Joint Decoding and Correlation Channel Parameter Estimation in Wyner-Ziv Video Coding

Yangli Wang[1,2], Amal Punchihewa[2], Chengke Wu[1]
[1]Xidian University, [2]Massey University
yliwang@mail.xidian.edu.cn, G.A.Punchihewa@Massey.ac.nz

***Abstract*—In Wyner-Ziv (WZ) video coding, or decoding of video with decoder side information (SI), the correlation between the source and its SI at the decoder side is modeled as a virtual correlation channel, and decoding of the quantized source requires the estimation of the correlation channel parameter. In this paper, we present a joint decoding and correlation channel parameter estimation approach. In the approach, starting from an initial correlation channel parameter, channel decoding and correlation channel parameter estimation are carried out iteratively. In each iteration, channel decoding of a bit plane of the quantized source is carried out first. The probability that a channel output bit is different from the corresponding one in the SI is then collected. The correlation channel parameter is numerically estimated afterwards by solving the equation between the probability collected and another probability calculated with the SI, the already decoded bit planes of the quantized source, and the correlation channel parameter. The process is repeated till a bit plane of the quantized source is correctly decoded. Simulation results show that the proposed approach obtains very promising results.**

Keywords-component; Wyner-Ziv video coding; side information; correlation channel; parameter estimation

I. INTRODUCTION

The emergence of new applications such as wireless video surveillance and video sensor networks arouses the research interests in low complexity video encoding. As one of such promising techniques, Wyner-Ziv (WZ) video coding, or lossy coding of video with decoder side information (SI), has attracted much research efforts in the past decade.

WZ video coding has its roots in Slepian-Wolf (SW) theorem [1] for lossless distributed source coding, and the WZ rate distortion function [2] for source coding with decoder SI. The SW theorem states that for lossless compression of two correlated information sources X and Y, the joint entropy $H(X,Y)$ can still be achieved even when the two sources are separately encoded, but are jointly decoded. And WZ rate distortion function states that, the rate distortion function of X, $R^*(D)$ when SI Y is only available at the decoder, and that of X, $R_{X|Y}(D)$ when SI Y is available at both the encoder and decoder satisfy $R^*(D) \geq R_{X|Y}(D)$, which indicates that there is generally a rate loss for scenarios where SI Y is not available at the encoder. But in certain cases, e.g., Gaussian sources with mean squared error distortion metric, the equality can be achieved. The two theorems suggest that by exploiting SI only at the decoder side, low complexity encoding, with a high complexity decoder is possible.

In [3], Wyner pointed out that SW coding is closely related to channel coding. Due to the lack of powerful channel codes at that time, practical SW coding and WZ coding was only possible after Pradham and Ramchandran published their work in 1999 [4], in which, practical SW and WZ coding is realized by scalar quantization followed by trellis codes based channel coding. After that, different WZ video coding approaches have emerged based on SW coding using more powerful channel codes such as turbo codes and low density parity check (LDPC) codes [5] [6]. For the background and development of WZ video coding, the readers are referred to [7][8][9], and also to [10] for more recent development in this field.

In a typical channel codes based WZ video codec, among other building blocks, such as SW coding, decoder SI estimation, and etc., correlation channel modeling and parameter estimation takes a very important role. In the literature, the most popular model for the correlation between source and its SI is the Laplacian model [5], in which, the probability density function of the residual between the source and its SI is described by a Laplacian distribution. Other models, for example, Laplacian mixture model [11], and exponential power distribution model [12] are also employed by some researchers. For correlation channel parameter estimation, the basic approach is offline estimation through training video sequences [5]. In [13], an online correlation channel parameter estimation method was proposed. In the method, the Laplacian distribution parameter is estimated by computing the variance of the frame difference between motion compensated backward and forward decoded frame which are also used for SI estimation. Intuitively, when the variance as computed is large, the Laplacian parameter should be small, otherwise, the Laplacian parameter should be big. However, there is no obvious analytical relation between them. There are also other online correlation channel parameter estimation methods [14][15], but they are only applicable for binary correlated sources. In this paper, we present a joint decoding and correlation channel parameter estimation approach. Starting from an initial parameter, channel decoding and correlation channel parameter estimation are carried out iteratively. In each iteration, channel decoding of a bit plane of the quantized source is carried out first. The probability that a

