

Two-Dimensional Iterative Source-Channel Decoding for Distributed Video Coding

Lajos Hanzo, Yongkai Huo, Rob Maunder, and Tao Wang

Abstract—Motivated by the Joint Source-Channel Decoding (JSCD) principle of exploiting the source redundancy, in this treatise we study the application of iterative source-channel decoding (ISCD) conceived for distributed video coding (DVC), where the video signal is modelled by our Iterative Horizontal-Vertical Scanline Model (IHVSM) relying on a first-order Markov process. The IHVSM technique is combined with the classic forward error correction (FEC) codecs employed in the state-of-the-art DVC systems for the sake of reducing the bitrate. Our simulation results show that up to 21.5% bitrate reductions are achieved by employing the proposed ISCD technique in a DVC scheme. Alternatively, a peak signal-to-noise ratio (PSNR) gain of 2.2 dB is achieved at a bitrate of 4.5 Mbps when considering the Foreman sequence.

Index Terms—

I. INTRODUCTION

SLEPIAN-WOLF theory [1] has attracted substantial research attention for the sake of shifting the computational complexity of the transmitter to a central mains-supplied receiver, where power-dissipation is less of a constraint. Suppose X and Y are two independently and identically distributed (i.i.d.) binary sequences, which may be generated either at the same or different locations. The Slepian-Wolf (SW) Theorem [1] states that it is possible to compress X and Y independently using two separate encoders and then to jointly decode them at the receiver, using a rate similar to that as though X and Y were encoded jointly, i.e. as a single sequence. Hence the SW philosophy may be viewed as the complement of using a high-complexity encoder and a low-complexity decoder. The punctured turbo codes and polar codes based SW schemes were investigated for compressing two distributed correlated sources in [2] and [3], respectively.

Wyner and Ziv (WZ) [4] further extended this theory to lossy coding scenarios. Motivated by the emerging source encoding requirements of low-complexity video sensor networks, Wyner-Ziv video coding, also known as distributed video coding (DVC), has attracted substantial attention in recent years, as documented both by Girod and his team [5], as well as by Puri *et al.* [6]. Brites and Pereira [7] proposed a more realistic WZ video coding approach, which performs online estimation of the channel-induced noise (CN) model parameters at the decoder. Moreover, the online estimation of channel model parameters is also studied in [8] in the context of DVC. The Slepian-Wolf and Wyner-Ziv theorems state that the sequences X and Y must be jointly decoded in order to

approach the theoretical bound of the joint source entropy $H(X, Y)$. Hence it is possible to apply iterative decoding principles for further exploiting both the intra-frame and inter-frame source correlation at the receiver.

A common feature in WZ video coding is that a high amount of redundancy is present in the transmitted video stream, which manifests itself in terms of a high adjacent-pixel correlation and should be exploited at the receiver for the sake of reducing the bitrate. A Markov random field approach was investigated in [9] for the sake of modeling and exploiting the residual correlation at the receiver. In this treatise, we propose an iterative source-channel decoding (ISCD) [10] aided WZ video coding system, which employs our new iterative horizontal-vertical scanline model (IHVSM) aided decoder. Against this background, our novel contribution is that in contrast to the conventional video scenario of [11], here we conceive a ISCD based distributed video coding system, where the two-dimensional IHVSM and the forward error correction (FEC) codes perform turbo-like iterative extrinsic information exchange for the sake of reducing the required bitrate.

This rest of this paper is organized as follows. In Section II, we introduce the architecture of the IHVSM aided iterative source decoding model. Then the system's architecture and the performance of the IHVSM aided pixel-domain Wyner-Ziv coding scheme will be detailed in Section III. Finally, our conclusions are offered in Section IV.

