Selective Reconstruction of Low Motion Regions in Distributed Video Coding

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

Department of Communication and Computer Engineering, University of Malta,
Msida, Malta

jeffrey.micallef@ieee.org
reuben.farrugia@um.edu.mt
c.debono@ieee.org

Abstract—**The Distributed Video Coding (DVC) paradigm offers lightweight encoding capabilities which are suitable for devices with limited computational resources. Moreover, DVC techniques can theoretically achieve the same coding efficiency as the traditional video coding schemes which employ more complex encoders. However, the performance of practical DVC architectures is still far from such theoretical bounds, mainly due to the inaccurate Side Information (SI) predicted at the decoder.**

The work presented in this paper shows that the soft-input values predicted at the decoder may not correctly predict the Wyner-Ziv coefficients, even for regions containing low motion. This generally degrades compression efficiency. To mitigate this, the proposed system predicts the quality of the SI for regions with low motion and then employs a technique which avoids correcting mismatch at locations where the SI and WZ falls within different quantization intervals but the prediction error is within an acceptable range. The experimental results demonstrate that the average Peak Signal-to-Noise Ratio (PSNR) is improved by up to 0.39dB compared to the state-of-the-art DVC architectures, like the DISCOVER codec.

Keywords: Adaptive quantization, correlation noise, distributed video coding, source representation, Wyner-Ziv coding.

I. INTRODUCTION

The traditional video coding schemes exploit complex motion estimation techniques to explore the source statistics at the encoder and achieve high compression efficiencies. This makes the encoder very computational intensive, about 10 times more complex than the decoder [1]. Although these schemes are suitable for services like video streaming and TV broadcasting, where the video is encoded once and decoded several times, they are inapt for new emerging mobile capturing devices or miniature cameras. These devices offer a limited power and little processing capabilities, making it a significant challenge to implement such coding techniques. Lighter encoding schemes are thus required, offering also benefits to other applications which require a large number of cameras, such as video surveillance and multi-view systems.

Over the past few decades Distributed Video Coding (DVC) has been considered as an alternative solution to reduce the complexity at the encoder. This was inspired from the Slepian-Wolf [2] and Wyner-Ziv (WZ) [3] theorems, which study the minimum rate required to encode two correlated sources independently. Provided that the two sources are jointly decoded, these theorems prove that the minimum rate required to encode the two sources independently is the same as that required when the sources are jointly encoded exploiting the correlation between them. This suggests that the computationally expensive tasks of motion estimation and compensation can be shifted to the decoder without compromising compression efficiency.

The temporal prediction in DVC is thus executed at the decoder using Motion Compensated Temporal Interpolation (MCTI) [4] between the adjacent key frames, and used as Side information (SI) to aid compression. The bit-planes of the original WZ frame are then transmitted to limit prediction errors. The correlation noise (*WZ-SI*) is treated as a virtual channel, allowing the bit-planes of WZ to be recovered efficiently using only a sub-set of parity information needed to correct the SI. The compression efficiency is thus heavily dependent on the quality of the side information.

The authors in [5]-[6] considered a uniform scalar quantizer, whose intervals can adapt with the characteristics of the transform coefficient band, to reduce quantization errors when reconstructing the WZ frame. Meanwhile, [7]-[8] used non-uniform quantizers which are optimally designed to minimize quantization noise. Jung and Karam [9] transmitted only the parity information required to correct the bit-planes which could lead to a high Rate-Distortion (R-D) performance. Furthermore, in [10], the same authors varied the quantization interval for different parts of the WZ frame to improve the local R-D characteristics. On the other hand, the authors in [11] anticipated some of the discrepancies in SI using the mismatch present within the previously decoded bit-planes, and correct them to improve the correlation between the WZ frames and their SI. Finally, [12]-[14] considered carefully selected codeword representations, such as gray codes, to improve the correlation between WZ frames and the SI.