This work was supported by China Scholarship Council, and National Natural Science Foundation of China (Grant No. 60672117).

channel output bit is different from the corresponding one in the SI is then collected. The correlation channel parameter is numerically estimated afterwards by solving the equation between the probability collected and another probability calculated with the SI, the already decoded bit planes of the quantized source, and the correlation channel parameter. The process is repeated till a bit plane of the quantized source is correctly decoded. Simulation results justify the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 gives a brief description of the codec involved in the joint decoding and correlation channel parameter estimation process. Section 3 presents in detail the proposed joint decoding and correlation channel parameter estimation approach. Simulation Results are presented in Section 4. And, we conclude the paper in Section 5.

II. THE WZ VIDEO CODEC

The codec structure considered in this paper is illustrated in Figure 1. Its basic structure is based on [5]. The codec works in the pixel domain, so the basic components of the encoder include a quantizer plus a SW encoder, while the decoder is composed by a corresponding SW decoder, an SI estimation module, and a reconstruction module. Different from the structure of any other pixel domain WZ video codec, a correlation channel parameter estimation module, which is denoted by "CCP estimation" in Figure 1, is added at the decoder side for joint decoding and correlation channel parameter estimation. The SW decoder is composed of a turbo decoder, and The SW encoder is made up by a turbo encoder and a buffer. The turbo encoder is constructed by two recursive systematic convolutional encoders of rate $1/2$, which are parallel connected through an interleaver. Puncturing is used in the codec for turbo encoder to obtain higher code rate. In this paper, puncturing is carried out by a 2×16 puncture matrix.

When coding, the frames in an input video sequence is divided into key frames and WZ frames. All even numbered frames are taken as key frames, and all odd ones are treated as WZ frames. As a result, there are two key frames, one before and one after every WZ frame. This makes the coding of a WZ frame very like a B frame in conventional video coding. All key frames are intra encoded and decoded by an H.264 intra codec, so there is also an intra codec in the system, which is shown at the bottom part of Figure 1.

Each WZ frame is encoded independently of any other frame. For a particular WZ frame X, pixels are first quantized with a uniform quantizer with 2^M levels, where $M = 1,2, \ldots$, bit planes of the quantized pixel values in the frame are then extracted. Obviously, with the quantizer in the form of 2^M, the extraction of the most significant bit plane of the quantized pixel values equivalent to extraction of the most significant bit plane from the pixel values of the WZ frame X. Other bit planes follow the same argument. Each extracted bit plane is then channel encoded by the turbo encoder. The systematic bits of the turbo encoder output are discarded and only the parity bits of the turbo coder output are stored in a buffer. The parity bits are sent to the decoder in chunks upon decoder requests.

At the decoder, the SI for X, Y is first estimated with the two adjacent already decoded key frames of X using motion compensated temporal interpolation. Upon reception of a chunk of parity bits from the encoder, joint turbo decoding and correlation channel parameter estimation are then carried out. This process is repeated till a bit plane is correctly channel decoded. After all bit planes of the quantized pixel values are decoded, reconstruction of the frame is performed using the decoded quantized pixel values and the SI.

III. JOINT DECODING AND CORRELATION CHANNEL PARAMETER ESTIMATION

In WZ video coding, the correlation model for the source pixel values and those of its SI is assumed to be Laplacian with probability density function (PDF):

$$p(x-y) = \frac{\lambda}{2} e^{-\lambda|x-y|} \tag{1}$$

where x and y are respectively the realization of source video frame X and its SI Y, and λ is the correlation channel parameter which is to be estimated.