II. ITERATIVE SOURCE DECODING MODEL

This section briefly introduces our IHVSM technique conceived for decoding 2D video frames, while more details, including parameter training and complexity analysis, can be found in [11], [12]. Let us commence by stipulating the following assumptions:

- x_i : the m -bit pattern representing a pixel scanned from the original video pixels at index i , which is expressed as $\{x_i(0), \dots, x_i(m-1)\} = x_i \binom{m-1}{0}$;
- m : the number of bits in each m -bit pattern x_i of pixels;
- $X_m = \{0, 1, \dots, 2^m - 1\}$: the set of all possible values in an m -bit pattern x_i ;
- $x_0^t = x_0, \dots, x_t$: the one-dimensional indexed bit patterns scanned from a frame of the original video;
- $y_0^t = y_0, \dots, y_t$: error-infested bit patterns of x_0^t ;

The model of 2D iterative source decoder is displayed in Fig. 1, which only accepts the softbit input and improves the soft pixel information. Below we will firstly introduce the preliminaries of our 2D IHVSM model, including the required softbit input and the relevant decoding techniques of first-order

Manuscript received September 26, 2013. The associate editor coordinating the review of this letter and approving it for publication was E. Paolini.

The authors are with the University of Southampton, United Kingdom (e-mail: lh@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/LCOMM.2013.111513.132180

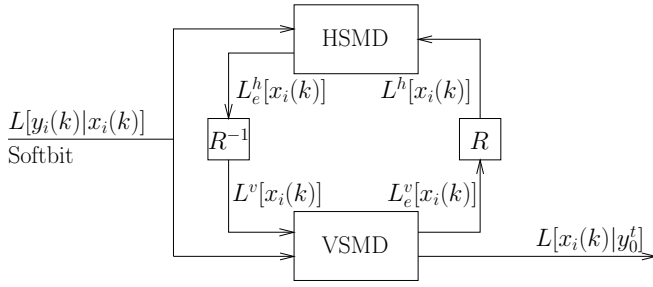


Fig. 1: Architecture of IHVSM, where R represents reordering of the video pixels.

Markov processes, followed by conceiving our 2D iterative source decoding technique.

A. Preliminaries

1) *Softbit Input*: Let us now focus our attention on the 1st original video frame carrying $(t + 1)$ consecutive and hence correlated bit patterns, e.g. x_0, \dots, x_t . Due to the channel-impairments, the receiver can only reconstruct the error-infested version of x_0, \dots, x_t . Again, the 2D iterative source decoder shown in Fig. 1 only accepts softbit information as its input. Then, the receiver has to estimate the softbit information of each bit representing the original pixels x_0, \dots, x_t , namely y_0, \dots, y_t , which carry the error-infested bit sequence $y_0 \binom{m-1}{0}, \dots, y_t \binom{m-1}{0}$. The softbit information represented by the log-likelihood ratios (LLR) may be readily derived from the softbit patterns y_0, \dots, y_t , yielding $L[y_i(k)|x_i(k)]$ in Fig. 1, which indicates the receiver's confidence in the original m -bit pixel $x_i(k)$.

2) *Decoding of First-Order Markov Chain*: Let us assume that the soft Markov process y_0, \dots, y_t are known at the Markov decoder, which represented the error-infested soft information of the original Markov process x_0, \dots, x_t . The objective of the Markov decoder is to estimate the quantized pixels x_0, \dots, x_t based on the soft pixels. Given the soft information y_0, \dots, y_t , the quantized pixel x_i may be determined by the a-posteriori probability (APP)

$$p(x_i|y_0, \dots, y_t) = p(x_i|y_0^t), \quad (1)$$

where we have $x_i \in X_m$, $0 \leq i \leq t$. The bit-based LLR version of the APP $p(x_i|y_0^t)$ may be expressed as [11]

$$L[x_i(k)|y_0^t] = \ln \frac{\sum_{\substack{x_i^{(k)}=0 \\ x_i \in X_m}} \beta(x_i) \cdot \chi(x_i) \cdot \alpha(x_i)}{\sum_{\substack{x_i^{(k)}=1 \\ x_i \in X_m}} \beta(x_i) \cdot \chi(x_i) \cdot \alpha(x_i)}, \quad (2)$$

where $\beta(x_i)$, $\alpha(x_i)$ are the backward and forward recursion probability parameters of the pixel x_i , while $\chi(x_i)$ is the channel information for the pixel x_i , as detailed in [10], [11]. Moreover, the extrinsic information $L_e[x_i(k)]$ of the bit $x_i(k)$ may be calculated for iterative decoding as detailed in [11].

