SPATIAL NON-STATIONARY CORRELATION NOISE MODELING FOR WYNER-ZIV ERROR RESILIENCE VIDEO CODING

Yongsheng Zhang, Hongkai Xiong, Li Song, Songyu Yu

Dept. of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, P.R. China

ABSTRACT

Most of the Wyner-Ziv (WZ) video coding schemes in literature model the correlation noise (CN) between original frame and side information (SI) by a given distribution whose parameters are estimated in an offline process. In this paper, an online CN modeling algorithm is proposed towards a more practical WZ-based error resilient video coding (WZ-ERVC). In ERVC scenario, the side-information is typically generated from the error concealed picture instead of bi-directional motion prediction. The proposed online CN modeling algorithm achieves the so-called classification gain by exploiting the spatially non-stationary characteristics of the motion field and texture. The CN between the source and error concealed SI is modeled by a Laplacian mixture model, where each mixture component represents the statistical distribution of prediction residuals and the mixing coefficients portray the motion vectors estimation error. Experimental results demonstrate significant performance gains both in rate and distortion *versus* the conventional Laplacian model.

***Index Terms*—** Wyner-Ziv coding, spatial non-stationary, correlation noise modeling

1. INTRODUCTION

In hybrid video coding schemes, motion estimation and motion compensation (ME/MC) is adopted to exploit temporal redundancy between successive frames. Although ME/MC achieves high compression efficiency, transmitting the encoded video may suffer from error propagations and lead to the well-known drifting phenomenon. Recently, inspired by the natural error resilient property of Wyner-Ziv (WZ) coding, WZ-based error resilient video coding (WZ-ERVC) schemes have been proposed in literature [1]. The schemes integrate a joint source-channel coding framework for video, and show significant RD performance gains over conventional error resilient schemes, e.g., Intra refresh (IR) and forward error correction (FEC) [2].

The coding efficiency of WZ coding depends critically on the capability to model the correlation noise (CN) between the original frame and side information (SI) [3]. Traditionally, the CN is modeled by a given distribution whose parameters are offline estimated by assuming that both the source data and corresponding side-information are available at the encoder side or the decoder side. It is undesirable since: (1) the encoder cannot have the error pattern in transmission, especially for applications with large end-to-end delay; (2) the decoder side cannot have the source data, otherwise transmission errors would be perfectly eliminated. To solve the problem, Brites et. al. proposed an online CN modeling algorithm for conventional WZ video coding, based on bi-directional ME/MC SI generation [4].

For ERVC applications, this paper proposes an online correlation noise modeling algorithm where SI is typically generated from the error concealed picture. In fact, the prediction residual has been de-correlated with the original frame, so the major task of error concealment resort to estimating the lost motion vectors according to the coherence property of the motion field and the spatial smoothness of neighboring blocks. The CN between the source and error concealed SI is modeled by a Laplacian mixture model, where each mixture component represents the statistical distribution of prediction residuals and the mixing coefficients portray the motion vectors estimation error. Essentially, the proposed CN model describes the spatially non-stationary characteristics and achieves the so-called classification gain [5].

The rest of this paper is organized as follows: Section 2 provides a brief summary of WZ-ERVC framework. The proposed online CN modeling algorithm is studied in Section 3. Experimental results validate the efficiency of the proposed algorithm in Section 4. Section 5 concludes this paper.

2. THE WZ-BASED ERROR RESILIENT VIDEO CODING

A general architecture of WZ-ERVC is shown in Fig. 1. X at the encoder side denotes the current frame through the conventional predictive encoder, e.g., MPEG or H.26x engine. To eliminate temporal error propagation, the waveform of some P-frames is protected by WZ coding. At the decoder side, the MPEG/H.26x bit-stream is firstly decoded. If current frame is not contaminated by transmission error, the error concealment (EC) module and WZ decoding module will be skipped. Otherwise, the EC module would be

* The work has been partially supported by the NSFC grants No. 60632040, No. 60772099 and the National High Technology Research and Development Program of China (863 Program) (No. 2006AA01Z322).

activated to approximate error-free X by exploiting the temporal and spatial correlation of video. Error concealed frame X' would serve as the SI for WZ decoding of X. And then, the WZ decoding module analyzes the CN between X and X'. With the estimated CN, WZ decoder would decode the WZ bit-stream and outputs the final reconstruction X''. Theoretically, X'' is a better description of X than X' since the WZ decoder can at least correct some of the transmission errors by parity-check codes. Then, X'' updates X' in the frame buffer of the decoder. The following parts are dedicated to investigating the CN model by correctly decoded information.