Given y and the already decoded bits of x, b_x^{i-1},b_x^{i-2},…,b_x^0 with $i = 0$ being the most significant bit, the probability that the $i-th$ bit of x, b_x^i, different from that of y, b_y^i can be calculated by

$$P\left(b_x^i \neq b_y^i \middle| b_x^{i-1}, b_x^{i-2}, \ldots, b_x^0, y\right) = \int_{L^i}^{U^i} \frac{\lambda}{2} e^{-\lambda|x-y|} dx \tag{2}$$

where L^i and U^i are respectively the lower and upper integral bound for the $i-th$ bit of x. Denote the quantized pixel value for x before the $i-th$ bit is decoded by $\hat{q}_x^i$, obviously, when

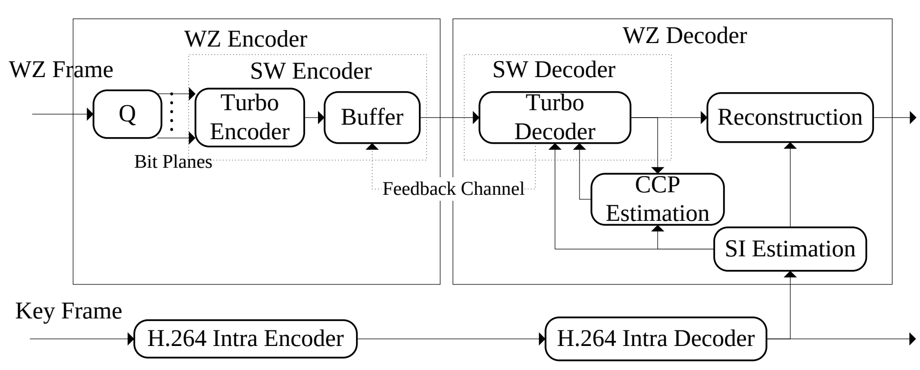

Figure 1. The structure of the WZ video codec

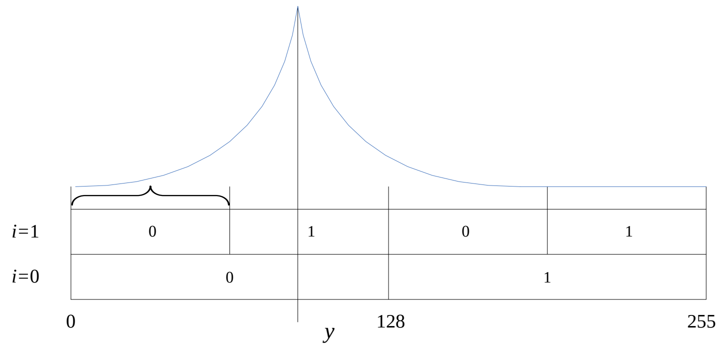

Figure 2. Determination of the region in which a bit of x and that of its SI y are different

$i = 0$, $\hat{q}_x^i = 0$, and define $s = 256/(1 \ll (i+1))$, where $\ll$ denotes left bit shift. Further denote the $i-th$ bit of y by b_y^i, then $L^i = (1 - b_y^i + \hat{q}_x^i) \times s$, and $U^i = (1 - b_y^i + \hat{q}_x^i + 1) \times s - 1$. For example, for $i = 0$, $s = 128$, if $b_y^i = 0$, then $L^i = 128$, $U^i = 255$, which is the region that the $0-th$ bit of x and that of y are different. The same rule applies for other bits of x. Figure 2 illustrates the determination of the region in which the $i-th$ bit of x and that of y are different given the already decoded bits of x and its SI y. In the figure, $i = 1$, $b_x^0 = 0$, and $y \in [64, 128)$, so the $1-th$ bit of y is 1, as a result, the region defined by L^1 and U^1 is $[0, 63]$, as shown in the figure by a brace. At the upper part of Figure 2, the distribution of the residual of x and its SI y is also shown.

The average probability that the $i-th$ bit plane of X and its SI Y are different is

$$P_{avg}^i = \frac{1}{N}\sum_{y \in Y}\int_{L^i}^{U^i}\frac{\lambda}{2}e^{-\lambda|x-y|}dx \equiv f(\lambda) \qquad (3)$$

where N is the frame size. The right hand of the above equation is a function of λ, and is denoted by $f(\lambda)$ in equation 3. When P_{avg}^i is known, the correlation channel parameter λ can be obtained by taking the inverse of f. As the analytical inverse of f is difficult to find, so numerical method is used to obtain the solution to λ.