This paper studies the effect of variations in light intensity, particularly for regions of low motion. It is shown that the transform coefficients which are found at the boundary of the quantization intervals may be affected by small variations in light, causing the SI to fall within the adjacent interval. In this case, the soft-input values predicted at the decoder are incorrect, degrading coding efficiency, even when the area
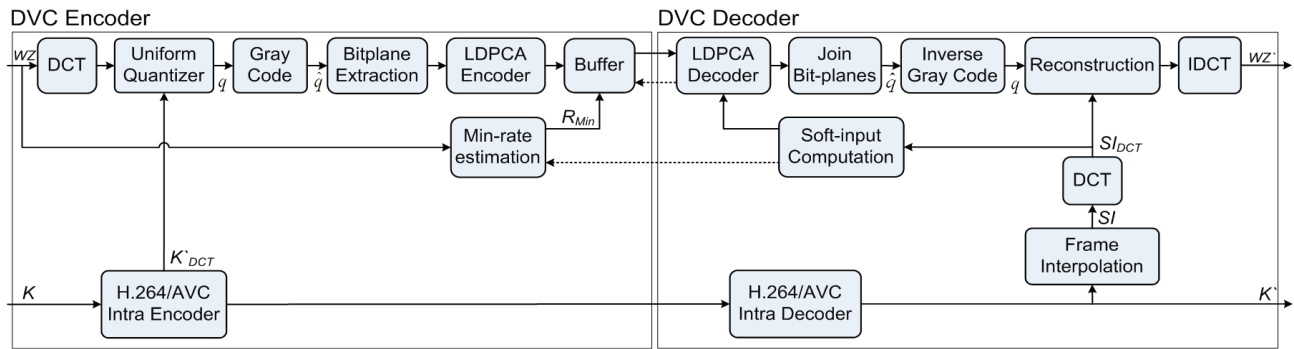
Fig. 1. Wyner-Ziv video coding architecture.

contains little or no motion at all. Such mismatch can be avoided by observing that the encoder can easily predict the coefficient values generated by the SI for areas of low motion. These predictions can then be used to prevent the decoder from correcting certain coefficient values which have a good SI quality, hence avoiding wasting bits to correct areas which can lead to small improvements in quality. Simulation results illustrate that the overall R-D performance can be improved by up to 0.39dB, compared to the traditional algorithms adopted by the state-of-the-art architectures like the DISCOVER Codec [15]-[16].

The next section introduces the Wyner-Ziv video coding architectures considered in this work. Section III illustrates how the compression efficiency is degraded by small variations in light intensity causing the SI to fall within the adjacent interval, even when the correlation noise is very small. Conversely, Section IV proposes a coding scheme to avoid such discrepancies, in areas of low motion. The experimental results are then presented in Section V, whilst Section VI provides the final comments and conclusion.

II. WYNER-ZIV VIDEO CODING FRAMEWORK

The transform domain Wyner-Ziv video coding architecture considered for this work is shown in Fig. 1. It is similar to the DISCOVER Codec in [16], namely that the Key frames (odd frames for a GOP of 2) are encoded using H.264/AVC Intra coding techniques, whilst the remaining WZ frames will first undergo a 4×4 block-based Discrete Cosine Transform (DCT). The resulting coefficients are then organized into bands of the same frequencies, and uniformly quantized into 2^M levels using the dead zone quantizer in [5]. However, opposed to the DISCOVER architecture [16], the *uniform quantizer* module considers the adjacent key frames to identify the blocks which contain low motion. It then tries to anticipate the SI coefficients predicted at the output of the decoder for such blocks. These predictions are used, as described in Section IV, to generate a new set of WZ symbols q, where the WZ coefficients are similar to the quantized SI coefficients for all regions where the discrepancy error is smaller than the quantization interval. The new quantized symbols q are then mapped into gray code representation $\hat{q}$

and their bit-planes are fed, one by one, into an LDPCA encoder whist the resulting syndrome bits are stored in a buffer. The *Min-rate estimation* module calculates the initiation number of parity bits R_{min}, to be transmitted promptly to the decoder, whilst the remaining parity bits are transmitted incrementally upon request through the feedback channel [17].

The decoder will first recover the adjacent key frames using H.264/AVC Intra decoder and subsequently used them to generate the Side Information *SI,* by applying MCTI techniques between these two frames [4]. The correlation noise between the WZ frame and the SI is modeled, at band level, using the difference between the forward and backwards motion compensated frames as in [18]. The resulting Laplacian parameters are then transmitted to the encoder to calculate R_{min}, and later used to estimate the soft-input information considering both the SI and the previously decoded bit-planes [19]. The soft-input values are fed into the LDPCA decoder [20], which request parity information from the encoder's buffer to correct the bit-planes of the SI. Once all the bit-planes are recovered successfully, these are joined together to form the quantized stream $\hat{q}$, and then mapped back into the original stream q, using inverse gray code mapping. The symbol stream q is used together with the SI to obtain the best reconstruction of the WZ DCT coefficients [21], which will finally undergo an Inverse DCT (IDCT) to reproduce the WZ frame in pixel domain.

