

An Error Resilient Video Coding Scheme Using Embedded Wyner–Ziv Description with Decoder Side Non-Stationary Distortion Modeling

Yongsheng Zhang, Hongkai Xiong, *Senior Member, IEEE*, Zhihai He, *Senior Member, IEEE*, Songyu Yu, and Chang Wen Chen, *Fellow, IEEE*

Abstract—In this paper, we propose a generic error resilient video coding (ERVC) scheme using embedded Wyner–Ziv (WZ) description. At the encoder side, a joint source-channel R-D optimized mode selection (JSC-RDO-MS) algorithm with WZ-coded anchor frames is statistically studied and developed. Given a stationary first-order Markov Gaussian source, the proposed mode optimization is justified by an analysis of the RD impact on the WZ bit-rate. JSC-RDO-MS involves in the estimation of expected rate and distortion of WZ coding with the unavailable side information, and the WZ bit-rate of each coding mode is determined based on the error correction capability of the specific WZ codec. At the decoder side, an online correlation noise model between the source and the side-information is proposed with a mixture of Laplacians whose parameters are attained to reflect the coherence of the motion field of successive frames and the energy of prediction residual. Each mixture component represents the statistical distribution of prediction residuals, and the mixing coefficients represent the amount of errors in motion compensation. The proposed scheme achieves the so-called classification gain by exploiting the spatially non-stationary characteristics of the motion field and texture. Extensive experimental results show that the proposed WZ-ERVC scheme achieves a better overall RD performance than existing ERVC schemes, and the proposed modeling algorithm also significantly outperforms the conventional Laplacian model by up to 2 dB.

Index Terms—Correlation noise modeling, distributed video coding, error resilient video coding, non-stationary, Wyner–Ziv coding.

I. INTRODUCTION

BECAUSE of the sophisticated spatio-temporal prediction schemes used in H.264/AVC video coding, compressed

Manuscript received March 4, 2010; revised May 25, 2010 and August 20, 2010; accepted October 21, 2010. Date of publication March 17, 2011; date of current version April 1, 2011. This work was supported in part by the NSFC, under Grants 60632040, 60772099, 60802019, 60928003, by the Program for New Century Excellent Talents in University (NCET-09-0554), and by the National High Technology Research and Development Program of China (2006AA01Z322). This paper was recommended by Associate Editor R. Bernardini.

Y. Zhang, H. Xiong, and S. Yu are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yongsheng.zhang.cn@gmail.com; xionghongkai@sjtu.edu.cn; syyu@sjtu.edu.cn).

Z. He is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: hezhi@missouri.edu).

C. W. Chen is with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: chencw@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2011.2129050

H.264/AVC bit streams are very sensitive to transmission errors. Error-resilient H.264/AVC video streaming has become an important topic of research. The approaches proposed in the literature [1]–[4] can roughly be classified into the following categories. The first category focuses on link-layer reliability, typically, forward error correction (FEC) [2] and/or automatic repeat request; the second category considers the intrinsic source dependence, attempting to minimize the quality deterioration caused by transmission errors using error concealment and error resilient video coding (ERVC) methods. While error concealment approaches try to estimate the lost blocks in a video frame at the decoder side using the spatio-temporal correlation of video data [3], ERVC approaches try to generate more robust encoded bit-stream at the encoder side, e.g., using more intra-coded macro-blocks (MBs) [5], robust motion estimation [4], redundant slices, or flexible macro-block ordering (FMO) [6], and so on. Recently, distributed source coding, more specifically, Wyner–Ziv (WZ) coding [7], emerges as a promising scheme for error-resilient video coding, which integrates the encoder-driven error resilience and the decoder-driven error concealment.

Although WZ coding for error resilience also depends on channel codes, e.g., Turbo and LDPC, there are major differences between WZ-based and FEC-based error resilience tools. First, conventional FEC schemes aim to protect the coded bit-stream and cannot eliminate error propagation from previous frames, whereas the WZ-based approaches attempt to protect the waveform of frames and remove errors either occurring on current frame or propagated from previous frames. Second, conventional FEC schemes only consider transmission errors, while the WZ-based approaches jointly analyze transmission errors and intrinsic correlation of video sequences to produce more robust bit-streams. In order to investigate statistical correlation with distributed source coding and the rate-distortion optimization over the conventional video compression framework, the dependency modeling of lossy video communication and the embedded WZ-based error resilient video coding are two major issues to be addressed in this paper.