The P_{avg}^i in equation 3 can be estimated after each run of turbo decoding, by

$$\hat{P}_{avg}^i = \frac{1}{N}\sum_{j=1}^{N}(\hat{b}_{x_j}^i - b_{y_j}^i) \qquad (4)$$

where $\hat{P}_{avg}^i$ is the estimated P_{avg}^i. x_j and y_j are respectively the $j-th$ pixel in X and Y. $\hat{b}_{x_j}^i$ is the turbo decoder output for the $i-th$ bit of x_j, and $b_{y_j}^i$ is the $i-th$ bit of y_j.

The detailed joint decoding and correlation channel parameter estimation algorithm can now be best described as follows.

Initialization: Before the running of turbo decoding, an initial correlation channel parameter λ_0 is specified. For the first bit plane of the first WZ frame in a video sequence, an arbitrary parameter, say for example 0.1, is specified, while for the first bit plane of any other WZ frame, the correlation channel parameter estimated for the previous WZ frame is used. For any other bit plane, the correlation channel parameter estimated for the previous bit plane is taken as the initial correlation channel parameter. Set iteration counter $k = 0$.

Main iteration: Increment k by 1 and perform the following steps:

- Request parity bits from the WZ encoder.
- Run turbo decoding using the correlation channel parameter λ_{k-1}, the already requested parity bits till now from the WZ encoder and the SI.
- Estimate P_{avg}^i with equation 4..
- Update correlation channel parameter λ_k by numerically solving equation 3.

Stopping Rule: If the bit plane is correctly decoded, stop. Otherwise, apply another iteration.

IV. SIMULATION RESULTS

The proposed joint decoding and correlation channel parameter estimation approach is implemented in C++ and is integrated into a pixel domain WZ video codec [16]. Its performance is evaluated using test video sequences. The first 101 frames of four QCIF sequences: Foreman, Hall Monitor, Carphone, and Stefan, with different motion activity, are used in the evaluation. The frame rate for the sequences is 30fps, and the resulting frame rate for WZ frames is about 15fps. The performance of the proposed approach is compared with a reference WZ video codec also working in the pixel domain. In the current implementation, the proposed joint decoding and correlation channel parameter estimation approach is only applied for the first bit plane of a WZ frame to reduce the computational load, while the correlation channel parameter estimated from the first bit plane of a WZ frame is applied for the turbo decoding of all other bit planes of the WZ frame. In the reference WZ video codec, the correlation channel parameter is estimated on a frame by frame basis at the decoder by maximum likelihood (ML) method, assuming that the source video sequence is also available at the decoder. This is not possible in practice, so the reference WZ video codec works in an ideal condition. As the results for each video sequence have similar trends, so we only present simulation results for the two test video sequences: Foreman and Stefan to save space.

Figure 3 and Figure 4 show the average PSNR versus bit rate for Foreman and Stefan, respectively. The various bit rates

