NOISE MODELING FOR EFFICIENT TRANSFORM DOMAIN WYNER–ZIV VIDEO CODING

Sachidananda Samal

Corresponding author: snsamal29@gmail.com

Abstract: In recent years, practical Wyner–Ziv (WZ) video coding solutions have been proposed with promising results. Most of the solutions available in the literature model the correlation noise (CN) between the original frame and its estimation made at the decoder, which is the so-called side information (SI), by a given distribution whose relevant parameters are estimated using an offline process, assuming that the SI is available at the encoder or the originals are available at the decoder. The major goal of this paper is to propose a more realistic WZ video coding approach by performing online estimation of the CN model parameters at the decoder, for transform domain WZ video codecs. In this context, several new techniques are proposed based on metrics which explore the temporal correlation between frames with different levels of granularity. For transform-domain WZ (TDWZ) video coding, DCT bands and coefficients are the two granularity levels proposed. The higher the estimation granularity is, the better the rate-distortion performance is since the deeper the adaptation of the decoding process is to the video statistical characteristics, which means that coefficient levels are the best performing for TDWZ solutions.

Keywords: Correlation model, distributed video coding (DVC), online estimation, pixel domain, transform domain, Wyner–Ziv (WZ) video coding.

1. Introduction

In typical video coding applications, e.g., broadcasting or video streaming, the video codec relies on the powerful hybrid block-based motion compensation and DCT transform architecture which was primarily driven by the one-to-many model with a single complex encoder and multiple light decoders. The complexity burden of the encoder (which is typically 5–10 times higher than the decoder [1]) is mainly associated with the motion estimation task, which is primarily responsible for the high rate-distortion (RD) performance achieved. However, this architecture is being challenged by several emerging applications such as wireless video surveillance, multimedia sensor networks, wireless PC cameras, and mobile camera phones. These applications have different requirements from those targeted by more traditional video delivery systems, e.g., in wireless video surveillance systems, low-cost encoders, or codecs, allowing a flexible allocation of complexity between the encoder and decoder are important since there is a high number of encoders and only one or few decoders. Distributed video coding (DVC) fits well these emerging scenarios since it enables the exploitation of the video statistics, partially or totally, at the decoder only and allows to make the encoder more or less intelligent and thus complex. A flexible allocation of complexity between the encoder and the decoder is therefore enabled by the DVC paradigm. From the Information Theory, the Slepian–Wolf Theorem [2] states that it is possible to compress two statistically dependent discrete random sequences and that are independently and identically distributed in a distributed way (separate encoding and joint decoding) using a rate similar to that used in a system where the sequences are encoded and decoded together, i.e., like in traditional video coding schemes. The extension of Slepian–Wolf coding for lossy compression is well known as Wyner–Ziv(WZ) coding [3], which deals with the lossy source coding of when some side information is available only at the decoder. In [3], Wyner and Ziv show that there is no increase in the transmission rate if the statistical dependency between and is only explored at the decoder compared with the case where it is explored both at the decoder and the encoder, notably, if and are jointly Gaussian and a mean-square error distortion measure is considered. The side information is usually interpreted as an attempt made by the decoder to estimate the original frame to be WZ encoded. In the WZ coding scenario, error correcting codes are typically used to improve the quality of the side information until a target quality for the final decoded frame is achieved. Two of the most interesting DVC approaches are the pixel- and transform-domain turbo coding-based WZ video coding schemes presented in [4], where the decoder is responsible for exploring all (or most of) the source statistics and, therefore, to achieve compression following the WZ coding paradigm; the schemes presented in [4] make use of a feedback channel to perform rate control at the decoder. While the pixel-domain (PD) codec is simpler (in terms of complexity), the trans- form-domain (TD) codec provides better RD performance (at the cost of a slightly higher coding complexity.

