ON THE SIDE-INFORMATION DEPENDENCY OF THE TEMPORAL CORRELATION IN WYNER-ZIV VIDEO CODING

Nikos Deligiannis*, Adrian Munteanu, Tom Clerckx, Jan Cornelis and Peter Schelkens

Department of Electronics and Informatics (ETRO) – Interdisciplinary Institute for Broadband Technology (IBBT), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium.
*ndeligia@etro.vub.ac.be

ABSTRACT

Current models in Wyner-Ziv video coding consider the temporal correlation noise to be *side-information independent* (SII). This paper goes beyond this assumption and proposes a novel model, of which the parameters are *side-information dependent* (SID). The proposed model is experimentally validated showing remarkable accuracy improvement over the conventional SII model. Moreover, a novel SID technique for the accurate estimation of the correlation channel in video is introduced. The proposed technique enables the design of a novel pixel-domain Wyner-Ziv video coding system operating without a feedback channel. Preliminary experimental results show that the proposed codec achieves superior performance compared to the state-of-the-art in pixel-domain Wyner-Ziv coding.

Index Terms— Video coding, Wyner-Ziv coding, temporal correlation, statistical modeling, quantization.

1. INTRODUCTION

Nowadays, the forthcoming class of uplink-oriented, low-power applications has cast attention to the design of novel digital video coding architectures, simultaneously meeting requirements like low-complexity encoding, robustness against transmission errors and high compression efficiency. Potential solutions to satisfy these challenges move towards systems commonly referred to as Distributed Video Coding (DVC) or Wyner-Ziv video coding. DVC frameworks are engendered by the information theoretic findings of *Slepian and Wolf* [1] on noiseless coding of correlated information sources with side-information (SI) available at the decoder, which were later extended for lossy coding by *Wyner and Ziv* [2].

In contrast to the traditional Motion Compensated Prediction Coding (MCPC), Wyner-Ziv video coding exploits at the decoder side the inherent temporal correlation present in the video sequence and thus, causes a complexity shift from the encoder to the decoder. In typical DVC architectures [3-5], conventional intra codecs, such as H.263+ Intra, are employed to encode the key frames, based on which the SI is generated by the decoder. Considering a virtual dependency channel between the original data and the SI, efficient channel codes are used to remove the correlation noise and therefore encode the Wyner-Ziv frames. The strength of the channel codes is estimated at the decoder using a statistical model for the correlation noise and becomes known to the encoder via a feedback channel. However, the presence of such a channel introduces structural latency and renders this system unsuitable for unidirectional applications.

State-of-the-art feedback channel-based Wyner-Ziv approaches model the temporal dependencies in video as SII Laplacian noise. In this paper, we introduce a novel SID model to capture the temporal dependencies, offering a substantial accuracy improvement over the classical SII model. Furthermore, we demonstrate that the entire correlation, and thus the reconstructed Wyner-Ziv frames can be accurately estimated starting from a partial knowledge of the dependency channel at the decoder. This novel technique enables the design of a feedback channel-free spatial-domain Wyner-Ziv codec achieving superior performance compared to the state-of-the-art IST-PDWZ system of [4].

The remainder of the paper is structured as follows. The proposed SID model is described in Section 2 while the technique for estimating the actual dependency channel at the decoder is presented in Section 3. The validation of the proposed model and the experimental results are given in Section 4. Finally, Section 5 draws the conclusions of this work.

2. TEMPORAL CORRELATION MODELING

In DVC, the side-information Y is considered as a noisy version of the source X, corrupted by additive noise $N = X - Y$. The temporal correlation channel is described by the probability density function (PDF):

$$f_{X|Y}(x,y) = f_{X-Y|Y}(x-y,y) = f_{N|Y}(n,y). \quad (1)$$

2.1. SI independent modeling

In previous works [3-8] the correlation noise is assumed to be SII, hence, the temporal correlation model is given by:

$$f_{X|Y}(x,y) = f_{N|Y}(n,y) = f_N(n). \quad (2)$$

Initially, a temporally- and spatially-stationary, SII Laplacian model was employed to characterize the channel [3]. Later on, *Brites et al.* [4] proposed a spatially-stationary SII