III. EFFECTS OF VARIATIONS IN ILLUMINATION

Opposed to the traditional predictive video coding schemes, DVC architectures have to predict the SI at the decoder without knowing the original WZ frame. The predictions might therefore be incorrect, especially for regions of high motion which are difficult to predict using MCTI. The WZ coefficients are thus quantized and transmitted to the decoder, so that the SI values falling outside the intervals defined by the WZ coefficient can be truncated during reconstruction, limiting the prediction errors within a fixed range. These coefficients are compressed efficiently at the limit $H(WZ|SI)$, using SI as an initial prediction.

It is however observed that for some WZ coefficients which are found close the boundary of the quantization interval, a

small variation in light intensity can place the SI coefficient in the adjacent interval. For such locations, the decoder estimates poor soft-input values at the input of the LDPCA decoder, degrading the compression efficiency. The bit-probabilities are estimated by modeling a Laplacian distribution around the value of SI and later integrate the appropriate area under the graph as indicated by the previously decoded bit-planes. These are then used to calculate the Log Likelihood Ratio (LLR):

$$LLR = \log\left(\frac{P(x_i = 0)}{P(x_i = 1)}\right) \qquad (1)$$

where the sign value of LLR represents the bit-value predicted at the decoder, whilst the magnitude gives the confidence for such decision. Considering an SI coefficient value of 123, and assuming that the first two bit-planes are decoded as 1, the shaded regions in Fig. 2 are thus used to calculate the soft-input values for the 3rd bit-plane. Note that gray code representation has been adopted in this figure to improve correlation between the WZ and the SI. Yet, if the WZ coefficient has a value of 198, then the 3rd bit-plane is incorrectly assumed to have a bit-value of 1 ($P(x_2=1) > P(x_2=0)$) despite the low correlation noise. In general, an incorrect SI is always predicted for one of the bit-planes, whenever the SI and WZ fall within adjacent intervals due to variations in light intensity. Such mismatch forces the decoder to request a lot of parity information to correct regions where the prediction error is low, but this provides very little improvements in quality.
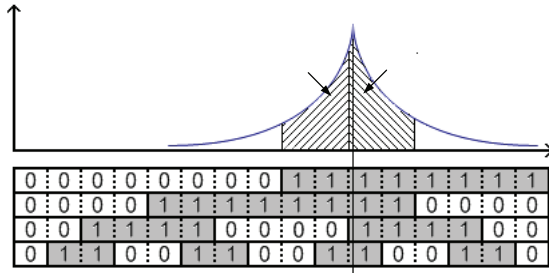

Fig. 2: Predicting the soft-input values for the 3rd bit-plane.

The correlation noise between the 15th frame of the *Hall Monitor* sequence, and its corresponding SI predicted at the decoder, is illustrated in Fig. 3(b). On the other hand, Fig. 3(c) shows the locations where the coefficients of WZ and SI fall within different quantization intervals, producing incorrect soft-input values. These results were obtained after applying a 4×4 DCT transform and quantizing both the WZ and SI coefficients using a dead-zone quantizer [5] with the number of levels set by the quantization matrix in Fig. 3(a). It is clear that the SI has a lot of mismatch in areas where the correlation noise is low.

(a) Quantization Matrix Q_7

(b) Correlation Noise (*WZ–SI*).
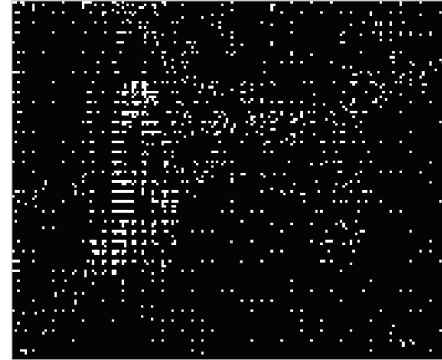
(c) Mismatch between the quantized WZ and *SI* coefficients.

Fig. 3. Correlation noise and resultant mismatch in SI for the 15th frame.