Recently, WZ coding has been concerned to protect the waveform of video signal for error resilience [8]–[12]. Rane *et al.* [8], [9] proposed to correct transmission errors using a coarsely quantized version of video signal with WZ coding. It could be thought as an extension of the FEC scheme by

protecting the bit-stream of a coarsely requantized version of video frames. Sehgal *et al.* [10] attempted to terminate temporal error propagation by periodically inserting WZ protected frames. Later, Zhang *et al.* [11] encoded the peg frames of the Sehgal’s scheme with a unified WZ codec not only to exploit the correlation between the current WZ frames and its reference frames, but also to protect against transmission errors. Bernardini *et al.* [12] used an auxiliary stream containing parity bits for each frame to protect video signal. These schemes, however, do not provide a correlation modeling algorithm at the decoder side. In WZ coding applications, side-information Y is considered as a corrupted version of the source X by channel errors, and the WZ decoding is used to correct these channel errors. In general, a correlation model to describe the virtual channel between source X and side-information Y has been estimated with a statistical distribution whose parameters are estimated offline [7], [13]. Different correlation models lead to different levels of performance [14]. Although the encoder could not have access to the error pattern during data transmission, those approaches commonly assume that both the source data and corresponding side-information are available at the encoder side or the decoder side. An online correlation noise modeling, to model the correlation noise at the decoder side without the source data, is desirable to be concerned. Based on the analysis of side-information generation with bi-directional motion prediction, an online correlation noise modeling algorithm has ever been developed [15]. Toward a more practical WZ-based error resilient video coding scenario where the side-information is typically generated from the error concealed picture instead of bi-directional motion prediction, our previous efforts in [16] proposed an online Laplacian mixture model by exploiting the spatially non-stationary characteristics of the motion field and texture. It is superior to estimate the correlation noise based on the coherence property of the motion field and the spatial smoothness of neighboring blocks.

In this paper, we propose a generic WZ-based error resilient video coding (WZ-ERVC) scheme where a subset of frames, called *anchor frames*, are protected with WZ coding to prevent temporal error propagation. It refers to two major contributions. At the encoder side, a joint source-channel R-D optimized mode selection (JSC-RDO-MS) algorithm with WZ coding is developed to optimize the overall R-D performance for a given total bit-rate of the primary description (H.264/AVC bit-stream) and the redundant description (embedded WZ bit-stream) of *anchor frames*. Given a stationary first-order Markov Gaussian source, the proposed mode optimization is motivated by an analysis of the RD impact on the WZ bit-rate. Different from the rate and distortion estimation in conventional video coding scheme, the expected rate and distortion of WZ coding for JSC-RDO-MS depend on the side information that is only available at the decoder side. Instead of modeling the correlation noise between the source and the side-information using a binary symmetric channel [17], [18], we use a widely adopted symbol-level correlation noise model [14] and estimate the minimal achievable rate, i.e., the conditional entropy $H(X|Y)$ for WZ video coding. Based on this model, we can determine the minimum WZ coding

bit-rate which guarantees a maximum decoding failure probability P_e . Statistically, we estimate the number of the parity bits from a combination of P_e and $H(X|Y)$ with a training set over a set of video sequences. At the decoder side, we propose an online correlation noise model to capture the spatially non-stationary characteristics of source correlation in WZ coding. Without using the source data, it leads to a more realistic WZ-ERVC solution when no feedback channel is available or a stringent end-to-end delay constraint exists. Since the correlation noise model depends on the coherence of motion field and the energy of prediction residual, it can model the spatially non-stationary characteristics of the correlation noise in WZ video coding, and achieves the so-called classification gain [19]. In the proposed online correlation noise modeling algorithm, the correlation noise between the source and side-information is modeled in both the pixel and the DCT domains, by a mixture of Laplacians whose parameters are obtained by analyzing the coherence of the motion field of successive frames and the energy of prediction residual. More specifically, when packet loss occurs, the transmission error e consists of two parts: e_w , the error caused by prediction residual loss, and e_ρ , the error caused by motion vectors (MVs) loss. The distribution of e can be written as

$$p(e) = p(e_w + e_\rho) = \sum_{e_\rho} p(e_w|e_\rho)p(e_\rho). \quad (1)$$

Since e_w follows a Laplacian distribution [20], the $p(e_w|e_\rho)$ in (1) also follows Laplacian distribution because e_w and e_ρ are uncorrelated, this assumption will be justified later in this paper. Therefore, $p(e)$ can be modeled by a mixture of Laplacian distributions, where each Laplacian distribution $p(e_w|e_\rho)$ is called a mixture component with its own mean e_ρ . $p(e_\rho)$ in (1) is called the mixing coefficients which determine the weight of each mixture component. The energy of prediction residual determines the variance of $p(e_w|e_\rho)$.