2. WZ Video Codecs

Fig. 1 illustrates the architecture of the TDWZ video codec proposed in [11] and that will be used in this paper; this codec is an improved TDWZ video coding solution which basically follows the same architecture as the one proposed by Aaron et al. in [7]. However, the solution adopted here brings some major advances regarding the coding solution in [7], namely in the DCT transform, the quantizer, and the frame interpolation (FI) modules. The main differences are in the key frames codec, the quantizer, the turbo decoder, the frame interpolation, and the correlation noise modeling modules. • The key frames are Intra coded with a H.264/AVC codec since it is a very efficient coding solution. • In terms of AC coefficients quantization, it is proposed here to have quantized bins evenly distributed around zero in- stead of an unbalanced AC quantization approach used so far, e.g., as in [11]. By doing this, one bin of the current quantization approach is suppressed and the central (zero) bin is doubled in size, maintaining the remaining bins size. Since most of the AC coefficients are concentrated around zero, by doubling the zero bin size, the matching probability between corresponding quantized bits of the WZ and SI frames increases bringing bit rate savings. Some distortion loss is, however, expected since the bigger the quantization bin, the worst the decoded frame quality is, but overall the RD performance improves. •

3. Correlation Noise Model in WZ Video Coding

To make good usage of the SI obtained through the FI framework, the decoder needs to have a reliable knowledge of the model that characterizes the correlation noise between the original WZ frame and the corresponding SI frame. The correlation noise can be interpreted as a virtual channel with an error pattern characterized by some statistical distribution (or model) since SI may be seen as a "corrupted" version of the original information. In the TD scenario, to the residual between corresponding DCT bands of the WZ and SI frames. If the model accurately describes , the coding efficiency may be higher; however, if the model fails, a coding efficiency loss will be observed. In the context of Fig. 1, this would correspond to less accurate data at the input of the turbo decoder (SISO de- coders) which would make the turbo decoder to spend more parity bits to correct the same amount of errors; the data at the SISO decoders input/output is called soft information, i.e., confidence information (e.g., probabilities) from which it is possible to make a decision about an event.

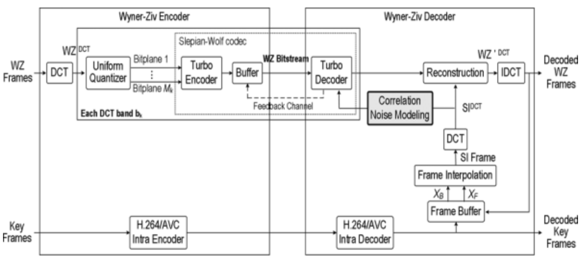

Fig. 1

$\mathrm{p}[\mathrm{WZ}(\mathrm{x},\mathrm{y}) - \mathrm{SI}(\mathrm{x},\mathrm{y})] = \frac{\alpha}{2}\exp[\alpha|WZ(x,y) - SI(x,y)|]$ (1)

to model the residual between the original WZ frame and the corresponding SI frame. In (1), is the probability density function, is the position to be evaluated within the WZ and SI frames, and is the Laplacian distribution parameter is defined as $\alpha = \sqrt{\frac{2}{\sigma^2}}$ (2)

can vary along time (e.g., different values for each frame) and space(e.g., different values for each pixel within a frame), as will be shown in Sections IV–VII. In (2), is the variance of the residual.

4. Offline Transform-Domains Correlation Noise Modeling

The TDWZ video codec exploits the spatial redundancy within a frame by applying a DCT transform over the frame blocks; therefore, instead of pixel values, DCT transform coif- cents are quantized and turbo coded Since DCT coefficients are turbo coded, the noise distribution to be taken into account regards the residual between corresponding DCT bands of the WZ and the corresponding SI frames. Based on previous work, the authors use a Laplacian distribution as in [1] to model the statistical correlation between corresponding DCT bands of the original WZ and the corresponding SI frames. Following the same coarse to fine strategy as adopted for the PDWZ codec to offline calculate the parameter, three granularity modeling levels are proposed in the following: DCT band/sequence level, DCT band/frame level and coefficient/frame level.

