function of the source. A more precise log-likelihood ratio is consequently computed thereby reducing the number of bits requested by the turbo decoder for correcting the SI [9]. Moreover, we propose a refined WZ reconstruction which allows usage of the correct SI to compensate for the errors appearing in the others.

We additionally propose two selective feedback mechanisms in order to reduce the number of requested parity bits. In most DVC systems this request for additional parity bits involves the update of all regions of the coded frame irrespective of the quality levels of the local SI. As a result unnecessary bits are sent leading to a significant increase of the total bitrate. The methods rely either on

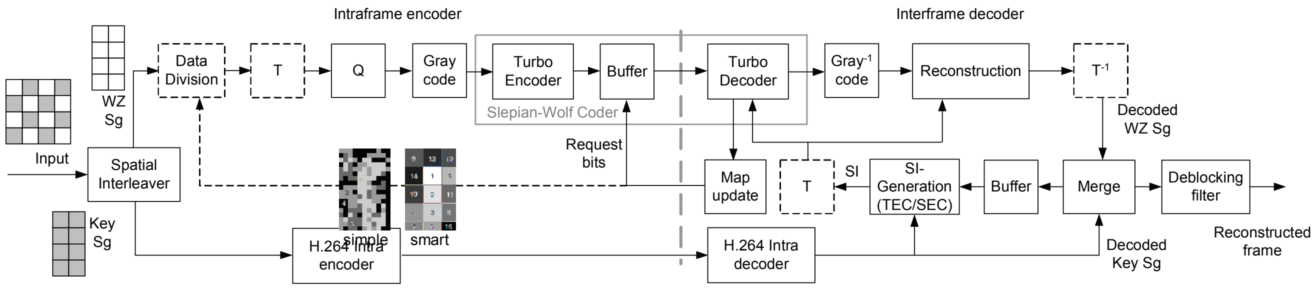

Fig. 1: The proposed SPI-DVC codec

reliability maps sent by the decoder to the encoder or on prior agreement of the encoder/decoder which allow selective parity bit transmissions for those regions that are associated with poor SI.

The rest of paper is organised as follows: Section 2 describes the proposed codec. Experimental results and discussion are presented in Section 3. The conclusions and future work are presented in Section 4.

2. PROPOSED CODING SCHEME

The proposed SPI-DVC codec is illustrated in Fig. 1 and described in this section.

2.1. Spatial Interleaving

The first step involves splitting of the current input frame into KEY and WZ groups in a similar fashion to the dispersed type of flexible macroblock ordering (FMO) specified in H.264 [10]. The block size is fixed for the entire input sequence and can range from 16x16 to 4x4 pixels. Smaller interleaving steps than 4x4 should not be used as they can interfere with the de-correlation property of the 4x4 transformation. Bigger block sizes can benefit KEY coding as they allow better utilisation of context information, while the smaller block sizes benefit the performance more, as better prediction (concealment) of the SI should be achieved especially in areas of high motion. After the interleaving process, the KEY blocks are, normally, horizontally shifted to make a new frame of the same height but half the width of the original which is then encoded with H.264 in intra mode. When spatial prediction is used, further temporal interleaving of the KEY groups of two consecutive frames can be applied to avoid any performance loss relative to full frame KEY coding, at the cost of a small additional delay. The same can be applied to the respective WZ groups in order for the frame length of the input to the turbo encoder to be adequate for good performance.

2.2 A Gray Code

At the WZ encoder, the quantised symbols are converted into binary data. Subsequently, before extracting bit-planes, the binary codes are converted to Gray codes by XORing the binary values with their logical-shift-right values. At the decoder, the decoded Gray codes are converted back to binary values in a similar fashion. The Gray code conversion obviously adds a small complexity to the system, but improves the codec's performance.