B. Iterative Source Decoding

Given the softbit LLR input $L[y_i(k)|x_i(k)]$ of Fig. 1, our 2D iterative source decoder may be invoked for correcting the

effects of channel-errors on the error-infested m -bit sequence of $(t + 1)$ pixels $y_0 \binom{m-1}{0}, \dots, y_t \binom{m-1}{0}$. The decoding architecture of the IHVSM is displayed in Fig. 1.

Specifically, two similar source decoders are employed, namely the Horizontal Scanline Model Decoder (HSMD) operating in the horizontal direction and the Vertical Scanline Model Decoder (VSMD) proceeding in the vertical direction. Let us now assume that we receive H horizontal scanlines and V vertical scanlines, where each scanline indicates a first-order Markov process. Then the IHVSM decoding may be performed based on the $(H \times V)$ -pixel block, which is represented by the $(H \cdot V)$ - m -softbit patterns y_i . Here we consider the HSMD as an example for highlighting the decoding process. The m -softbit-based pixel value y_i is used for generating the systematic LLR information $L[y_i(k)|x_i(k)]$, which is then input to the HSMD without the assistance of any channel decoding. However, our technique may be conceived with arbitrary channel codecs. The HSMD also exploits the *a-priori* LLR information $L^h[x_i(k)]$ generated from the extrinsic information $L_e^v[x_i(k)]$ provided by the vertical Markov decoder. The horizontal decoder independently performs source-modelling aided soft decoding of each of the H horizontal scanlines. Hence, following horizontal decoding, the extrinsic LLR information $L_e^h[x_i(k)]$ may be generated for all H scanlines. Then the relevant $(H \times V)$ -line extrinsic information block will be appropriately reordered in the block R^{-1} of Fig. 1 for generating the *a-priori* information $L^v[x_i(k)]$ for the vertical decoder.

III. IHVSM AIDED WYNER-ZIV VIDEO CODING

In the Wyner-Ziv video coding schemes of Girod *et al.* [5] and of Brites and Pereira [7], the side information (SI) of the WZ frame is estimated from the key frames, which are intra-coded by standard codecs, such as H.264 for example. Furthermore, a part of the parity bits of WZ frames is transmitted to the receiver upon request for the sake of assisting the receiver to recover the WZ frames. However, in the spirit of [11] the residual pixel-correlation within the WZ frames may be further exploited for the sake of improving the system's performance.

In this section, we will combine our IHVSM with the turbo-coding aided WZ codec [7], in order to iteratively conceal the softbit errors returned by the turbo decoder, thereby implicitly reducing the requested bitrate. Note that both the key frames and the WZ frames are transmitted to the receiver through a perfect channel. Hence the errors of the WZ frames are purely introduced by the SI estimation process at the receiver, where our proposed IHVSM model will exploit the intra-frame correlation of the WZ frames for correcting the errors, thereby reducing the amount of parity bits required for the WZ frames.

A. System Architecture

Two types of techniques exist in the literature for encoding the WZ frames, namely pixel-domain WZ (PDWZ) [5], [7] and transform-domain WZ (TDWZ) [7] video coding. The PDWZ codec imposes a lower complexity on the encoder than the TDWZ scheme, hence - as expected - it exhibits a

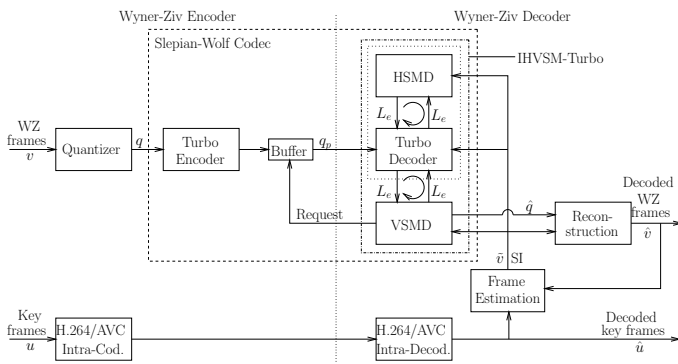


Fig. 2: System architecture of our proposed IHVSM-PDWZ system, which is based on the PDWZ system proposed in [7].