4.1 Correlation Noise Model at DCT Band/Frame Level

Step-1 Compute first the residual frame between the WZ frame and the corresponding SI frame as

$$R(x, y) = WZ(x, y) - SI(x, y) \qquad (3)$$

Step-2 Apply a 4 x4 block-based discrete cosine transform over the residual frame to obtain the DCT coefficients frame T

$$T(u, v) = DCT[R(x, y)]. \qquad (4)$$

Step-3 Compute the R frame variance. We know, the variance of a random variable Z can be computed using

$$\sigma_z = E[z^2] - (E[z])^2 \qquad (5)$$

Where E [.] is the expectation operator. So for our case the T frame DCT band b coefficient variance computation,

$E[Z^2]$ and $E[Z]$ corresponds $E_b[T_b^2]$ and $E_b[T_b]$ respectively.

$$E_b[T_b] = \frac{1}{J}\sum_{j=1}^{J} T_b(j)]$$

$$E_b[T_b^2] = \frac{1}{J} \sum_{j=1}^{J} T_b(j)]$$

T_b represents all of the DCT coefficients of T frame band b and J stands for the DCT band size; the DCT band size is given by the ratio between the frame size and the number of different DCT coefficients bands.

Step -4 Compute the DCT-band b variance for a certain WZ frame via

$$\sigma_b^2 = E_b[T_b^2] - (E_b[T_b])^2 \qquad (6)$$

Step -5 Compute the DCT band αvalue using $\alpha_b = \sqrt{\frac{2}{\sigma_b^2}}$ (7)

Despite this computation process being more efficient (in terms of RD performance) than the one described in Section V-A, due to the exploitation of the (varying) temporal correlation, the offline computation procedure can be even more efficient by also exploring the (varying) spatial correlation

4.2 Correlation Noise Model at Coefficient/Frame Level

Step 1 Compute first the residual frame between the WZ frame and the corresponding SI frame as

$$R(x, y) = WZ(x, y) - SI(x, y) \qquad (8)$$

Step 2 Apply a 4 x4 block-based discrete cosine transform over the residual frame to obtain the DCT coefficients frame T

$$T(u, v) = DCT[R(x, y)]. \qquad (9)$$

5. Online Transform

Domain Correlation Noise Modeling

It was used a Laplacian distribution to model the correlation noise in the TDWZ video coding context where the Laplacian distribution parameter $\boldsymbol{\alpha}$ is estimated offline at the encoder and at the DCT band level, i.e., each DCT band has a (constant) value associated. This offline process is not acceptable and realistic because it requires the encoder to replicate the side information. Following an approach similar to the one proposed for the PDWZ video codec, two novel online Laplacian distribution parameter estimation techniques, which work at the decoder at two different granularity levels are proposed for the TDWZ video codec: the DCT band/frame level and the coefficient/frame level. In the DCT band/frame-level approach, one value is estimated for each DCT band.

5.2 Correlation Noise Estimation at DCT Band /Frame Level

Step 1 - Residual frame generation: Compute first the residual frame between R the motion compensated versions of the framesX_B and $X_F$as follows:

$$R(x,y) = \frac{X_F(x+dx_f\ ,y+dy_f)-X_B(x+dx_b\ ,y+dy_b)}{2} \qquad (10)$$

where $X_B(x + dx_f\ , y + dy_f) and X_F(x + dx_b\ , y + dy_b)$ and represent the backward and the forward motion- compensated frames, respectively, and (x,y) corresponds to the pixel location in the frame R. (dx_f, dy_f) and$(dx_{b,} dy_b)$and represent the motion vectors for the and framesX_F and X_B, respectively

Step 2 - R frame k^{th} block variance computation: Compute the residual frame th block variance using

$$\sigma_{R_k}{}^2 = E_{R_k}[R_K(x,y)^2] - \left(E_{R_k}[R_k(x,y)]\right)^2 \qquad (11)$$

Step 3 |T| frame generation: Compute the |T| frame whose elements are the absolute value of the corresponding elements in the |T| frame

Step 4 - |T| frame DCT band variance computation, :Compute the DCT band variance using equation (6)

Step 5 - DCT band parameter estimation, : Estimate the parameter for each DCT band using equation (7).