2.3 Side Information Generation

The generation of the SI is equivalent to an error concealment process for missing blocks (WZ blocks) in the presence of their 4-neighbours (KEY blocks). We employ temporal and spatial error concealment methods (TEC/SEC) controlled by a mode selection algorithm as proposed in [11] for generating the SI. For TEC, each missing WZ block is divided into four sub-blocks. Each sub-block is concealed using two motion vectors coming from neighbouring KEY blocks – where motion estimation has previously been performed – and two motion vectors generated using an external boundary matching error (EBME) process, fused via a cosine weighted overlapping step. The SEC module uses bordering KEY pixels to conceal the WZ blocks through bilinear or directional interpolation depending on the directional entropy of neighbouring edges. The mode selection algorithm switches between TEC and SEC based on the levels of motion compensated activity and spatial activity in the neighbourhood of the processed block. If forward reference frames are available, bi-directional motion estimation is employed so that the replacement (SI) block can additionally result from averaging a forward and backward replacement block. The multiple prediction blocks resulting from this process form the basis for the multi-hypothesis SI coding, described below.

2.4 Multi-hypothesis Decoding

The WZ decoder uses the correlation noise to calculate bit probabilities. We estimate the correlation noise on a block basis using the 4-neighbouring KEY blocks of each WZ block. We model the noise as a Laplacian distribution with a specific variance that changes from block to block. The resulting distribution should follow closely that of the difference between the SI block and the transmitted WZ block as it employs actual received pixels in the vicinity of the processed block, as opposed to frame based interpolated values.

2.4.1 Log-Likelihood Ratio

Multiple SI is generated and defined as SI_i, $i \in \{1,..,N\}$, with increasing i denoting the ranking of the SI, from best to worst, according to the MSE of the predicted KEY group associated with each SI group. As multiple SI data are available, multiple bit probabilities can lead to a more precise log-likelihood ratio (LLR). Previously reconstructed WZ data based on available decoded bit planes are additionally used in the LLR computation expressed as:

$$LLR^l = \log \frac{\Pr\left[X^l = 0 \middle| SI_1, SI_2, ..., SI_N, Y^{l-1}\right]}{\Pr\left[X^l = 1 \middle| SI_1, SI_2, ..., SI_N, Y^{l-1}\right]} \quad (1)$$

where X^l is a bit value at bit-plane l and Y^{l-1} is a reconstructed WZ data Y using decoded bit-planes 1-(l-1). Note that the turbo decoder is sensitive to extreme LLR values, and we therefore average as opposed to summing the value of the multiple LLRs.

2.4.2 Multi-hypothesis WZ Reconstruction

The decoded bit-planes are converted to quantisation symbol $\widetilde{Q}$. The pixel Y is then reconstructed by firstly using the best SI based on its quantisation bins as follows:

$$Y_1 = \begin{cases} \Delta Q \cdot (\widetilde{Q}+1) & Q_{SI_1} < \widetilde{Q} \\ SI_1 & Q_{SI_1} = \widetilde{Q} \\ \Delta Q \cdot \widetilde{Q} & Q_{SI_1} > \widetilde{Q} \end{cases} \quad (2)$$

where ΔQ is a quantisation step size and Q_{SI} is the quantisation bin of the SI. Next, the pixels of SI_2 with $Q_{SI_2} = \widetilde{Q}$ are used to replace the clipped pixels (where $Q_{SI_1} \neq \widetilde{Q}$). The other SI is sequentially applied to the clipped pixels as follows.

$$Y_p = \begin{cases} SI_p & Q_{SI_p} = \widetilde{Q} \text{ and } Q_{SI_{p-1}} \neq \widetilde{Q} \\ Y_{p-1} & \text{otherwise} \end{cases} \quad (3)$$

where Y_p is the pixel value after the p^{th} SI has been received, $p \in \{2,..,N\}$. Thus the final refined version of the reconstructed WZ data ($Y=Y_N$) exhibits better image quality.

2.5 Selective Feedback

At each bit-plane, the decoder requests a set of parity bits from the encoder via a feedback channel if the estimated bit error rate (BER) is more than a threshold. To avoid sending unnecessary bits (e.g. for pixels with good SI), we propose two methods for introducing some granularity to the feedback process. We identify these two methods as simple and smart, according to the additional complexity that they introduce to the encoder. Note that in the transform domain this approach is not applied to the DC coefficients, and for the rest of the frequency bands a prior grouping according to quantisation level takes place.