The rest of this paper is organized as follows. Section II presents the generic framework of the WZ-ERVC scheme. The joint source-channel (WZ coding) RDO-MS and the online correlation noise modeling problems are studied and developed in Sections III and IV, respectively. Section V summarizes the major steps of the proposed scheme. Extensive experimental results are validated in Section VI. Finally, Section VII concludes this paper.

II. OVERVIEW OF THE PROPOSED SCHEME

Fig. 1 shows the bit-stream structure of the proposed encoding scheme. After a conventional predictive video coding (e.g., the state-of-the-art H.264/AVC), a subset of the P-frames, called *anchor frames* (e.g., frame X_T and X_{2T} in Fig. 1), are protected by embedded WZ coding with a joint source-channel R-D optimized mode selection (JSC-RD-MS). Frames X_{nT+1} to $X_{(n+1)T}$ form a recovery unit, and its length is called recovery unit length (RUL) hereafter. We use the group of pictures (GOP) structure of “I-P-P-P- . . .” where only the first frame is encoded as Intra (I) frame and the subsequent frames are all encoded as Inter (P) frames. Transmission errors which occur between frames X_{nT+1} and $X_{(n+1)T}$ are supposed

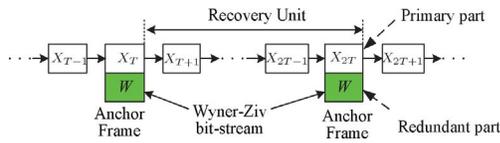
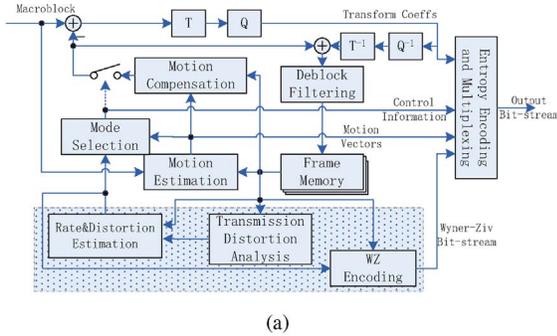
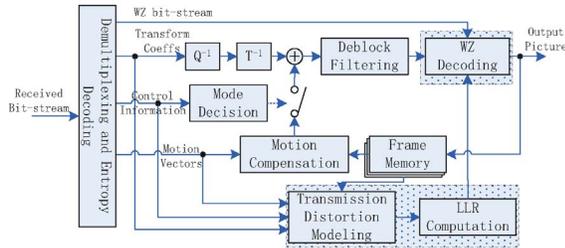


Fig. 1. Bit-stream structure of the proposed WZ-ERVC scheme.



(a)



(b)

Fig. 2. Encoding and decoding diagrams of the proposed WZ-ERVC scheme. (a) Encoder. (b) Decoder.

to be eliminated by the *anchor frame* $X_{(n+1)T}$. At the decoder side, the primary bit-stream of *anchor frames* is first decoded. If the *anchor frame* is error free, the embedded WZ bit-stream will be discarded and the WZ decoding process will be skipped. Here, “error free” implies that not only the specific *anchor frame* is correctly decoded but also it is not corrupted by errors propagated from its previous frames. Otherwise, the proposed WZ decoding with an online correlation noise model will enable us to correct errors and stop error propagation.

Fig. 2 shows the codec structure of *anchor frames* where WZ coding is integrated into standard H.264/AVC architecture.

A. Encoding Process

The encoding architecture of *anchor frames* is shown in Fig. 2(a), while the encoding of non-*anchor frames* is identical with the standard H.264/AVC encoder engine. All the frames will be partitioned into slices and MBs, and the cost of different coding modes (e.g., Intra modes and Inter modes) of each MB will be calculated. During the mode selection process, the expected bit-rate and residual distortion of WZ decoding are taken into account in Section III. After all the MBs in the *anchor frame* are coded with the H.264/AVC engine, the WZ rate $R_{WZ}(n)$ of each MB which is estimated in the “rate and distortion estimation” module, will be summed to determine the coding rate of embedded WZ description in the “WZ encoding” module. It is worth mentioning that

the actual transmission error pattern is unknown but the error free decoded video frame is available. Thus, the co-located “transmission distortion analysis” module is to analyze the expected transmission distortion for a packet loss ratio [21].