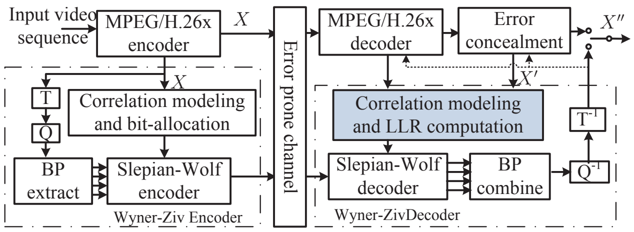


Fig. 1: The diagram of WZ-based error resilient video coding

3. THE PROPOSED CORREALTION NOISE MODELING ALGORITHM

A packet loss in transmission damages the decoded frame in two folds: it loses both the motion (motion vector, MV) and the texture (prediction residual) information of the frame. The task of inter-frame EC is to estimate MV of the lost blocks because MVs of neighboring blocks are highly correlated while the prediction residual is unpredictable [6]. In the proposed algorithm, the transmission error is modeled by two parts: error caused by MV estimation error, denoted as N_c^ρ; and error caused by the loss of prediction residual, denoted as N_c^w.

The transmission error which is regarded as the CN in WZ coding setup, is here modeled by a Laplacian mixture distribution which is composed of a series of Laplacian distributions with mean μ_k, i.e.

$$p(c|c_s) = \sum_k u_k e^{-\alpha_w |c-\mu_k|} \tag{1}$$

where c is a possible value of the DCT coefficient of the original frame and c_s is the corresponding SI generated by EC; α_w is the parameter of a mixture component, which represents the energy of prediction residual; μ_k is the coefficient of the mixture model, which represents MV estimation error and k is the index of mixing coefficient μ_k; u_k is normalized to ensure $\sum p(c|c_s) = 1$. Since the actual MVs of the lost slice are unavailable at the decoder side, both c and μ_k could be regarded as random variables for the decoder.

For the mixture model, to estimate parameters μ_k and α_w is a major concern of the next Section. Because EC is generally implemented in pixel domain, we first analyze the transmission-induced error and then derive the DCT domain CN for WZ decoding.

3.1 Pixel domain error caused by MV loss

The i-th pixel of frame X_n can be written as

$$\begin{aligned} x_n(i) &= x_n^\rho(i) + w_n(i) \\ &= f_{MC}\big(x_{n-1}(i), v_n(i)\big) + w_n(i) \\ &= \sum_{l=1}^{L} a_l x_{n-1}\Big(u_l\big(i + v_n(i)\big)\Big) + w_n(i), \end{aligned} \tag{2}$$

where $x_n^\rho(i)$ is the motion-compensated (MC) value from previous frame X_{n-1}, and $w_n(i)$ is the prediction residual. $v_n(i)$ is the MV of the block in which $x_n(i)$ is located. The motion-compensated interpolation coefficient a_l satisfies $\sum_l a_l = 1$, and $u_l(i)$ denotes the spatial index of the l-th pixel in frame X_{n-1} which is used to predict $x_n(i)$.