5.3 Correlation Noise Estimation at Coefficient/Frame Level

In the coefficient/frame-level approach, an online adaptation of the Laplacian distribution parameter both temporally, i.e., along the video sequence, and spatially, i.e., for each DCT coefficient inside the DCT coefficients frame, is performed. Basically, the DCT coefficient is classified into one of two classes: (1) inlier coefficient and (2) outlier coefficient: the inlier coefficients are those whose value is close to the corresponding DCT band average value and the outlier coefficients are those whose value is far away from . In order to determine how close a certain coefficient is to the corresponding DCT band average value, it is proposed here to compare the distance between that coefficient and with the DCT band variance because the variance is a measure of how spread the coefficient values are regarding its average value. The coefficient/frame level correlation noise estimation technique is described in the following

Step -1 Residual frame generation: Compute first the residual frame between R the motion compensated versions of the framesX_B and $X_F$as follows:

$$R(x,y) = \frac{X_F(x+dx_f\ ,y+dy_f)-X_B(x+dx_b\ ,y+dy_b)}{2} \qquad (12)$$

Where $X_B(x + dx_f\ , y + dy_f) and X_F(x + dx_b\ , y + dy_b)$ represent the backward and the forward motion-compensated frames, respectively, and (x,y) corresponds to the pixel location in frame R.

(dx_f,dy_f) and (dx_b,dy_b) represents the motion vector for the frames X_F and X_B respectively.

Step-2 R frame kth block variance computation: Compute the residual frame kth block variance using

$$\sigma_{R_k}{}^2 = E_{R_k}[R_K(x,y)^2] - \left(E_{R_k}[R_k(x,y)]\right)^2 \qquad [13]$$

Step-3 |T| frame generation: Compute the |T| frame whose elements are the absolute value of the corresponding elements in the |T| frame

Step-4 |T| frame DCT band variance computation, :Compute the DCT band variance using (6)

Step 5 |T| frame *(u,v)*DCT coefficient distance computation: Compute, for the|T| frame DCT band b , the distance D(*u*,*v*) between the coefficient and the |T| frame DCT band b average value μ, , using(26) where represents the DCT coefficient at the position of the frame DCT band .

D(*u*,*v*) = |T| (u,v)- μ,

Where |T| (u,v)- represents the DCT coefficient at the *(u,v)*position of the frame |T| DCT band b

Step -6 DCT coefficient parameter estimation, Estimate the parameter for the DCT coefficient .