IV. EFFECTS OF VARIATIONS IN ILLUMINATION

The encoder can locate low motion regions using the adjacent key frames. The forwards and backwards frames are divided into 8×8 blocks and the quantizer computes the Mean Square Error (MSE) difference in (2), between the co-located blocks of the two frames. If the MSE is smaller that a threshold (TH$_1$ = 15), the block is considered to be a low motion block. Fig. 4 illustrates the low motion blocks using the original pixel intensities of SI, whilst the other blocks are depicted using a darker shade of grey. Clearly, all the blocks with an MSE smaller than TH$_1$ form part of the background and experience a low motion.

$$MSE = \sum_{i=1}^{8}\sum_{j=1}^{8}\frac{\left(K_{2x-1}(i,j) - K_{2x+1}(i,j)\right)^2}{64} \qquad (2)$$
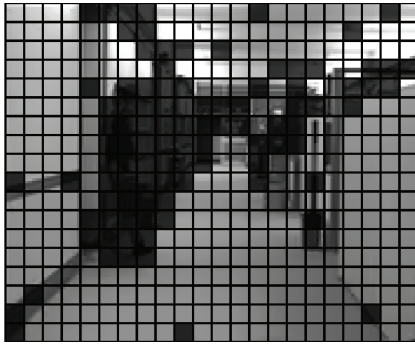
Fig. 4. Blocks where the MSE between adjacent key frames is smaller than the threshold TH_1.

The proposed method exploits the fact that, for regions of low motion which are usually affected by variations in light intensity, the decoder considers a Motion Vector (MV) of zero and works out the SI by considering the average of the co-located blocks in the forward and backwards key frames. This is especially true when adopting spatial motion smoothing filters [14] to reduce the number of false motion vectors. In fact, all the blocks with MSE smaller than TH_1 in Fig. 4 consider a zero MV at the decoder. For such regions of low motion, the encoder can thus use the adjacent key frames to anticipate how the SI coefficients are reconstructed at the decoder. These predictions are then used by the *uniform quantizer* to modify the quantized WZ symbol stream q, to avoid discrepancies at locations where the difference between the WZ and SI is smaller than the width of the quantization interval but the SI still falls within a different interval compared to that containing the WZ coefficient, as shown in Fig. 5. This is achieved by replacing the WZ coefficients at such locations with the quantized SI coefficients predicted at the encoder. This ensures that the decoder generates the correct soft-input information, hence reducing the amount of parity information needed to recover the new bit-plane of WZ, whilst knowing that the reconstructed coefficient is still within an acceptable range.
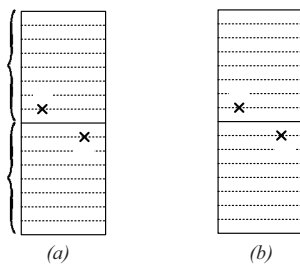

Fig. 5. *WZ* and *SI* coefficients are close to each other but they still fall within different bins.

Fig. 6 considers the discrepancies in SI when the low motion blocks (MSE < TH_1) are encoded using the proposed scheme, showing that most of the discrepancies in SI are now avoided. When reconstructing the WZ frame using the new coefficients, the resulting frame has a loss in quality of 0.09dB

compared to the frame reconstructed using the DISCOVER codec, where the original WZ coefficients are used at all times. Yet the new bit-planes of WZ can be recovered using 10% less parity information than that required to recover the original bit-planes of the WZ frame with the discrepancies shown in Fig. 3(c). This is one of the highest motion frames in the Hall Monitor sequence; better results are obtained for the other frames where a higher percentage of blocks consider a zero MV.
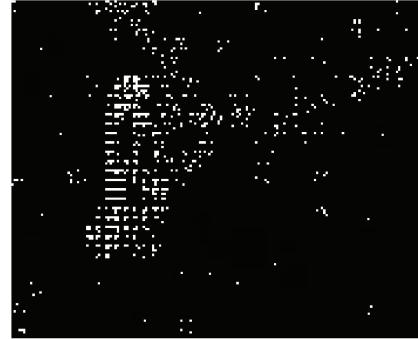

Fig. 6. New mismatch between the quantized WZ and *SI* coefficients.

(a) Reconstructed frame using the proposed technique.

(b) Reconstructed frame using DISCOVER Codec [16].

Fig. 7. WZ frame reconstructed using different schemes.