worse RD performance. In this section, we apply the proposed IHVSM algorithm in the context of a PDWZ scheme for improving the attainable RD performance. We note however that the transformation technique and quantization techniques applied to the WZ frames only partially exploit the intra-frame correlation, since the entropy coding is not employed. Nonetheless, our proposed Markov modeling techniques will also be investigated in a TDWZ system in the future. We opted for the PDWZ system of [7] as our benchmark, while some further details are illustrated in [13]. Note that the proposed IHVSM may also be employed in other WZ video coding schemes provided that the video decoders are capable of estimating the softbit information of the video pixels.

Let us now proceed by detailing the IHVSM-turbo coded PDWZ system of Fig. 2. At the transmitter, the key frame u is intra-coded by the H.264/AVC video codec, whilst the WZ frame v is uniformly quantized for generating the 2D signal q . Then the bit-pattern q is encoded by the classic turbo codec [14], where only the parity bits q_p will be buffered and transmitted to the receiver. At the receiver, the key frame is reconstructed using the H.264/AVC intra-decoder. The decoded key frame \hat{u} will then be utilized by the frame estimator of Fig. 2 for estimating the SI frame \tilde{v} . Then, our IHVSM-Turbo decoder of Fig. 2 will perform three-stage iterative decoding [12] using the estimated SI frame \tilde{v} and the received parity bits q_p . For the IHVSM-Turbo scheme, the SI frame \tilde{v} may be used as the systematic information of the corresponding WZ frame v , which will be exploited by all the three decoding components, namely the HSMD, the turbo codec, and the VSMD of Fig. 2. For the HSMD and the VSMD components, the technique introduced in Section II-B is utilized for iterative decoding. Furthermore, the HSMD and the turbo codec constitute the inner decoding component, while the VSMD is the outer one. Similarly to most proposals found in the literature [5], [7], the "request-and-decode"¹ process of [5] is repeated, until an acceptable bit error ratio is reached. Finally, the quantized WZ frame is recovered as \hat{q} , which will be utilized together with the estimated SI frame \tilde{v} for reconstructing the WZ frame \hat{v} .

Below, we benchmark the IHVSM-PDWZ scheme's per-

¹When failing in decoding a WZ frame, the receiver sends a request to the transmitter, which will transmit additional parity bits to the receiver. Then the receiver re-decodes the WZ frame upon the accumulated parity bits.

TABLE I: Table of parameters employed for the *Hall*, *Foreman* and *Coastguard* sequences represented CIF.

	Hall	Foreman	Coastguard
Representation	YUV 4:2:0	YUV 4:2:0	YUV 4:2:0
Format	CIF	CIF	CIF
Bits Per Pixel	8	8	8
FPS	30	30	30
GOP	2	2	2
Number of Frames	300	300	300
Channel	Perfect	Perfect	Perfect