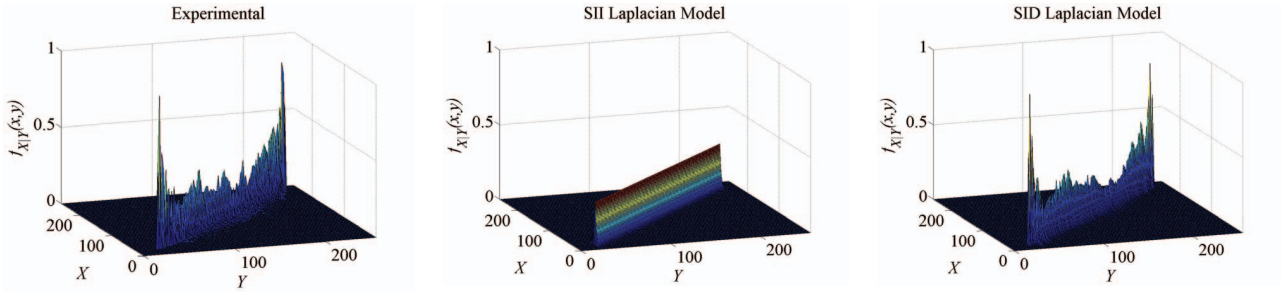

Fig. 1. Instantiation of (a) the experimental temporal correlation dependencies, $f_{X|Y}(x,y)$, (b) the spatially stationary SII Laplacian PDF, and (c) the proposed SID Laplacian PDF for the 85th frame of the Foreman (QCIF) sequence at full frame-rate (30Hz).

Laplacian model and performed an online estimation of the standard deviation at the decoder side. Assuming spatio-temporal non-stationarity, *Westerlaken et al.* [6] improved the performance by discriminating the occluded and non-occluded areas in the SI. In [8], *Brites et al.* modeled the correlation noise as SII Laplacian, block-wise stationary, with constant standard deviation within each block. In addition, assuming the SII noise to be spatially-dependent, they proposed a pixel-based Laplacian model $f_N(n)$ and estimated the standard deviation from the local value of the residual between the forward and backward motion-compensated frames at each pixel position [8]. A similar approach has been followed in [5], extending the ideas of [8] in the DCT domain. In [7], *Westerlaken et al.* showed that the two-sided Gamma and the Generalized Gaussian distributions are more insensitive to modeling ambiguities than a Laplacian distribution. Nonetheless, the most widely-accepted model in the literature for the temporal correlation channel is the zero-mean SII Laplacian distribution, given by:

$$f_N(n) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}|n|}{\sigma}} = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}|x-y|}{\sigma}} = f_{X|Y}(x,y). \quad (3)$$

2.2. SI dependent modeling

In this work, we do not make the assumption that the temporal correlation noise in video is SII. On the contrary, in order to capture more accurately the exact statistics of the temporal correlation channel, we introduce an SID model, expressed by the following probability density function:

$$f_{X|Y}(x,y) = \frac{1}{\sigma(y)\sqrt{2}} e^{-\frac{\sqrt{2}|x-y|}{\sigma(y)}}. \quad (4)$$

We notice that, in contrast to an SII Laplacian model (3), which is a function only of the residual random variable, $N = X - Y$, the proposed model is a function of two random variables, X, Y, and must be plotted as a three-dimensional (3-D) surface. The dependency on the SI is expressed by considering the standard deviation as a function of y. For any given value of the SI, y_0, the proposed model simplifies to a Laplacian distribution with standard deviation σ_0 and a mean value of y_0. An example illustrating the histograms of the actual dependency channel, the conventional spatially-stationary SII model (3), and the proposed SID model is given in Figure 1. It is clear that the proposed model can follow accurately the actual spatio-temporal dependencies between the Wyner-Ziv frame and the SI, compared to the existing model which approximates the channel by its average over the statistics of the SI.

2.3. Successive Approximation Quantization in SID modeling

In this section, we extend the proposed SID modeling scheme, and provide the estimation of the temporal correlation when the Wyner-Ziv source is subjected to embedded scalar quantization. We investigate the particular case of Successive Approximation Quantization (SAQ), which is widely applied in scalable image and video coding. We notice that the optimality of uniform scalar quantization (USQ) in Wyner-Ziv coding regarding IID Laplacian correlation noise has been studied in [9].

Here, we consider SAQ in which every source sample X is quantized to $q_n = Q_n(X) = \lfloor X/2^n\Delta \rfloor$ and reconstructed to $\hat{X} = \lfloor (\lfloor X/2^n\Delta \rfloor + 1/2)2^n\Delta \rfloor$, where $\Delta > 0$, $\lfloor \cdot \rfloor$ is the integer part, $n, 0 \le n < M$ represents the number of discarded bitplanes (or quantization level), and M is the total number of bitplanes. Supposing ideal Slepian-Wolf coding of the quantization indices, the Wyner-Ziv rate and L-2 distortion are calculated using the conditional probability $p_k(q_n(x)\,|\,y)$ of the source belonging to the k^{th} quantization bin, knowing the value y of the SI. Under high rate assumptions, in order to model the partial temporal correlation $p(q_n(x)\,|\,y)$, we propose the following Probability Mass Function (PMF):