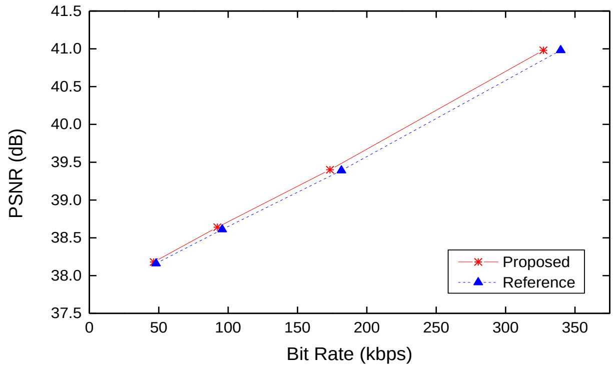

Figure 3. PSNR versus bit rate comparison, Foreman sequence

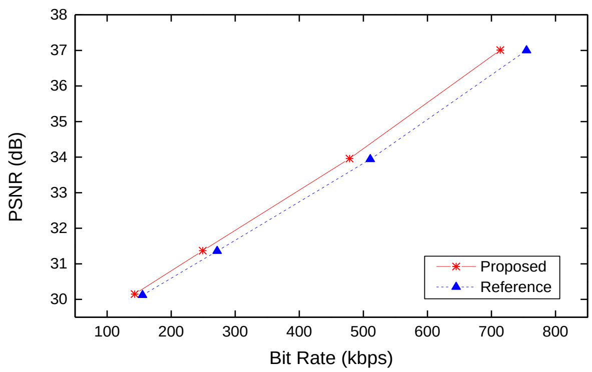

Figure 4. PSNR versus bit rate comparison, Stefan sequence

are obtained by varying M from 1 to 4. In each figure, the results obtained by the codec integrating the proposed joint decoding and correlation channel parameter estimation approach are denoted by "Proposed", while those obtained by the reference WZ video codec are denoted by "Reference". In the figures, PSNR and bit rate are counted only for the luminance component of the WZ frames.

From the figures, we can see that, the WZ video codec integrating the proposed approach achieves similar PSNR as the reference WZ video codec with similar or lower bit rate for all sequences. For example, for Foreman sequence, when the reference WZ video codec achieves 38.16dB at 47.99kbps, the WZ video codec integrating the proposed approach achieves 38.18dB at 46.34kbps. The WZ video codec integrating the proposed approach achieves a moderate bit rate saving of about 3.4%. Similarly, for Stefan sequence, when the reference WZ video codec achieves 30.12dB at 155.26kbps, the WZ video codec integrating the proposed approach achieves 30.15dB at 143.03kbps. In this case, the WZ video codec integrating the proposed approach achieves a bit rate saving of about 7.9%. The fact that the WZ video codec integrating the proposed approach achieves even lower bit rate than the reference WZ video codec seems somewhat unreasonable, however this can actually be explained by the fact that, for an un-ideal SW coder like turbo codes, the ML method gives too much confidence to the correlation between the source and its SI. As the correlation channel parameters used in the reference WZ video codec are not available in practice, so the results obtained by the WZ

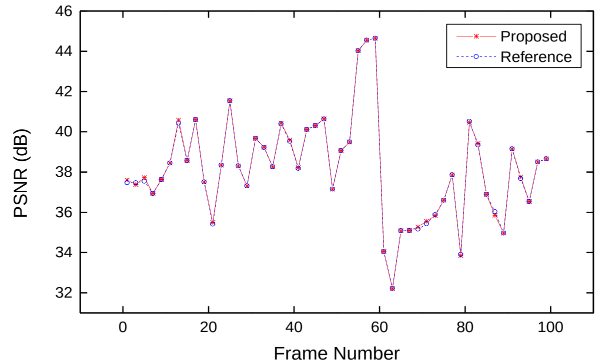

Figure 5. Frame by Frame PSNR comparison, Forman sequence with $M = 1$

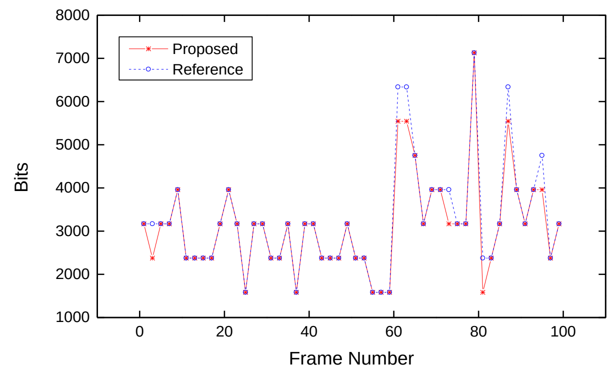

Figure 6. Frame by frame bits comparison, Foreman sequence with $M = 1$

video codec integrating the proposed approach are very promising.