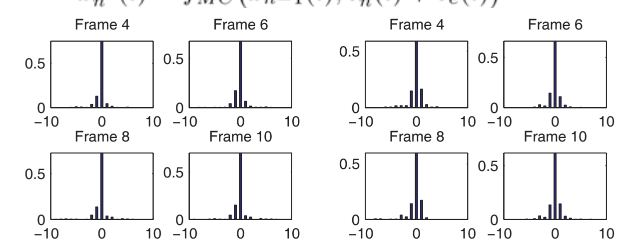
When transmission erasure occurs, the EC operation approximates $x_n(i)$ by

$$\begin{aligned} x_n^{\rho'}(i) &= f_{MC}\big(x_{n-1}(i), v_n'(i)\big) \\ &= \sum_{l=1}^{L} a_l x_{n-1}\Big(u_l\big(i + v_n'(i)\big)\Big), \end{aligned} \tag{3}$$

where $v_n'(i)$ is the estimated value of $v_n(i)$. The mismatch is denoted as $v_e(i) = v_n(i) - v_n'(i)$. If $v_e(i)$ is available at the decoder side, we can obtain the error-free $x_n^\rho(i)$ with $f_{MC}\big(x_{n-1}(i), v_n'(i) + v_e(i)\big)$. By this means, the probability of $x_n^\rho(i)$ can be derived from the probability of the corresponding $v_e(i)$, i.e.,

$$p\big(x_n^\rho(i) = x_n^{\rho''}(i)\big) = p(v_e(i)). \tag{4}$$

where $x_n^{\rho''}(i) = f_{MC}\big(x_{n-1}(i), v_n'(i) + v_e(i)\big)$.



(a) Mobile (b) Flower

Fig. 2: The distribution of v_e along successive frames of sampled sequences.

Fig. 2 shows the distribution of v_e along successive frames of sampled "*Mobile*" and "*Flower*" sequences, where the scale interval of the horizontal axis is integer pixel. It shows that the PDF of $v_e(i)$ along successive frames have a similar distribution. It infers that we can estimate the probable distribution of $v_e(i)$ for frame X_n by simulating EC operation in frame X_{n-1}.

After the PDF of $v_e(i)$ is estimated, the possible value of $x_n^\rho(i)$ can be obtained from Eq. (4). Fig. 3 presents the possible value of $x_n^\rho(i)$ for a set of pixels of "*Mobile*" sequence, where the blue "o" denotes the error-free pixel value $x_n(i)$, the black "$\times$" stands for the EC result $x_n^{\rho'}(i)$, and the green "$+$" is the possible value of $x_n^{\rho}(i)$ estimated by Eq. (4).

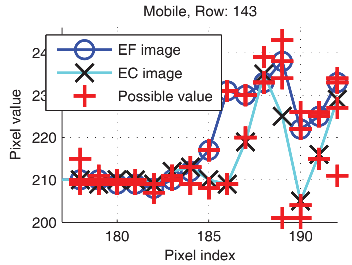


Fig. 3: The estimated possible value of $x_n^{\rho}(i)$

3.2 DCT domain error caused by MV loss

Because the DCT exploits the spatial correlation of source, most WZ video coding schemes are implemented in DCT domain. Here, we also analyze the error on DCT coefficients induced by MV loss.

The DCT is a linear transform and the DCT coefficient c_m could be expressed as a linear combination of pixels in a M-by-M block:

$$c_m = \sum_{i \in \mathcal{B}} b_l x(i), \tag{5}$$

where i is the pixel index in the M-by-M block $\mathcal{B}$ and b_l is transform coefficient. The possible value of DCT coefficient c_m could be obtained at the decoder as

$$p(c_m = c'_m) = \prod_{i \in \mathcal{B}} p(x_n^{\rho''}(i)), \tag{6}$$

where $c'_m = \sum_{i \in \mathcal{B}} b_l x_n^{\rho''}(i)$ is the DCT coefficient value, and $x_n^{\rho''}(i)$ is the possible pixel value which is estimated as Section 3.1.