In WZ coding, the reconstructed *anchor frame* from H.264/AVC encoder is partitioned into 4×4 blocks, and each block is transformed with 4×4 DCT transform. DCT coefficients at the same position are grouped to form a coefficient band. The DC band is quantized by uniform scalar quantizer, and AC bands are quantized by scalar quantizer with a deadzone around zero. Bit planes are extracted from the quantized DCT coefficient bands and fed into the Slepian-Wolf coder which utilizes the LDPCA [22] to generate the WZ bit-stream.

The two important steps during the encoding process: rate-distortion estimation for the WZ coding and joint source (H.264/AVC) and channel (WZ) RD-optimized mode selection, will be presented in Section III.

B. Decoding Process

The decoding structure of *anchor frames* is shown in Fig. 2(b), while the remaining non-*anchor frames* are the same as Fig. 2(b) except the “LLR computation” and “WZ decoding” module. After the corresponding bit-stream of a frame is received, the H.264/AVC decoding is performed. Once transmission errors occur, error concealment will enable. Considering that the actual transmission error pattern is available, but the error free video frame is inaccessible. In the “transmission distortion modeling” module at the decoder side, therefore, transmission errors occurring within this frame and/or propagated from preceding frames will be modeled for the specific transmission error pattern.

Within the proposed transmission error model, soft *a priori* likelihood information is computed in the “LLR computation” module using the conditional probability distribution partitioning algorithm [23], as shown in Section III-C1. With decoded bit-planes, minimum mean square error (MMSE) reconstruction algorithm is used to generate the output [24].

III. JOINT SOURCE-CHANNEL R-D OPTIMIZED MODE SELECTION

Essentially, the proposed scheme is a joint source channel coding (JSCC) scheme, where the bit-stream of an *anchor frame* is composed by not only the H.264/AVC bit-stream but also the WZ protective bit-stream. Although encoding an MB with Inter mode can achieve better RD performance than with Intra mode in error-free environment, it suffers larger transmission distortion and thus requires more WZ bits to correct the transmission error. Therefore, the mode selection process for *anchor frames* should take into account both the source coding and the WZ coding. Like a typical JSCC scheme [27], [28], the total rate and the end-to-end distortion are analyzed in the R-D optimized mode selection (RDO-MS) process.

As joint source channel rate-distortion optimization schemes [27], [28] and error resilient video coding schemes [9], [10] which have been widely assumed in the literature, we only

consider the joint source channel optimization in application layer. According to H.264/AVC [34], the network abstraction layer (NAL) adapts the generated bit stream to various network and channel environments. It covers all syntactical levels above the slice level, and a NAL unit is generally composed of one slice with the same number of MBs [27]. For IP-based packet-switched communication, RTP has been chosen for media transport. To be concrete, the bit-stream is encapsulated into a RTP packet, and the packets are randomly and independently erased in transmission.

Before presenting the detailed JSC-RDO-MS algorithm for the proposed scheme, we first analyze the RD characteristic of the WZ representation and discuss the impact of source and channel property on the WZ rate for a first-order Markov Gaussian source.

A. Analysis on a Stationary First-Order Markov Gaussian Source

Under the context of motion compensation (MC) in predictive video coding, a video sequence could be modeled as a Markov source. Typically, it is reasonable to model the standardized prediction as a first-order Markov source.

Let us consider a zero-mean stationary first-order Markov source $\{X_n\}_{n \in \mathbb{Z}}$. It is modeled by a linear predictor $X_n = \rho \hat{X}_{n-1} + W_n$, where $|\rho| < 1$ is the correlation coefficient between successive frames, W_n is the prediction residual which follows an i.i.d. zero mean Gaussian distribution, and is independent of X_n . In this model, $\rho \hat{X}_{n-1}$ is the best unbiased estimate of X_n given \hat{X}_{n-1} . In a predictive encoder, X_n is predicted from \hat{X}_{n-1} , instead of X_{n-1} to achieve perfect synchronization between the encoder and decoder. If part of one frame is lost in transmission, the decoder conceals this error with $\tilde{\rho} \hat{X}_{n-1}$, where $\tilde{\rho}$ is an estimate of ρ and the estimation error $\rho_e = \rho - \tilde{\rho}$ follows zero mean Gaussian distribution, \hat{X}_{n-1} is the decoded picture at the decoder side, it may be corrupted by transmission errors.