To get a clearer image of the performance of the proposed approach, we show in Figure 5 and Figure 6 respectively the frame by frame performance comparison of the two WZ video codec for Foreman when $M = 1$. Figure 5 shows the frame by frame PSNR comparison of the two WZ video codec, while Figure 6 shows the frame by frame bit rate comparison. From the two figures, we can see, the WZ video codec integrating the proposed approach always achieves the same or better performance than the reference WZ video codec, and no performance loss is incurred by the joint decoding and correlation channel parameter estimation approach proposed in this paper. This also justifies the effectiveness of the proposed approach.

V. CONCLUSION

We have proposed a joint decoding and correlation channel parameter estimation approach in this paper. In the approach, starting from an initial correlation channel parameter, channel decoding and correlation channel parameter estimation are carried out iteratively. In each iteration, channel decoding of a bit plane of the quantized source is carried out first using the already requested parity bits from encoder, the SI, and the correlation channel parameter estimated in the previous run. The probability that a channel output bit is different from the corresponding one in the SI is then collected. The correlation channel parameter is numerically estimated afterwards by solving the equation between the probability collected and another probability calculated with the SI, the already decoded bit planes of the quantized source, and the correlation channel parameter. The process is repeated till a bit plane of the quantized source is correctly decoded. Simulation results on test video sequences show that the proposed approach obtains very promising results.

References

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inform. Theory, vol. 19, July 1973, pp. 471-480.

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, vol. 22, January 1976, pp. 1-10.

[3] A. Wyner, "Recent results in the Shannon theory," IEEE Transaction on Information Theory, vol. 20, no.1, Jan. 1974, pp. 2-10.

[4] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in Proc. IEEE Data Compression Conference, Snowbird, Utah, Mar. 1999, pp.158-167.

[5] A. Aaron, R. Zhang, B. Girod, "Wyner-Ziv coding for motion video," in Proc. Asilomar Conference on Signals, Systems and Computers, Pacific Grove, U. S., November 2002.

[6] D. Varodayan, A. Aaron and B. Girod, "Rate-adaptive distributed source coding using Low-Density Parity-Check codes," Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, Nov. 2005.

[7] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," Proceedings of IEEE, vol. 93, no. 1, January 2005, pp. 71-83.

[8] Z. Xiong, A. D. Liveris, S. Cheng, "Distributed source coding for sensor networks," IEEE Signal Processing Magazine, vol. 21, no. 5, May 2004, pp. 80-94.

[9] R. Puri, A. Majumdar, P. Ishwar, and K. Ramchandran, "Distributed video coding in wireless sensor networks," IEEE Signal Processing Magazine, vol. 23, no.7, July 2006, pp. 94-106.

[10] C. Guillemot, et al, "Distributed monoview and multiview video coding: Basics, problems and recent advances," IEEE Signal Processing Magazine, vol. 24, no. 9, Sept. 2007, pp. 67-76.

[11] R. P. Westerlaken, R. K. Gunnewiek, R. L. Lagendijk, et al, "The role of virtual channel in distributed source coding of video," in Proc. 2005 IEEE International Conference on Image Processing, ICIP2005, pp. I-581-584.

[12] T. maguey, J. Gauthier, et al., "Using an exponential power model for Wyner-Ziv video coding," in Proc. 2010 IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP2010, pp. 2338-2341.

[13] C. Brites, F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," IEEE Trans. Circuits and Systems for Video Technology, vol. 18 no. 9, pp. 1177-1190.

[14] J. Garcia-Frias, Y. Zhao, "Compression of correlated binary sources using turbo codes," IEEE Communications Letters, vol. 5, no. 10, pp. 417-419.

[15] V. Toto-Zarasoa, A. Roumy, C. Guillemot, "Maximum likelihood BSC parameter estimation for the Slepian-Wolf problem," IEEE Communications Letters, vol. 15, no. 2, pp. 232-234.

[16] Y. Wang, J. Jeong, C. Wu, S. Xiao, "Wyner-Ziv Video coding with spatial-temporal side information," in Proc. 2007 IEEE International Conference on Multimedia and Expo, ICME 2007, pp. 132-135.