REDUCING CORRELATION NOISE IN WYNER-ZIV VIDEO CODING

Jeffrey J. Micallef, Reuben A. Farrugia and Carl J. Debono

Department of Communications and Computer Engineering,
University of Malta,
Msida, Malta.
jeffrey.micallef@ieee.org, reuben.farrugia@um.edu.mt, c.debono@ieee.org.

ABSTRACT

Over the past few decades, Distributed Video Coding (DVC) has been considered as a compression paradigm suitable for applications which require simple encoding. Yet, the performance obtained with practical architectures is still far from the theoretical bound, mainly due to the inaccurate Side Information (SI) predicted at the decoder. The work presented in this paper tries to improve the correlation between the bit-planes of the SI and the corresponding bit-planes of the Wyner-Ziv (WZ) frame. The proposed algorithm uses the adjacent key frames to predict the quantization intervals for the SI and for the WZ frame. It then chooses whether the quantization module should use the floor or the ceiling operation such that the indices generated for the two frames differ by the smallest number of bits possible. Experimental results show that, for a given target quality, the proposed method can reduce the WZ bit-rate by up to 8.4% compared to the traditional coding schemes.

***Index Terms*—**Adaptive quantization, correlation noise, distributed video coding, source representation, Wyner-ziv video coding

1. INTRODUCTION

With the emergence of mobile video capturing devices, compression of video is being performed with encoders having limited power and little processing capabilities. Yet, the traditional video coding paradigm invests in encoding schemes which explore the source statistics using very complex motion estimation tasks. This makes the encoder about 5 to 10 times more complex than the decoder [1], limiting the application of such schemes in mobile devices. Distributed Video Coding (DVC) is a coding paradigm which can potentially reduce the complexity of the encoder. DVC is based on Wyner-Ziv (WZ) theorem [2] which shows that two correlated sources, which are jointly Gaussian, such as two consecutive frames of a video sequence, can be encoded independently and jointly-decoded at the same coding efficiency of traditional schemes. This suggests that the computationally expensive tasks, that explore the source statistics at the encoder, can be shifted to the decoder. This facilitates the implementation of lightweight encoders, theoretically, without affecting the compression efficiency.

The authors in [3]-[5] have considered using different codeword representations to improve the correlation between the WZ frames and their Side Information (SI). On the other hand, our previous work in [6] predicted some of the discrepancies in SI, from the discrepancies present within the previously decoded bit-planes and corrected them to improve the predictions given by the SI. In [7], Jung and Karam transmitted only the syndrome bits associated with bit-planes having an estimated high Rate-Distortion (RD) performance, so as to avoid sending information which can lead to little improvement in quality. Later on, in [8] the same authors considered adapting the quantization interval according to the local RD characteristics. In contrast, the authors in [9] reduced quantization errors by adapting the quantization interval with the dynamic range of the transform coefficient bands. Non-uniform quantizers which are optimally designed to minimize quantization errors were then considered in [10]-[11].

This paper considers that some variation in the bin index can lead to a mismatch within multiple bit-planes. It thus presents a coding scheme where the encoder and the decoder alternate between the floor and the ceiling operation in order to seek the bin indices for WZ and SI that have the smallest hamming distance between them. This helps in reducing discrepancy between the bit-planes of WZ and those of SI, hence improving compression efficiency without affecting the video quality. Simulation results show that the proposed method can significantly reduce the WZ bitrates by up to 8.4%, compared to the traditional Instituto Superior Técnico (IST) Pixel-Domain (PD) architecture [12], with minimal increase in complexity at the encoder and decoder.

This paper continues by discussing the adopted Wyner-Ziv video coding architecture in Section II. Section III briefly reviews how variations in the bin index can cause mismatch within multiple bit-planes and Section IV introduces a method to select the most appropriate quantization function for each pixel location. Experimental results are then evaluated in Section V, whereas Section VI provides final comments and conclusions.