formance against that of the PDWZ system of Brites and Pereira [7], where the traditional turbo codec was employed. Moreover, as seen in most of the state-of-the-art contributions [5], the corresponding results recorded both for the JPEG and for the H.264/AVC schemes are also provided as baseline benchmark curves in Section III-B. More specifically, all the frames were intra-frame (I) coded by the JPEG and H.264 codecs, as well as by the H.264/AVC NoMotion mode, where all the frames may be coded as I frames and predicted frames (P) without inter-frame motion estimation. Three sequences are tested, namely the Hall monitor, the Foreman and the Coastguard sequences, whose parameters are shown in Table I. In all the experiments, the group of picture (GOP) parameter was set to 2; the bitrate/peak signal-to-noise ratio (PSNR) of both the WZ and the key frames was also taken into account. The RD results recorded for both the PDWZ and IHVSM-PDWZ schemes in Fig. 2 were parameterized by the number of WZ coded bitplanes, which was set to $m = 1, 2, 3$ or 4 most significant bit (MSB) planes, as usual in the pixel-domain WZ video coding literature. This was arranged by invoking the uniform quantizer shown in Fig. 2. In [13], each bitplane of the MSB planes was transmitted separately and each bitplane was then refined based on the previously decoded bitplanes. However, in our system, all bitplanes of the MSB planes were transmitted together, which allowed us to reduce the number of "request-and-decode" processes defined in [5]. The remaining parameters of our system were identical to those in [7], [13]. For the IHVSM-PDWZ schemes, the size of the Markov Model's State Transition Table (MMSTT) was of $(2^m \times 2^m)$ elements. The MMSTT may be trained at the transmitter using the entire video sequence or at the receiver using the key frames. More details for training of MMSTT can be found in [12].

B. Numerical Results

The simulation results recorded for the three sequences are displayed in Fig. 3. Observe from Fig. 3a that for the Hall-CIF sequence the IHVSM-PDWZ using a single iteration² reduces the bitrate by about 0%, 11%, 14.7% and 15% in comparison to the PDWZ scheme for $m = 1, \dots, 4$ MSB planes, respectively. This is due to the fact that in contrast to the turbo decoder the IHVSM-turbo decoder performs

²The performance is only modestly improved upon increasing the number of decoding iterations, whilst increasing the complexity. Hence, using a single iteration is preferred.

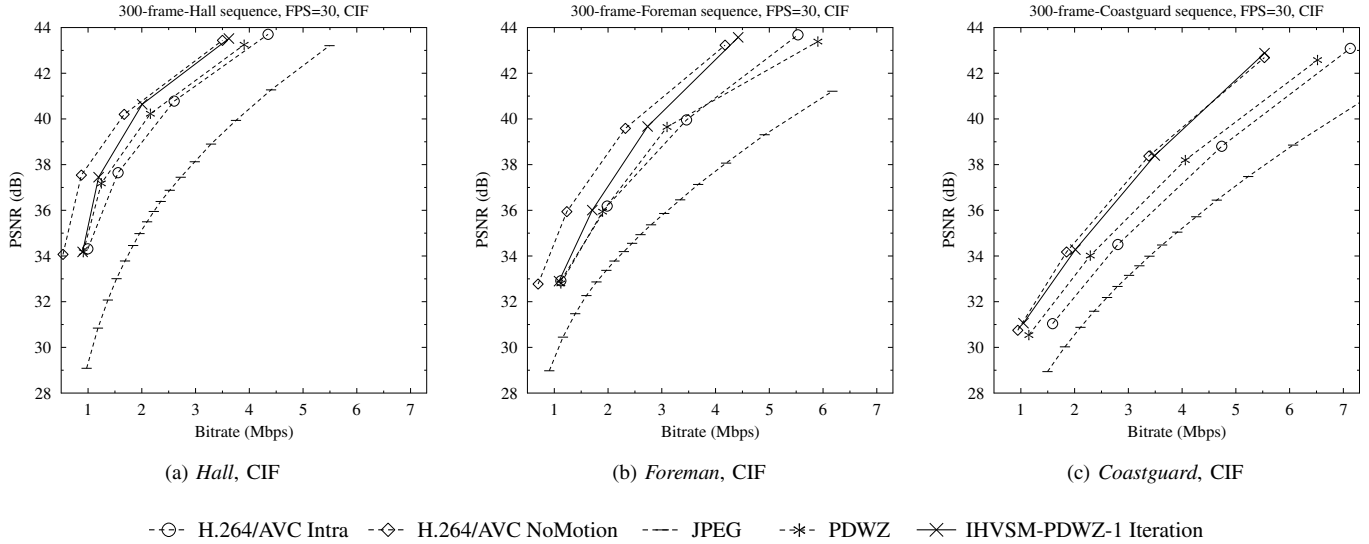


Fig. 3: Rate-distortion performance comparison of the IHVSM-PDWZ codec and the benchmarks: JPEG, H.264/AVC Intra, H.264/AVC NoMotion, PDWZ [7]. *Hall*, *Foreman* and *Coastguard* sequences.