$$p(q_n(x)|y) = \frac{1}{\sigma'(y)\sqrt{2}} e^{-\frac{\sqrt{2}|x_b - y_b|}{\sigma'(y)}}, \quad (5)$$

where x_b, y_b are the Wyner-Ziv source and the SI respectively, as represented using $b = M - n$ bitplanes. We notice the use of y_b in the modulus at the exponent. This is similar to the approach of [5], equalizing the quantization levels in the intra and the Wyner-Ziv frames. Furthermore, it is important to mention that for $n \neq 0$ the standard deviation function $\sigma'(y)$ in (5) is different from $\sigma(y)$ in (4). This is because the former refers to a correlation channel between the quantized Wyner-Ziv source and the SI, whereas the latter relates to a dependency channel between the un-quantized Wyner-Ziv source and the SI.

3. TEMPORAL CORRELATION ESTIMATION AT THE DECODER

In this section, we present Overlapped Block Motion Estimation & Probabilistic Compensation (OBMEPC), a novel decoder-oriented technique for the estimation of the entire correlation channel from a partial knowledge of it (resulting from a coarse SAQ of the Wyner-Ziv source).

3.1. Overlapped Block Motion Estimation

Let $X_{\mathcal{B}}(i_0, j_0)$ and $\hat{R}_{\mathcal{B}}(i_0, j_0)$ be some arbitrary blocks of size $B \times B$ pixels containing the b most significant bitplanes of the Wyner-Ziv and reference (previous reconstructed key or Wyner-Ziv) frame at the decoder, respectively, with top-left coordinates (i_0, j_0). The Wyner-Ziv frame is divided into overlapping spatial blocks $X_{\mathcal{B}}(i_0 + \varepsilon, j_0 + \varepsilon)$, where $\varepsilon \in \mathbb{Z}_+$, $1 \le \varepsilon \le B$, is the overlap step-size. The best match of the Wyner-Ziv block is searched in the reference frame within a specified search-range sr. The matching criterion maximizes the ratio $w_{\mathcal{B}}$, which is defined as the number of pixels in $X_{\mathcal{B}}$ of which the b most significant bitplanes are identical to those of the co-located pixels in $\hat{R}_{\mathcal{B}}$, divided by the total number of pixels in the block. The best match in the reference frame serves as the SI $Y_{\mathcal{B}}$ for the considered Wyner-Ziv block $X_{\mathcal{B}}$ and, $w_{\mathcal{B}}$ is an estimate of the amount of correlation between $X_{\mathcal{B}}$ and $Y_{\mathcal{B}}$.

3.1. Probabilistic Motion Compensation

After the execution of the OBME, each pixel in the Wyner-Ziv frame belongs to a number of overlapping blocks $X_{\mathcal{B}_k}$, and for each of these blocks one has a matching block $Y_{\mathcal{B}_k}$ serving as SI. This means that each pixel x in the Wyner-Ziv frame is linked to a number of candidate pixels y_k, each having a certain weight $w_{\mathcal{B}_k}$. Then, the residual reconstructed value $\hat{x}_{(M-b)}$ is calculated as a weighted average of the residual candidate-pixel values $y_{k,(M-b)}$, given by:

$$\hat{x}_{(M-b)} = \sum_k y_{k,(M-b)} w_{\mathcal{B}_k} \Big/ \sum_k w_{\mathcal{B}_k} . \quad (6)$$

4. EXPERIMENTAL RESULTS

In the first set of our experiments, we validate the proposed SID modeling scheme. Throughout the experimental part referring to the validation of the proposed model, SI for the current frame is generated by performing integer-pel motion estimation and compensation at the previous frame in the video sequence. Specifically, the current frame is split into non-overlapping 16×16 blocks, and full search is executed using a search range of ± 15 pixels. The joint statistics of the luminance component between the current frame and the SI are measured, and thus the dependency channel statistics are calculated by the Bayesian law:

$$p(x|y) = p(x,y) \Big/ \sum_x p(x,y) . \quad (7)$$

	Sequences	D_{KL}^{SII} (bits)	D_{KL}^{SID} (bits)	Gain (%)	$\frac{R_{SII} - R_{exp}}{R_{exp}}$	$\frac{R_{SID} - R_{exp}}{R_{exp}}$	Gain (%)
qcif	Akiyo	0.56	0.39	**29.68**	0.51	0.36	**29.31**
	Coastguard	0.29	0.21	**29.18**	0.10	0.07	**29.12**
	Hall	0.46	0.33	**28.22**	0.16	0.12	**28.38**
	Silent	0.76	0.56	**25.37**	0.32	0.24	**25.30**
	Soccer	0.72	0.60	**16.36**	0.19	0.16	**17.75**
cif	Container	0.29	0.21	**29.18**	0.10	0.07	**29.12**
	Flower	1.19	0.43	**63.93**	0.32	0.09	**71.50**
	Football	0.49	0.36	**26.58**	0.10	0.05	**47.43**
	Mobile	0.64	0.48	**24.79**	0.13	0.08	**39.74**
	Stefan	0.67	0.54	**20.60**	0.14	0.05	**61.72**

Table 1: Accuracy in predicting the original correlation channel of the proposed SID PDF in comparison with the corresponding SII spatially stationary PDF, for 10 sequences (5 at QCIF and 5 at CIF resolution) at 30Hz.

Sequences	QCIF					CIF				
	Akiyo	Coastguard	Hall	Silent	Soccer	Container	Flower	Football	Mobile	Stefan
$\bar{D}_{KL}^{SII}$ (bits)	0.17	0.20	0.26	0.29	0.35	0.20	0.50	0.26	0.30	0.33
$\bar{D}_{KL}^{SID}$ (bits)	0.11	0.13	0.17	0.20	0.26	0.13	0.16	0.17	0.19	0.21
Gain (%)	**35.86**	**37.48**	**32.44**	**31.32**	**24.42**	**37.48**	**68.08**	**37.54**	**36.63**	**37.86**
$\overline{(R_{SII} - R_{exp})/R_{exp}}$	0.55	0.46	0.49	0.39	0.30	0.46	0.35	0.17	0.18	0.23
$\overline{(R_{SID} - R_{exp})/R_{exp}}$	0.23	0.18	0.22	0.21	0.17	0.18	0.10	0.08	0.08	0.10
Gain (%)	**57.69**	**60.84**	**55.64**	**46.36**	**42.87**	**60.84**	**72.86**	**53.13**	**56.41**	**58.28**

Table 2: Accuracy of the proposed SID PMF in comparison with the corresponding SII spatially stationary PMF, averaged over all quantization levels of the source (from 1 to 8 bits depth) for 10 sequences (5 at QCIF and 5 at CIF resolution) at 30Hz.

In case of SII modeling we fit a 3-D Laplacian PDF on the entire dependency statistics, which minimize the Kullback-Leibler (KL) distance, denoted by D_{KL}^{SII}, between the actual and the modeled PDFs. On the contrary, in case of the proposed SID model, we fit a 2-D Laplacian PDF on each $P(x|y_i)$ distribution, defined by each value of the SI, y_i, under the requirement of minimizing the marginal D_{KL}^{SID}. The standard deviation of the Laplacian distribution is estimated by a brute-force approach with a precision of two decimal digits. Apart from the entire statistics, this method is repeated for each quantization level of the Wyner-Ziv source in order to fit the proposed PMF given by (5).

Table 1 and 2 summarize the numerical results of the proposed SID PDF and PMF, respectively, averaged over 100 frames of 10 test video sequences (5 QCIF and 5 CIF) at full frame-rate. Table 1 shows that the SID PDF offers a substantial reduction of the overall KL distance which can be translated into a gain of up to 63.93% depending on the sequence statistics. Similarly, the proposed SID model can predict the actual rate (R_{exp}) much more accurately, offering improvements of up to 71.5%. Also, Table 2 reports the KL distance and the precision in predicting the Wyner-Ziv rate, averaged over all SA quantization levels of the Wyner-Ziv source, for the proposed PMF, given by (5). The results highlight the high accuracy of the proposed PMF. Overall, the proposed SID modeling scheme offers a substantial gain in the modeling accuracy over the conventional SII model.

In the following, we show the ability of the proposed OBMEPC technique in estimating the actual dependencies starting from a very coarse approximation of them,

corresponding to a very coarse quantization ($b=1$). Figure 2 depicts the KL distance between the actual temporal correlation channel, here denoted as $f_{\exp}$, and the one estimated by OBMEPC, denoted by $f_{estimated}$ for the Coastguard (QCIF) video sequence at GOP 4. Additionally, the KL distance between the partial knowledge of this channel, corresponding to $b=1$, and the actual channel is depicted. The tremendous drop from a very large to a very low (close to zero) KL distance reveals the capabilities of OBMEPC in estimating the actual channel from a very coarse knowledge of it.