3.3 Prediction residual loss

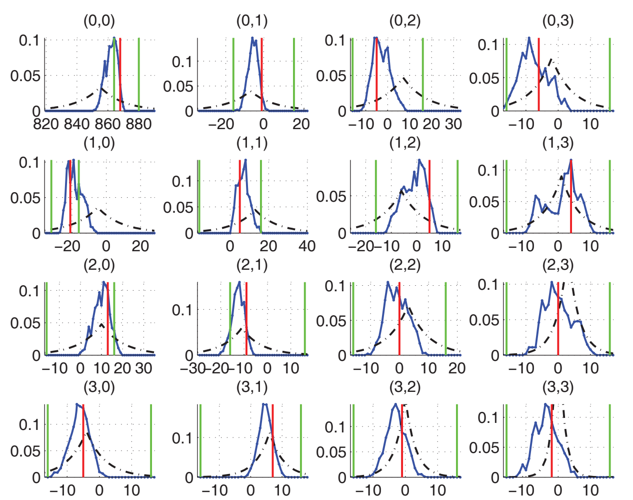


Fig. 4: The estimated PDF $p(c'_m)$ for each coefficient band.

In general, the prediction residual can be modeled by a zero mean additive Laplacian noise $p(w) = \frac{\alpha_w}{2} e^{-\alpha|w|}$, where

$$\alpha_w = \sqrt{2/\sigma_w^2}, \tag{7}$$

and σ_w^2 is the variance of the prediction residual [4]. Although the prediction residual of the lost block is unavailable at the decoder side, its variance σ_w^2 could be approximated from its correctly decoded neighbors.

Finally, the probability distribution of DCT coefficient c_m at the decoder side could be modeled by a mixture Laplacian noise as Eq. (1), where α_w can be obtained from Eq. (7), and $\mu_k = f_{MC}(x_{n-1}(i), v'_n(i) + v_e(i))$ could be calculated by Eq. (3) and (4).

Fig. 4 gives the estimated PDF of each DCT coefficient band. The title "(h,v)" of each subplot indicates the horizontal and vertical coefficient index, respectively (e.g., (0,0) is the DC coefficient). The blue line is the estimated $p(c'_m)$ with the proposed Laplacian mixture model, while the black dash-and-dot line represents that with conventional Laplace model. The vertical red line indicates the actual error-free value c_m, and the green vertical lines located on both sides of the vertical red line indicate the quantization interval Δ. Δ is set to 16 in Fig. 4.

A larger probability $p(c'_m \in \Delta)$ requires fewer bits for the decoder to correct the same amount of transmission error. It signifies that the proposed modeling algorithm outperforms the conventional Laplace model in a rate-distortion sense. A MMSE inverse quantizer reconstructs the coefficient c_m by the centroid of $p(c'_m)$ where $c'_m \in \Delta$. A superior quality of the reconstructed picture could hence be observed.

4. EXPERIMENTAL RESULTS

To evaluate the proposed CN modeling algorithm, four CIF (352×288) sequences @15Hz are considered: "*Flower*", "*Foreman*", "*Mobile*" and "*Stefan*". In experiments, JM12.2 with *Baseline* profile is adopted as the conventional video coder, and only the luminance component is considered for the RD performance evaluation. The transmission error is randomly inserted into both the standard coded bit-stream and the WZ bit-stream. The video is encoded with chess-board FMO pattern, and the extended boundary matching error concealment algorithm [7] in the reference software is enabled. LDPCA [8] is used to produce the WZ bit-stream, and MMSE inverse quantizer is used to minimize the mean square error (MSE) distortion of the reconstructed frames.

As in Section 3.3, the proposed CN modeling algorithm benefits the RD performance: to reduce both the encoding rate and the distortion of reconstructed frames. Table I gives the significant bit-rate saving of the proposed model *versus* the conventional Laplace model. The parameters in the proposed model are estimated as in Section 3, and the parameter of the Laplacian model is obtained with offline training. The term "QPW" denotes the quantization parameter of WZ video coding. As the setup in H.264, the quantization step increase by a factor of two for every increment of six in QP.

Table II shows that the PSNR gain decreases with the decrease of QPW. It could be observed that the PSNR gain of the proposed model is significant at low bit-rate because the distortion of the reconstructed frame highly depends on the accurate CN model. With the increase of bit-rate, the inaccurate effect of the CN modeling could be gradually mitigated with the more previous decoded bit-planes.