V. EXPERIMENTAL RESULTS

The *Hall Monitor* (15 Hz), *Foreman* (30 Hz), *Coastguard* (30 Hz) and the *Akiyo* (15 Hz) sequences, all having a QCIF resolution, were considered for Rate-Distortion (R-D) performance analysis. All the frames of these sequences were

compressed using the architecture described in Section II with a GOP size of 2, and the average R-D performance for the luminance component of the WZ frames and the key frames are illustrated in Fig. 9(a)-(d) respectively. For comparison purposes, the plots include the R-D performance obtained with the DISCOVER codec in [15]-[16] (considering Gray code representation). The DCT coefficient bands of the WZ frames were quantized using the number of levels defined by the 4×4 quantization matrices in Fig. 8, which were taken from [5] and [16]. A zero in the matrix means that the encoder transmits no information for that coefficient band and the reconstructed coefficients are copied directly from the SI at the decoder. Conversely, the corresponding key frames were Intra coded using H.264/AVC in Main profile, with the Quantization Parameters (QPs) set as shown in Table I. These QPs were chosen to ensure that the key frames and the WZ frames can maintain the same average quality throughout the whole sequence. The R-D performances obtained with H.264/AVC Intra coding and H.264/AVC No Motion techniques were also considered, since only these coding schemes share an important property in terms of low encoding capabilities.
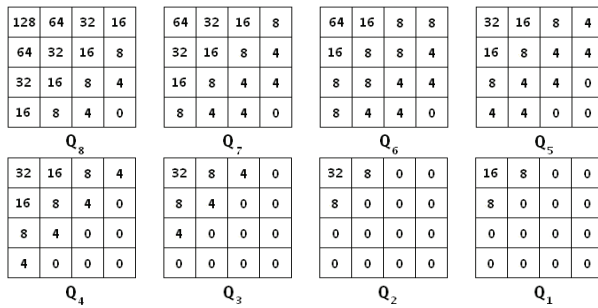

Fig. 8. 4×4 quantization matrices for different R-D points.

TABLE I: QUANTIZATION PARAMETERS USED FOR THE KEY FRAMES

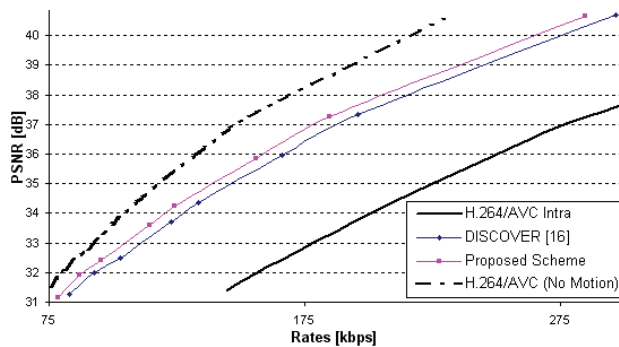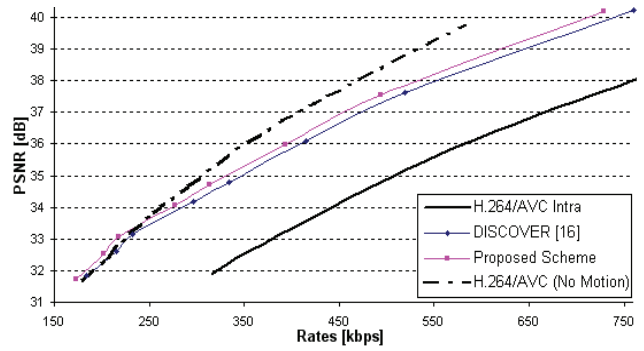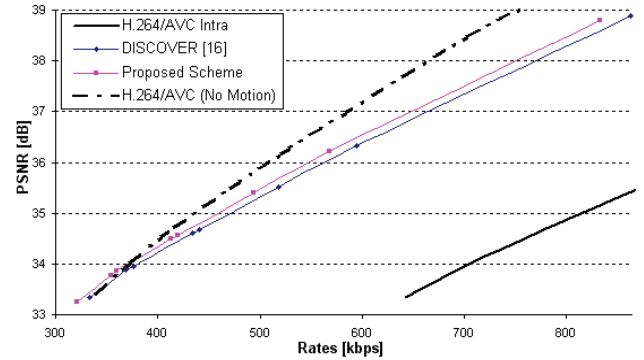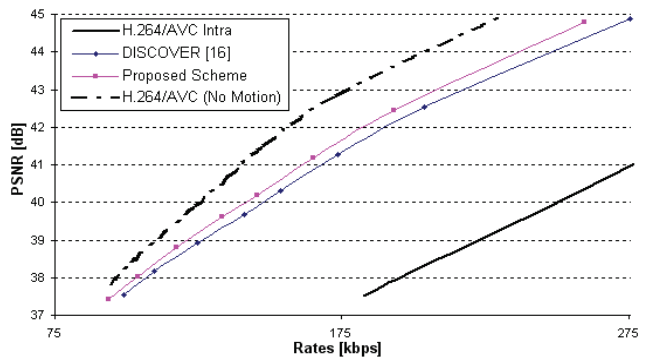|	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
Hall Monitor	37	36	35	34	33	31	29	24
Foreman	35	34	33	32	31	29	27	24
Akiyo	30	29	28	27	26	25	23	20
Coastguard	31	30	30	29	29	28	27	24