Since OBMEPC requires limited knowledge of the correlation channel, it can be employed in the design of a novel spatial-domain Wyner-Ziv (SDWZ) codec which does not require the use of a feedback channel. SDWZ's light encoding syntax consists of a differential entropy encoder (LZMA) for the b most significant Wyner-Ziv bitplanes and a H.263+ Intra encoder for the reference frames. In Figure 3, the proposed SDWZ is compared against the state-of-the-art IST-PDWZ codec of [4] for the first 101 frames of the Coastguard sequence at GOP 4. The results show that apart from low bitrates, where both Wyner-Ziv codecs exhibit equal performance, the proposed codec clearly outperforms the IST-PDWZ codec introducing a considerable gain, of up to approximately 1dB. Similar performance gains are obtained for the Foreman sequence. At the same time, the proposed scheme retains the advantage of being liberated from the use of a feedback channel.

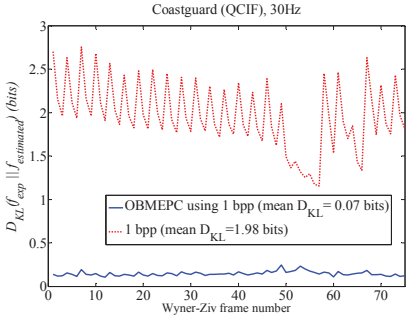

Fig. 2. KL distance per Wyner-Ziv frame, between the original temporal correlation and the one estimated by the OBMEPC for Coastguard GOP 4 at 30 Hz. In OBMEPC, a block size of 16×16 pixels, a stepsize of $\varepsilon=2$ pixels and a search-range of $sr=\pm10$ pixels have been used.

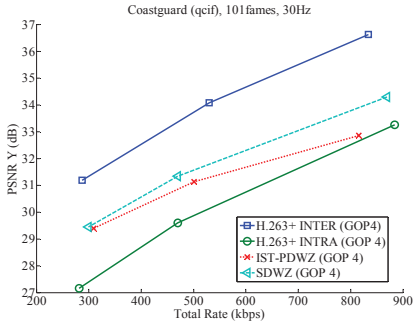

Fig. 3. Coding performance of the SDWZ instantiation, the IST-PDWZ and the H.263+ codec for GOP 4; the luma components from the first 101 frames of Coastguard (QCIF) at 30Hz were encoded. The IST-PDWZ results were reproduced from [4].

5. CONCLUSIONS

In this paper, we have proposed a novel SID modeling scheme to capture the temporal correlation statistics in video which shows substantial accuracy improvements over the existing SII model. Additionally, we have introduced OBMEPC, a novel SID, data-dependent technique which enables the accurate estimation of the actual correlation starting from a partial knowledge of it. OBMEPC enables the design of a spatial-domain Wyner-Ziv system which does not require the use of a feedback channel. Preliminary results show that the instantiation of the SDWZ codec compares favorably against the state-of-the-art feedback channel-based IST-PDWZ system of [4].

6. ACKNOWLEDGEMENTS

This work was supported by FWO Flanders (post-doctoral fellowships of A. Munteanu and P. Schelkens and project G.0391.07), IWT (GBOU project Resume and PhD bursary of Tom Clerckx) and IBBT via the ISBO-VIN project.

7. REFERENCES

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

[2] A. D. 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.

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

[4] C. Brites, J. Ascenso, and F. Pereira, "Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding," *Picture Coding Symposium*, PCS 2006, Beijing, China, pp. 1-6, April 2006.

[5] C. Brites, J. Ascenso, J. Q. Pedro, and F. Pereira, "Evaluating a feedback channel based transform domain Wyner-Ziv video codec," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 269-297, April 2008.

[6] R. P. Westerlaken, R. K. Gunnewiek, and R. L. Lagendijk, "The role of the virtual channel in distributed source coding of video," *IEEE International Conference on Image Processing*, ICIP 2005, Genova, Italy, vol. 1, pp. 581-584, September 2005.

[7] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. L. Lagendijk, "Dependency channel modeling for a LDPC-based Wyner-Ziv video compression scheme," *IEEE International Conference on Image Processing*, ICIP 2006, Atlanta, GA, USA, pp. 277-280, October 2006.

[8] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel-based Wyner-Ziv video Coding," *IEEE International Conference on Image Processing*, ICIP 2006, Atlanta, GA, USA, pp. 273-276, October 2006.

[9] V. Sheinin, A. Jagmohan, and D. He, "Uniform threshold scalar quantizer performance in Wyner-Ziv coding with memoryless, additive Laplacian correlation channel," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP 2006, Toulouse, France, vol. 4, pp. 217-220, May 2006.