2. WYNER-ZIV VIDEO CODING FRAMEWORK

Fig. 1 illustrates the Wyner-Ziv video coding framework adopted in this work. It is based on the IST PD architecture in [12], where the incoming frame sequence is divided into key frames (odd frames) and WZ frames (even frames). The key frames are encoded using H.264/AVC intra coding techniques, whereas the intermediate WZ frames are first divided into four quadrants and then uniformly quantized into 2^M levels. The appropriate quantization process to be utilized for each pixel location is identified by considering the binary map Q, generated using the algorithm described in Section IV after considering a reconstruction version of the adjacent key frames *K'*. The quantized symbol stream q is then fed, bit-plane by bit-plane, into an LDPCA encoder [13] and the generated syndrome information is stored in a buffer to be transmitted incrementally to the decoder.

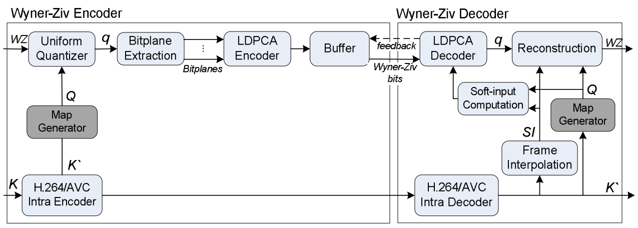


Figure 1. Proposed Wyner-Ziv video coding architecture.

For every WZ frame, the decoder decodes the forward and backward key frames and applies Motion Compensated Temporal Interpolation [12] (MCTI) to predict SI. At the same time, it also re-generates the same binary map Q used at the encoder, since this is required for the *soft-input computation module* and for the *reconstruction module* to maintain synchronization. The correlation noise between the WZ frame and the predicted SI is then modeled as a Laplacian distribution whose parameters are calculated, at pixel level, considering the difference between the forwards and backwards Motion Compensated frames [14]. The soft input values are calculated by summing up the relevant areas under the Laplacian model, as indicated by the previously decoded bit-planes [15]. These values are fed into the LDPCA decoder and syndrome information is requested from the encoder's buffer to correct the initial predictions given by the SI and recover the bit-planes of the WZ frame. All the decoded bit-planes are subsequently joined together to form the quantized symbol stream q and used together with SI to obtain the best reconstruction of the WZ frame.

3. PROBLEM OVERVIEW

Most of the indices of the WZ frame can be successfully predicted at the decoder by the SI, especially for areas of low motion such as the background. Only little parity information is therefore required to recover the bit-planes of the original WZ frame from those of the SI and hence compression is obtained. However, when the index value (*SI*) predicted by the side information differs even slightly from the index (*WZ*) of the WZ frame, the co-located indices may still differ by a number of bits, causing mismatch within multiple bit-planes. An illustrative example is shown in Fig. 2. When a *WZ* index of say 11 is being predicted with an SI of 12, it is noted that up to three different bit-planes are predicted incorrectly. Such variations in bin index are very probable and these may result in a poor rate-distortion performance

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BIT VALUES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Figure 2. Binary representation.

4. PROPOSED METHODOLOGY

This paper proposes a coding scheme which can limit the mismatch caused by such variations in bin index. The encoder and decoder predict the quantization interval for the WZ pixel and its SI (*WZ* and *SI*), and the quantization module will then choose a function, between the floor and the ceiling, which can assign indices that are likely to have the smallest number of bit changes. The appropriate quantization operation to be used for every pixel location is determined according to the following algorithm:

Step 1: The co-located pixel values of the adjacent key frames are used to predict the *SI* and *WZ* indices for every pixel location. Since neither the WZ frame, nor the SI, is available at both sides of the codec, to ensure that the same quantization function is chosen at every pixel location, the proposed algorithm considers the reconstructed version of the adjacent key frames *K'* to predict *SI* and *WZ*. The encoder can easily obtain a copy of the frames *K'*, present at the decoder, by performing an IDCT on the quantized DCT coefficients of the key frames. These frames are exactly the same at both sides of the codec and hence they can guarantee synchronization. Moreover, they are highly correlated to the side information and to the WZ frame, especially for areas of low motion.