As Section II, an MB in current *anchor frame* X_n would be protected by embedded WZ coding. If transmission error exists, the side-information Y_n for WZ decoding would be the expectation of current frame based on its previously decoded frames $E\{X_n | \hat{X}_{n-1}, \hat{X}_{n-2}, \dots\}$, i.e., $E\{X_n | \hat{X}_{n-1}\}$ with the first-order Markov model. Without loss of generality, we have

$$Y_n = \tilde{X}_n = \begin{cases} \rho \hat{X}_{n-1} + W_n, & \text{with prob. } 1-p \\ E\{X_n | \hat{X}_{n-1}\}, & \text{with prob. } p \end{cases} \quad (2)$$

where p is the packet loss ratio. Each slice is assumed to be encapsulated into a packet for transmission, and p is the slice loss ratio. In addition, each slice has the same loss probability in IP-based packet-switched transmission and p is also the MB loss ratio.

If an MB in *anchor frame* X_n is originally coded with Inter mode, the expected transmission distortion is

$$D_c^P(n) = p \frac{1 - \rho^{2T}}{1 - \rho^2} \left(\sigma_{\rho_e}^2 E\{\tilde{X}^2\} + \sigma_w^2 \right) \quad (3)$$

where T is the RUL, i.e., the interval between two successive *anchor frames*, $\sigma_{\rho_e}^2$ is the variance of the correlation coefficient

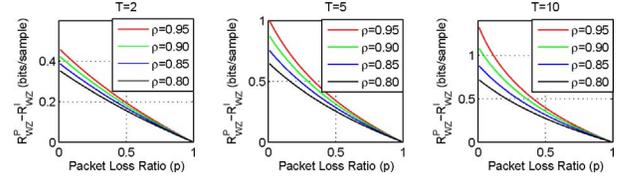


Fig. 3. WZ bit-rate saving for Intra mode compared with Inter mode.

estimation error ρ_e , and σ_w^2 is the variance of prediction residual. If it is originally coded with Intra mode, the expected transmission distortion is

$$D_c^I(n) = (p + p^2 \rho^2 \frac{1 - \rho^{2T-2}}{1 - \rho^2}) \left(\sigma_{\rho_e}^2 E\{\tilde{X}^2\} + \sigma_w^2 \right). \quad (4)$$

For the detailed derivations of (3) and (4), please refer to the appendix.

With regard to WZ coding for an MB in *anchor frames*, its R-D function can be expressed, under mean square error (MSE), by [25]

$$R_{WZ}(D) = \frac{1}{2} \log^+ \left(\frac{D_c}{D_r} \right) \quad (5)$$

where $\log^+(x) = \max\{\log(x), 0\}$, D_c is the transmission distortion, and D_r is denoted as the residual transmission distortion after WZ decoding. In this sense, to achieve the same D_r , the MB in *anchor frames* with Inter mode would require more WZ bits than Intra mode, namely, $R_{WZ}^P(D) > R_{WZ}^I(D)$ because

$$\begin{aligned} & D_c^P(n) - D_c^I(n) \\ &= \frac{(1-p)\rho\rho^2(1-\rho^{2T-2})}{1-\rho^2} \left(\sigma_{\rho_e}^2 E\{\tilde{X}^2\} + \sigma_w^2 \right) > 0. \end{aligned} \quad (6)$$

To better understand the impact of source and channel characteristic on the WZ bit-rate, we implement numerical simulations in Fig. 3 to demonstrate the difference between $R_{WZ}^P(D)$ and $R_{WZ}^I(D)$ to achieve the same D_r . In simulation, $E\{X^2\}$ is set to 25 000 and $\sigma_{\rho_e}^2$ is set to 0.1, σ_w^2 is derived according to ρ . The parameters are set for *Foreman* sequence with CIF (352 × 288) resolution. It could be seen that different source and channel factors would induce different WZ bit-rate saving performance: a larger recovery unit length T , and a larger successive frame correlation coefficient ρ , or a smaller packet loss ratio p will result in a larger WZ bit-rate saving. From source coding perspective, an MB with Inter mode behaves much better RD performance than Intra mode. Thus, a tradeoff in mode selection needs to be made by jointly optimizing source coding and WZ coding.