2.5.1 Simple Encoder Approach

At the decoder, a feedback buffer is employed for collecting successive clipping maps for a number of consecutive frames which identify pixels with $Q_{SI} \neq \widetilde{Q}$ for each decoded bit-plane (12 frames are used herein). These WZ pixels are then segmented into (BP_{max}) groups, where BP_{max} is the total number of bit-planes used in WZ coding, with each group signifying the bit-plane where quantisation mismatch first took place. An additional group is formed for areas where $Q_{SI} = \widetilde{Q}$ for all bit-planes. The WZ pixels are then segmented on a 16x16 block basis according to these pixel groups. Parity bit request is then applied at the group level until the desired BER is reached, with priority given to most significant bit-plane (Fig. 2). A trade off between update frequency and segmentation accuracy can clearly be made.

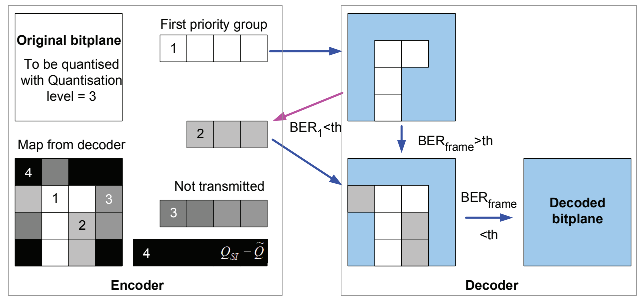

Fig. 2: Example of coding order of the simple encoder approach

2.5.2 Smart Encoder Approach

With this approach, grouping of the WZ pixels in fixed size blocks (e.g. 32x32) first takes place. Parity bit requests are then made for each block in raster scan order. Both encoder and decoder keep track of the number of parity bit requests made for each block, thus building a block-wise reliability map for each frame with blocks being ranked according to parity bit requests. This ranking is then employed for deciding the coding order of those blocks in the following frame, starting with the region with the largest number of requests. The block BER is calculated for deciding when to move to the next block in the coding order, and the frame BER for deciding when to stop requesting bits for the whole of the frame.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed codec was compared to the turbo-based Wyner-Ziv coding scheme of [6] (a frame-based SI interpolation scheme), to H.264 Intra coding as well as H.264 predictive coding with one motion vector and a limited or zero search range. Two standard CIF video sequences: *foreman* and *Breakdancers-cam3*, were coded with a GOP of 2 and 3. KEY frames and KEY slice groups were intra-coded with the quantisation parameters of 40, 36, 32 and 26 so that the quality of the WZ data at quantisation levels of 2, 4, 8 and 16 respectively was close to the quality of the KEY data. For transform-domain coding we employed the quantisation set introduced in [4]. A turbo encoder with two identical ½ rate constituent convolutional encoders was employed and the acceptable BER threshold was set to 10^{-3}

for each bit-plane. In the simulation, we used three SI data generated from the backward reference, the forward reference and the average of the two.

The rate-distortion performances for Foreman and Breakdancers are shown in Fig. 3 and Fig. 4 respectively. The bitrates shown in the plots are for all coded data, i.e. both KEY and WZ. The plots show that the proposed codec outperforms the frame-based DVC approach, particularly for the Breakdancers sequence that exhibits fast motion (up to 3 dB). The Gray code and multi-hypothesis coding can further reduce the WZ bitrate by up to 15% and improve the PSNR by approximately 1.5 dB. The performance of our DVC codec is comparable to Intra H.264 for Breakdancers, and superior for the case of Foreman by up to 1.5 dB. For the same sequence, the proposed DVC codec also outperforms H.264 Inter coding with zero motion (~0.8 dB at high bitrates).