(a) R-D performance for *Hall Monitor* sequence.

(b) R-D performance for *Foreman* sequence.

(c) R-D performance for *Coastguard* sequence.

(d) R-D performance for *Akiyo* sequence.

Fig. 9. R-D performances for the tested sequences.

As illustrated in Fig. 9, the proposed scheme can result in an overall improvement in R-D performance relative to the traditional DISCOVER Codec [16]. The numerical average difference between the R-D curves was calculated using the Bjontegaard-Delta (BD) metric in [23], and used for comparative study. For the *Hall Monitor* and *Akiyo* sequences the average BD-PSNR (Peak Signal to Noise Ratio) was found to improve by 0.39dB and 0.32dB respectively, whereas the performances of the *Foreman* and *Coastguard* sequences were improved by 0.26dB and 0.21dB respectively.

The proposed algorithm offers smaller gains for high motion sequences, like the *Foreman* and *Coastguard* sequences, where some block experience high motion and cannot be predicted at the encoder. Better performance gains are observed for medium to low motion sequences, like the *Hall Monitor* and *Akiyo* sequences, since the majority of the blocks assume a zero MV and can be successfully predicted at the encoder to prevent discrepancies caused by small variations in light intensity. This scheme is thus more suitable for applications dealing with low motion videos, such as video surveillance systems, which is a main application of DVC.

VI. CONCLUSION

This paper showed that in areas of low motion the SI may still not predict the WZ coefficients correctly, even when the correlation noise is very small. This is usually due to small variations in light intensity affecting coefficient values which are close to the boundary of the quantization intervals. It is assumed that the decoder considers a zero MV for such areas of low motion and that the SI is estimated as the average of the adjacent key frames. The encoder can thus estimate the SI coefficients predicted at the decoder and use them to prevent situations where the correlation noise is very low but the WZ and SI coefficients are still within different quantization intervals. Experimental results show that the proposed methodology can improve the DB-PSNR by up to 0.39dB compared to the state-of-art architectures.

ACKNOWLEDGMENT

REFERENCES

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

[2] J. Slepian, and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471-480, July 1973.

[3] A. 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, pp. 1-10, January 1976.

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

[5] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,* May 2006.

[6] C. L. Xiao, L. M. Po, and W. H. Mo, "Adaptive quantization for distributed video coding," in *Proc. Data Compression Conference*, March 2011.

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

[8] 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 Letters*, vol. 45, no. 5, pp.261-262, February 2009.

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

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

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

[12] Z. He, L. Cao, and H. Cheng, "Correlation estimation and performance optimization for distributed image compression," in *Proc. SPIE*, vol. 6077, pp 1-8, January 2006.

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

[14] A. L. Becker, and K. M. Leung, "New counting codes for distributed video coding," *Computing Research Repository*, 2007.

[15] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, "The discover codec: architecture, techniques and evaluation," in *Proc. of Picture Coding Symposium*, November 2007.

[16] The Discover Codec [online]. Available: www.discoverdvc.org.

[17] D. Kubasov, K. Lajnef, and C. Guillemot, "A Hybrid encoder/Decoder Rate Control for a Wyner–Ziv video codec with a feedback channel," in *Proc. IEEE Multimedia Signal Processing Workshop*, October 2007.

[18] C. Brites, and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, September 2008.

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

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

[21] D. Kubasov, J. Nayak and C. Guillemot, "Optimal reconstruction in Wyner-Ziv video coding with multiple side information", in *Proc. International Workshop on Multimedia Signal Process*ing, October 2007.

[22] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 598-603, July 2003.

[23] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," in *VCEG Meeting*, April 2001.