These frames are therefore quantized, using the same number of levels used for the WZ frame, and the quantization indices of the forward and backward key frames are denoted as:

$$B_{2X-1} = \left\lfloor \frac{K_{2X-1}}{W_i} \right\rfloor \text{ and } B_{2X+1} = \left\lfloor \frac{K_{2X+1}}{W_i} \right\rfloor \quad (1)$$

where K_{2X-1} and K_{2X+1} represent the co-located pixel values in the adjacent key frames and W_i represents the width of the quantization intervals. The WZ frame is then divided into three regions which are processed separately as follows:

Region 1: considers the locations where the pixel values of the adjacent key frames fall within the same quantization interval (B_{2X-1}=B_{2X+1}). These regions are considered as areas of low motion, where correlation between *WZ* and *SI* is usually very high, i.e. if *SI* and *WZ* differ they will probably still fall in adjacent bins. Here, the prediction algorithm considers the distance between the pixel value and the edge of the quantization interval, which is determined using:

$$F_1 = K_{2X-1}\ MOD\ W_i \text{ and } F_3 = K_{2X+1}\ MOD\ W_i \qquad (2)$$

Then, if the average of F_1 and F_3 (F_{Avg}) is located at the top of the quantization interval as in Fig. 3(a), it is assumed that *SI* and *WZ* are found within bins B_{2X-1} and/or $B_{2X-1} + 1$, i.e. they are either 1) both within bin B_{2X-1} (as in Fig. 3(b)), 2) both within bin $B_{2X-1} + 1$ (as in Fig. 3(c)), or 3) one of them is within B_{2X-1} whereas the other is within $B_{2X-1} + 1$ (as in Fig. 3(d)-(e)). On the other hand, if the average F_{Avg} is nearer to the bottom of the quantization interval, as in Fig. 3(f), then it is likely that *SI* and *WZ* are within bins B_{2X-1} and/or $B_{2X-1} - 1$.

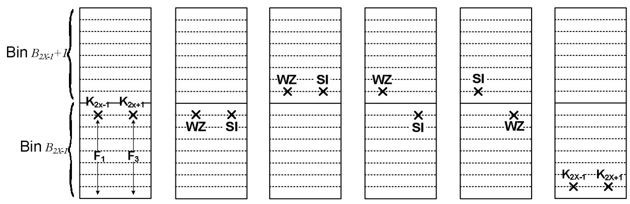


Figure 3: Predictions for WZ and SI corresponding with F_1 and F_3.

Region 2: Considers the locations where the pixels of the adjacent key frames are within adjacent bins (abs(B_{2X-1} – B_{2X+1})=1). Such regions can still represent areas of low motion and it is assumed that SI and WZ are found within the same quantization interval as those of the adjacent key frames, i.e. within B_{2X-1} and/or B_{2X+1}.

Region 3: Considers locations where the pixels of the two key frames are more than one quantization interval apart. These are considered as regions of high motion, where WZ and SI are less predictable and hence the proposed scheme uses the traditional (default) floor operation at all times.

Step 2: After predicting the quantization intervals for *WZ* and *SI,* within both *Region 1* and *Region 2*, the algorithm can then choose whether the floor function or the ceiling function generates the set of bin indices having the smallest hamming distance between them. This algorithm exploits the fact that these two functions will truncate the same pixel value towards opposite ends of the quantization interval, as shown in Fig. 4, to create a degree of freedom in the corresponding indices assigned to each bin. For every pixel location, the algorithm can therefore choose between two sets of successive indices (ex. between {*SI*=7, *WZ*=8} for the floor operation and {*SI*=8, *WZ*=9} for the ceiling operation), and tries to choose the set of indices having the smallest number of bit changes between them. The soft-input values are still calculated by modeling a Laplacian distribution around the original pixel value of SI. Yet, the new indices assigned for the intervals, which affects the regions summed up during the calculation of the bit-probabilities, helps reducing the amount of mismatch in areas where SI and WZ differ.