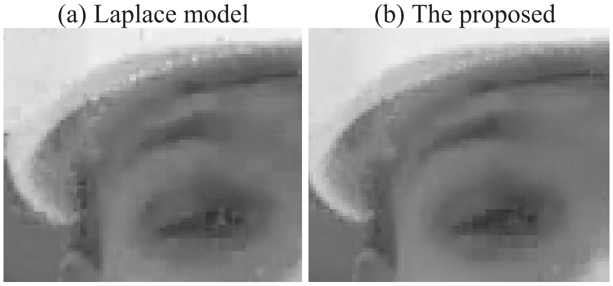
Table I: Rate saving with the proposed model

QPW		38	34	31	28	25	22
Δ_{Rate} (Kbps)	Foreman	-66	-96	-146	-146	-200	-147
	Flower	-185	-138	-171	-187	-354	-162
	Mobile	-158	-167	-245	-291	-431	-315
	Stefan	-78	-120	-182	-206	-321	-360

Table II: PSNR improvement with the proposed model

QPW		38	34	31	28	25	22
Δ_{PSNR} (dB)	Foreman	1.50	1.18	0.95	0.66	0.42	0.26
	Flower	1.70	1.36	1.07	0.77	0.46	0.27
	Mobile	1.84	1.46	1.11	0.70	0.45	0.26
	Stefan	1.40	1.16	0.93	0.68	0.43	0.27

(a) Laplace model (b) The proposed



(c) Laplace model (d) The proposed

Fig. 5: The subjective quality comparison

Finally, the subjective quality of "*Mobile*" and "*Foreman*" sequences are contrasted in Fig. 5 (a)(b) and (c)(d), respectively. The results are obtained with the coding rate of 840kbps for "*Foreman*" and 2, 077kbps for "*Mobile*". It could be obviously seen that the reconstructed picture with the proposed CN model achieves superior subjective quality than the conventional Laplacian model.

5. CONCLUSIONS

This paper addresses a key issue in WZ video coding: online CN modeling at the decoder side. For ERVC applications where SI is generated from the error concealed picture, the proposed algorithm models the transmission error by a Laplacian mixture model which is dynamically consistent with the motion field and the spatial smoothness. In the Laplacian mixture model between the source and error concealed SI, each mixture component represents the statistical distribution of prediction residuals and the mixing coefficients portray the motion vectors estimation error. Essentially, the proposed CN model describes the spatially non-stationary characteristics and achieves the so-called classification gain. Experiments demonstrate significant performance gain in both rate and PSNR *versus* the conventional Laplace model. In our future work, an advanced model exploiting correlations between neighboring pixels in a spatial domain will be investigated.

6. REFERENCES

[1] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-Ziv coding of video: An error-resilient compression framework," IEEE Transactions on Multimedia, vol. 6, pp. 249-258, Apr. 2004.

[2] R. Bernardini, M. Fumagalli, M. Naccari, et al., "Error concealment using a DVC approach for video streaming applications," EURASIP Signal Processing Conference, Poznan, Poland, 2007.

[3] *R. P. Westerlaken*, *R. K. Gunnewiek*, and *R. I. L. Lagendijk*, "The role of the virtual channel in distributed source coding of video," IEEE International Conference on Image Processing, 2005.

[4] C. Brites and F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner–Ziv Video Coding," IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, pp. 1177-1190, 2008.

[5] X. Li, "On the Importance of Source Classification in Wyner-Ziv Video Coding," Internationl Conference on Video Communation and Image Processing, 2008.

[6] I. E. G. Richardson, *H.264 and MPEG-4 video compression*: John Wiley & Sons Ltd, 2003.

[7] W.-M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," IEEE International Conference on Acoustics, Speech, and Signal Processing, USA, pp. 417-420, 1993.

[8] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," Eurasip Signal Processing, vol. 86, pp. 3123-3130, Nov. 2006.