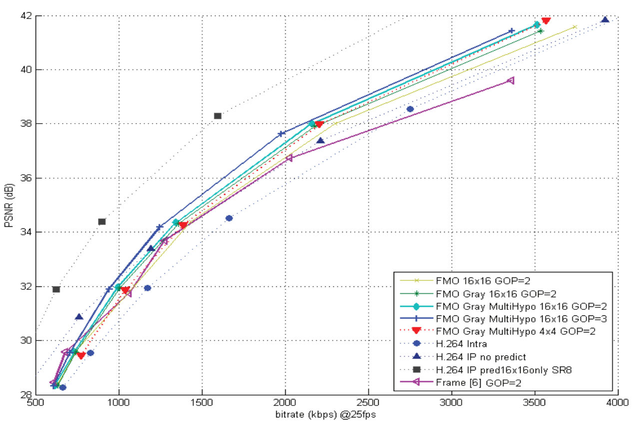

Fig. 3: Transform domain SPI-DVC performance of Foreman.

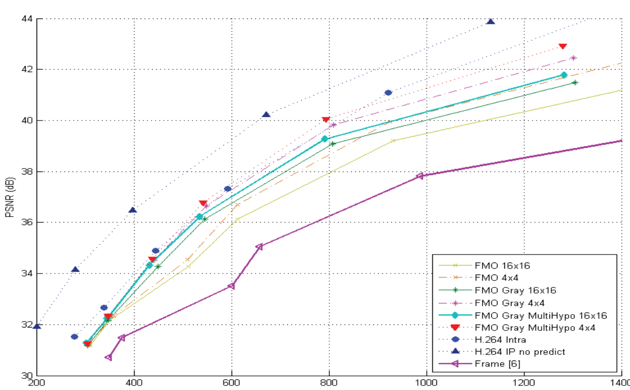

Fig. 4: Transform domain SPI-DVC performance of Breakdancers.

The performance with the two selective feedback methods is shown in Table 1, in the form of number of WZ bits requested for each bit-plane. The two proposed techniques, simple and smart, reduce the WZ bitrates, by 14.7% and 8.9% respectively. The simple encoder approach provides slightly better performance, at the cost of increased feedback overhead.

Table 1: Parity bits of Breakdancers (pixel-domain coding)

Bit-plane	without enhancement (kbits/frame)	with enhancement			
		Simple encoder		Smart encoder	
		kbits/frame	%	kbits/frame	%
1	6.28	4.89	22.1	5.19	17.3
2	14.55	12.41	14.7	13.32	8.4
3	19.44	17.07	12.2	18.17	6.5
total	40.27	34.37	14.7	36.68	8.9

4. CONCLUSIONS AND FUTURE WORK

This paper has presented the current state and performance of a DVC codec that is being developed by the authors. The proposed codec combines and proposes a number of ideas and algorithms that borrow from and contribute to the current state of the art in DVC coding. The performance achieved closes the gap to H.264 coding as shown in the results section, and surpasses that of other compared DVC schemes. This improvement can be traced to the use of spatial interleaving, which allows better SI generation and more accurate correlation noise estimation; the use of a Gray code and multi-hypothesis decoding; and the proposal of selective feedback mechanisms.

5. ACKNOWLEDGMENT

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

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

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

[5] N. Anantrasirichai, D. Agrafiotis and D. R. Bull, "A Concealment based Approach to Distributed Video Coding," accepted to ICIP2008.

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

[7] L. Natario, C. Brites, J. Ascenso and F. Pereira, "Extrapolating Side Information for Low-Delay Pixel-Domain Districuted Video Coding," *Visual Content Processing and Tepresentation*, vol. 3893, 2006, pp.16-21.

[8] Y.Tonomura, T. Nakachi, T. Fujii, "Efficient index assignment by improved bit probability estimation for parallel processing of distributed video coding," ICASSP, 2008, pp. 701-704.

[9] K.Misra, S. Karande, H. Radha, "Multi-Hypothesis Distributed Video Coding Using LDPC Codes," Proc. Allerton Conference on Communication, Control, And Computing, 2005.

[10] I.Richardson,"H.264 and MPEG-4 video compression-Video Coding for Next-generation Multimedia," John Wiley & Sons Ltd., 2003.

[11] D. Agrafiotis, D. R. Bull, N. Canagarajah, "Enhanced error concealment with mode selection", IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 8, pp. 960-973, 2006.