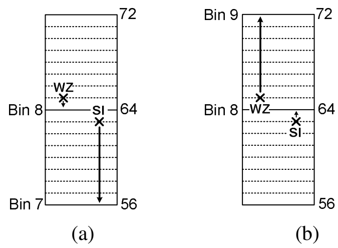


Figure 4. Quantization using (a) floor and (b) ceiling operation.

Given that both the encoder and decoder use a similar reconstructed version of the adjacent key frames, *K',* to determine the quantization operation to be used, the two sides of the CODEC are synchronized. The same indices are therefore generated when *SI* and *WZ* fall within the same bin, but for a very small difference in bin index, the proposed scheme can always find a set of indices which differ from each other only in the Least Significant Bit plane (LSB). This is illustrated in Fig. 2, where it is evident that every index has one of its adjacent values which differ from it by just one bit (the LSB). This improves significantly the correlation at the higher order bit-planes, whilst maintaining the discrepancies at the LSB unaffected. Less parity information is therefore required to correct the higher order bit-planes of SI, since the predicted bit-planes match better those of the original WZ frame. Furthermore, since both the floor and the ceiling operations consider the same quantization intervals, the quality of the resulting video is the same for either one of the two operations or a combination of both. Hence, such improvement in compression efficiency is obtained without affecting the resulting quality of the video compared to the traditional coding schemes.

The 57th frame of the *Hall monitor* sequence is considered as an illustrative example. Fig. 5(a) shows the discrepancies present between the 3rd bit-plane of the WZ frame and that predicted at the decoder using the sign of the soft input values. In this case, quantization is performed using just the floor operation. For the region marked by the red circle, the co-located pixels of the adjacent key frames, K_{2X-1} and K_{2X+1}, have values ranging from 126 to 129. Hence, when quantized at 32-levels these pixels are found at the top of bin 15 and/or at the bottom of bin 16 (falling within *Region 2*). It is therefore very probable that when *WZ* and *SI* differ, they change between a value of 15 and 16, incurring a mismatch within 5 bit-planes, including the 3rd bit-plane. In fact such change was observed for all the white pixels in the red circle (in Fig. 5(a)).

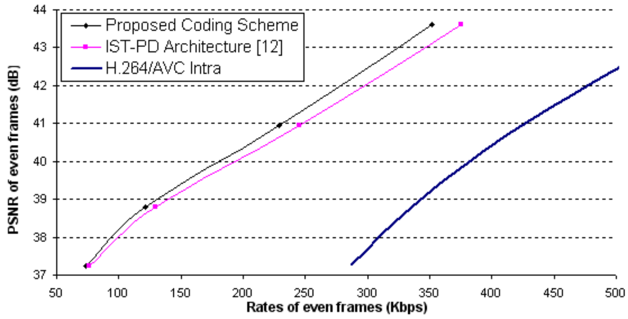
If the same pixel locations were quantized using a ceiling function, *WZ* and *SI* would probably change between 16 and 17 instead. This would incur only a mismatch within the LSB and avoid all the discrepancies at the higher order bit-planes. Fig. 5(b) illustrates the binary map *Q* generated for the considered video frame, marking the regions which should be quantized with the ceiling operation as white pixels. Meanwhile, Fig. 5(c) shows the discrepancies occurring at the 3rd bit-plane, when using the quantization operation indicated by map *Q*. It is shown that the amount of discrepancies were significantly reduced, including those which were present in the red circle. Discrepancy was also reduced in all the higher order bit-planes, decreasing the total bit-rate required to correct this particular video frame by up to 10% compared to the traditional schemes which consider just the floor operation.