6. Performance Analysis

RD PERFORMANCE FOR OFFLINE MODEL

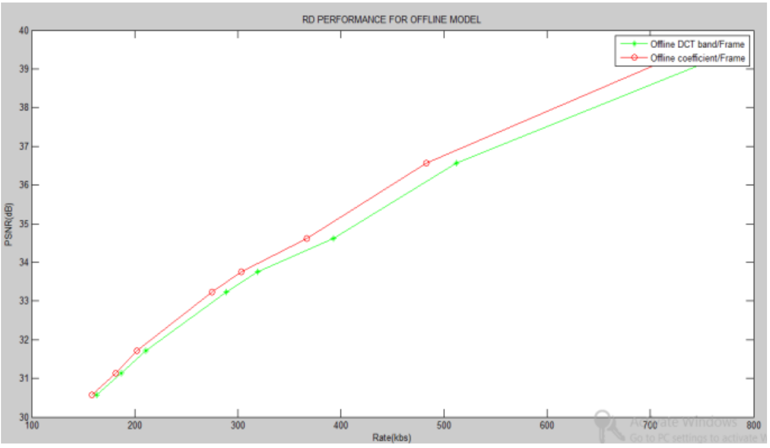

Fig.2

7. RD Performance for Onnline Model

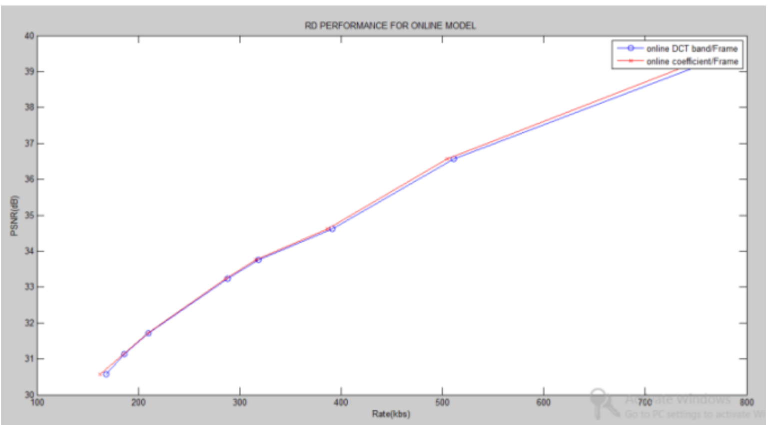

Fig. 3

8. Conclusion

This paper addresses a key issue in turbo-based WZ video coding: the correlation noise modeling. The techniques pro- posed in this paper, for both pixel and transform domains, alleviate the encoder from the high computational and cumbersome task of replicating the side information and allow the decoder to perform online estimation of the CNM parameters. These methods enable practical WZ video coding solutions where the encoder really has low complexity. This paper proposes both offline (encoder-generated) and online (decoder-generated) techniques, which work at different gran- ularity levels, to estimate the CNM. Experimental results show that better RD performance is achieved for the lowest granularity level, the pixel level for the PD solution, and the DCT coefficient/frame level for the TD solution, since both temporal and spatial correlation are explored at the deepest granularity. In a general way, comparing the online correlation noise estimation methods proposed in this paper with the offline equivalent ones, allows to conclude that there is a small coding loss justified by the fact that the offline methods use the original WZ data (which is in practice not viable). One of the next research challenges will be to combine the online estimation techniques proposed here with information about the spatial coherence of each block in its neighborhood within the side information to improve the RD performance. Another challenge will consist on extending these techniques to a multiview scenario where multiple correlated views are available, and thus where interview, temporal and spatial correlations should be exploited to estimate the CNM parameters.

References

[1] Catarina Britesand FernandoPereira "Correlation noisemodelling for efficient pixel and time domain Wyner-ziv video coding". IEEE transactions on circuits and systems for video technology, vol. 18, no. 9, September 2008

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inf. Theory, vol. 22, no. 1, pp. 1–10, Jan. 1976. [4] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," Proc. IEEE, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[3] M. Dalai, R. Leonardi, and F. Pereira, "Improving turbo codec integration in pixel-domain distributed video coding," in *Proc.* IEEE ICASSP, Toulouse, France, May 2006, vol. 2, pp. 537–540.

[4] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domainWyner-Ziv video coding performance," in *Proc. IEEE ICASSP*,Toulouse, France, May 2006, vol. 5, pp. II-525–II-528.

[5] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol.9, no. 10, pp. 1661–1666, Oct. 2000

[6] The H.264 Advanced Video Compression standard ,second edition,Iain E. Richardson ,*Vcodex Limited, UK*

[7] .Wyner,"Recent Results in the Shannon Theory,"*IEEETrans.on Information Theory*, vol. 20,no. 1, pp. 2–10, January 1974," IEEE Trans. on Information Theory, vol. 42, no. 6, pp. 2073–2084,November 1996.

[8] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between Source Coding and ChannelCoding and Its Extension to the Side Information Case," IEEE Trans. on Information Theory,vol. 49, no. 5, pp. 1181–1203,May 2003.

[9] C. Brites, J. Ascenso, and F. Pereira, "Modeling correlation noise sta- tistics at decoder for pixel based Wyner-Ziv video coding," in Proc. PCS, Beijing, China, Apr. 2006.

[10] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, "DistributedMonoview and Multiview Video Coding," IEEE Signal Processing Magazine, vol. 24, no. 5,pp. 67–76, September 2007.

[11] A. Aaron and B. Girod, "Wyner-Ziv video coding with low encoder complexity," in Proc. PCS, San Francisco, CA, Dec. 2004.

Biographical Notes

Sachidananda Samal received the BE degree in Electronics and Telecommunication Engineering from North Odisha University in 2001 and M.Tech. degree in Electronics and Communication from IIT Kharagpur in 2014. Currently he is working as Assistant Professor and Head of the Department of Electronics and Telecommunication in DRIEMS, Cuttack, Odisha. He is having the teaching experience of 14 years. His fields of research have covered signal processing, image processing and advanced video processing related to 4G communication.