ISCD, thereby achieving a lower BER for the decoded WZ frames, which iteratively exploits the pixel correlation within the WZ frames. Explicitly, the IHVSM-PDWZ is capable of maintaining the same video quality for the WZ frames with the aid of less parity bits. Similarly, observe from Fig. 3b that for the *Foreman*-CIF sequence the IHVSM-PDWZ using $I = 1$ iteration reduces the bitrate by 3%, 10.1%, 11.6%, 25.1% for $m = 1, \dots, 4$ MSB planes, respectively. For the *Coastguard*-CIF sequence, the IHVSM-PDWZ using $I = 1$ iteration reduced the bitrate by 8.7%, 12.1%, 14%, 15.1% for $m = 1, \dots, 4$ MSB planes, respectively. For all the test sequences, the bitrate reduction increased with the number of MSB planes considered. The reason for this is that the receiver is more unlikely to be able to estimate the SI of WZ frame accurately from the received key frames, while our IHVSM scheme is capable of effectively correcting the errors in the turbo decoding process.

Moreover, as seen in Fig. 3, the IHVSM-PDWZ system of Fig. 2 is comparable to the H.264/AVC-NoMotion system at sufficiently high bitrates corresponding to high PSNRs. For example, in the PSNR range in excess of 41 dB, the IHVSM-PDWZ slightly outperforms the H.264/AVC-NoMotion system, when employing the *Coastguard*-CIF sequence.

IV. CONCLUSIONS

In this paper, we conceived an iterative source-channel decoding technique for Wyner-Ziv video coding, where the IHVSM was combined with the turbo decoder for the sake of reducing the required bitrate. Our simulation results showed that we were able to reduce the bitrate by up to 21.5% compared to the PDWZ benchmarker system.

Our future work will focus on refining the IHVSM decoder, which may further improve the attainable system performance.

REFERENCES

- [1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.
- [2] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, 2001.
- [3] X. Lv, R. Liu, and R. Wang, "A novel rate-adaptive distributed source coding scheme using polar codes," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 143–146, 2013.
- [4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, pp. 1–10, Jan. 1976.
- [5] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, pp. 71–83, Jan. 2005.
- [6] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," *IEEE Trans. Image Process.*, vol. 16, pp. 2436–2448, Oct. 2007.
- [7] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 18, pp. 1177–1190, Sept. 2008.
- [8] S. Wang, L. Cui, L. Stankovic, V. Stankovic, and S. Cheng, "Adaptive correlation estimation with particle filtering for distributed video coding," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 22, no. 5, pp. 649–658, 2012.
- [9] Y. Zhang, H. Xiong, Z. He, S. Yu, and C. W. Chen, "Reconstruction for distributed video coding: a context-adaptive Markov random field approach," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 21, pp. 1100–1114, Aug. 2011.
- [10] M. Adrat and P. Vary, "Iterative source-channel decoding: improved system design using EXIT charts," *EURASIP J. Applied Signal Process.*, vol. 2005, pp. 928–941, Jan. 2005.
- [11] Y. Huo, T. Wang, R. G. Maunder, and L. Hanzo, "Iterative source and channel decoding relying on correlation modelling for wireless video transmission," to appear in *IET Commun.*, 2013. Available: <http://eprints.soton.ac.uk/350839/1/IETCom.pdf>.
- [12] Y. Huo, C. Zhu, and L. Hanzo, "Spatio-temporal iterative source-channel decoding aided video transmission," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1597–1609, 2013.
- [13] C. Brites, "Exploiting correlation noise modeling in Wyner-Ziv video coding." Ph.D. thesis, Instituto Superior Técnico, Technical University of Lisbon, 2011.
- [14] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes," in *Proc. 1993 International Conference on Communications*, pp. 1064–1070.