(a) *(b)* *(c)*

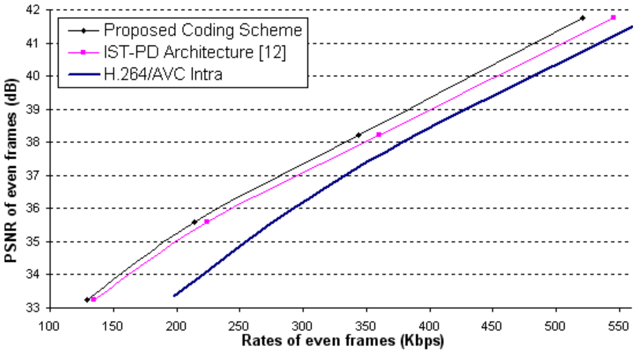
Figure 5. Discrepancies in SI (a) before and (c) after adopting the quantization process indicated by the map in (b).

4. EXPERIMENTAL RESULTS

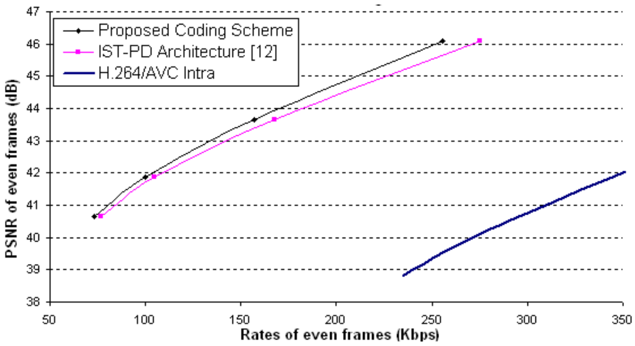
The Pixel-Domain WZ video codec described in Section II was used to compress all the frames of the *Hall monitor, Foreman*, *Mother & Daughter* and *Coastguard* sequences. The RD performances, obtained when considering the Luminance component over all the four quadrants of the WZ frames, are shown in Fig. 6 (a)-(d) respectively. All these video sequences were encoded at a WZ frame rate of 15 fps with a GOP size of 2 and have a QCIF resolution. The WZ frames were quantized at $2^M \in \{4, 8, 16, 32\}$ levels for QP_1 to QP_4 respectively, whist the corresponding key frames were encoded using H.264/AVC Intra coding (in Main Profile) and had their Quantization Parameter (QP) set as shown in Table I. These QPs were chosen such that the average quality of the WZ frames is similar to that of the key frames throughout the whole video sequence. For comparison purposes, the plots also include the performance obtained using the IST-PD architecture [12], where quantization is performed using just the floor operation, and the performance obtained when all the WZ frames are encoded using H.264/AVC Intra coding techniques.



(a) Rate-Distortion performance for the *Hall Monitor* sequence.



(b) Rate-Distortion performance for the *Foreman* sequence.



(c) Rate-Distortion performance for the *Mother & Daughter* sequence.

(d) Rate-Distortion performance for the *Coastguard* sequence.

Figure 6. Rate-Distortion performance for the various sequences.

TABLE I. QPs set for the Key frames.

	QP_1	QP_2	QP_3	QP_4
Hall Monitor	29	27	24	20
Foreman	33	29	26	22
Mother & Daughter	25	23	21	18
Coastguard	29	27	24	21

The results show that the indices assigned by the proposed coding scheme can improve the correlation between the bit-planes of the WZ frame and those of the corresponding SI. This reduced the WZ bit-rates required to correct SI and shifted the RD curves to the left for the same video quality. As expected, the rates required to correct the LSBs were left unaltered and all the gains were observed at the higher order bit-planes. Table II summarizes the overall percentage reductions in WZ bit-rates, obtained by the proposed coding scheme over the whole video sequences, when compared to the traditional coding scheme. From these results it is observed that the WZ bit-rate can be reduced by up to 8.4%, without affecting the resulting video quality.

TABLE II. Reduction in WZ bit-rates compared to IST PD architecture.

	QP_1	QP_2	QP_3	QP_4
Hall Monitor	4.8 %	6.3 %	8.1 %	8.4 %
Foreman	4.0 %	4.5 %	4.6 %	4.9 %
Mother & Daughter	4.6 %	5.1 %	6.8 %	7.2 %
Coastguard	4.2 %	4.5 %	5.0 %	5.3 %

5. CONCLUSION

This paper has provided an analysis of the variations in bin indices and showed how the traditional quantization scheme can results in a mismatch within multiple bit-planes, degrading compression efficiency. A novel scheme was then proposed, where the encoder and decoder distinguish between adopting a ceiling function or a floor function during quantization, to promote the use of indices whose binary representation changes by the least possible number of bits. Experimental results demonstrated that the proposed approach can reduce the WZ bit-rates by up to 8.4%, compared to the traditional schemes, with minimal increase in the encoder's complexity. Future work will consider the application of this coding scheme for transform domain architectures.

6. ACKNOWLEDGMENT

The research work disclosed in this publication is partially funded by the Strategic Educational Pathways Scholarship Scheme (Malta). The scholarship is part-financed by the European Union – European Social Fund. (ESF 1.25).

7. REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

[2] A. Wyner, and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transaction on Information Theory*, vol. 22, no. 1, pp. 1-10, January 1976.

[3] Z. He, L. Cao, and H. Cheng, "Correlation estimation and performance optimization for distributed image compression," in *Proc. SPIE Visual Communication and Image Processing Conference*, January 2006.

[4] Y. Tonomura, T. Nakachi, and T. Fujii, "Efficient index assignment by improved bit probability estimation for parallel processing of distributed video coding," in *Proc. IEEE Int. Conference Acoustic, Speech, Signal Processing*, April 2008.

[5] G. Huchet, Y. Chouinard, D. Wang, and A. Vincent, "Adaptive source representation for distributed video coding," in *Proc. of Int. Conference on Image Processing*, November 2009.

[6] J. J. Micallef, R. A. Farrugia, and C. J. Debono, "Improved Wyner-ziv video coding efficiency using bit plane prediction," in *Proc. of Int. Conference on Image Processing*, September 2011.

[7] C. W. Jung, and L. J. Karam, "BitpLAne SelecTive distributed video coding," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, October 2008.

[8] C. W. Jung, and L. J. Karam, "AQT-DVC: Adaptive Quantization for transform-domain distributed video coding," in *Proc. of Int. Conference on Image Processing*, September 2010.

[9] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain wyner-ziv video coding performance," in *Proc. of Int. Conference on Acoustics, Speech and Signal Processing*, May 2006.

[10] B. Wu, X. Guo, D. Zhao, and F. Wu, "An optimal non-uniform scalar quantizer for distributed video coding," in *Proc. of Int. Conference on Multimedia and Expo*, July 2006.

[11] W. A. R. Weerakkody, W. A. C. Fernando, M. B. Badem, and A. M. Kondoz, "Nonlinear quantisation for pixel domain distributed video coding," *IET Electronic Letter*, vol. 45, no.5, pp. 261-262, February 2009.

[12] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proc. of EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, June 2005.

[13] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Signal Processing Journal*, vol. 86, no. 11, November 2006.

[14] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Improved correlation noise statistics modeling in frame-based pixel domain Wyner-ziv video coding," in *Proc. of Very Low Bitrate Video Coding*, September 2005.

[15] S. Cheng, and Z. Xiong, "Successive refinement for the Wyner-ziv problem and layered code design," *IEEE Transaction on Signal Processing*, vol. 53, no. 8, pp. 3269-3281, August 2005.