B. Joint Source-Channel RDO-MS

In hybrid video coding schemes, a number of candidate coding modes are available for each MB. The RD-optimized mode selection (RDO-MS) can be posed as a well-known Lagrangian optimization problem [26]

$$J_s(k, m) = D_s(k, m) + \lambda R_s(k, m) \quad (7)$$

where the Lagrange multiplier λ finds the tradeoff between the source-coding distortion $D_s(k, m)$ and the encoding rate

$R_s(k, m)$, k is MB index and m indicates different coding modes. To reflect the impact of lossy transmission channel, joint source-channel RDO-MS (JSC-RDO-MS) analysis [27], [28] adopts the end-to-end distortion $D_{sc}(k, m)$ into (7)

$$J_{sc}(k, m) = D_{sc}(k, m) + \lambda R_s(k, m) \quad (8)$$

where $D_{sc}(k, m) = D_s(k, m) + D_c(k, m)$ since $D_s(k, m)$ and transmission distortion $D_c(k, m)$ are uncorrelated with each other [27].

For the proposed JSC-RDO-MS process, the Lagrangian cost of the WZ representation is

$$J_{WZ}(k, m) = D_s(k, m) + D_{WZ}(k, m) + \lambda (R_s(k, m) + R_{WZ}(k, m)) \quad (9)$$

where $D_{WZ}(k, m)$ denotes the residual transmission distortion after WZ coding, and $R_{WZ}(k, m)$ is the WZ bit-rate. Since the source distortion $D_s(k, m)$ and source rate $R_s(k, m)$ in (9) can be estimated with traditional rate and distortion modeling algorithms [27], [29], and there are a variety of algorithms to estimate the transmission distortion $D_c(k, m)$ [21], [27], we will focus on the problem about how to estimate $D_{WZ}(k, m)$ and $R_{WZ}(k, m)$.

C. Bit-Rate Estimation for WZ Video Coding

Because the channel codes (e.g., Turbo codes and LDPC codes) in WZ coding often require a long code length which typically involves with multiple MBs, the corresponding bit-rate and distortion of an MB is not straightforward available. In this paper, we estimate the WZ bit-rate based on the error correction capacity of channel codes.

As we know, the WZ coding bit-rate is determined by the correlation noise level between source X and its side information Y , i.e., the transmission distortion $D_c(k, m)$ in the context of this paper. Specifically, $D_c(k, m)$ is approximated by the maximum transmission distortion of possible transmission error patterns, that is

$$D_c(k, m) = \max_l \{D_c(k, m; l)\} \quad (10)$$

where $D_c(k, m; l)$ is the transmission distortion with error pattern l . Without loss of generality, the number of lost slices in error pattern l is assumed to be L , which would induce $\sum_{i=L+1}^S C_S^i p^i (1-p)^{S-i}$, the probability that more than L packets loss between two successive *anchor frames*, to be less than a margin p_t , e.g., 0.05. S is the total slice number between two successive *anchor frames*. Given the error concealment algorithm, the transmission distortion $D_c(k, m; l)$ for a specific error pattern l can be computed at the encoder side.

1) *Compute the Conditional Entropy of Bit-Planes*: Since the exact value of $H(X|Y)$ derived at the decoder side is unavailable at the encoder side, it is approximated by its expectation $H^E(X|Y)$. Since WZ coding operates on a bit plane basis, the encoding rate is estimated for each bit-plane. $H^E(X|Y)$ of the t th bit-plane can be computed by

$$H_t^E(X|Y) = \sum_{y \in Y} p(y) H(b_t^x | \mathcal{H}) \quad (11)$$

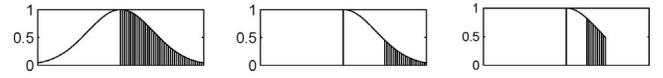


Fig. 4. Illustration of the computation procedure of the probability $p(b_t^x | \mathcal{H})$.

where \mathcal{H} represents available information at the decoder side, specifically, previously decoded bit-planes $\{b_{N-1}^x, \dots, b_{t+1}^x\}$ and side-information y , N is the number of bit-planes and $t \in [0, N-1]$. $p(e) = Ae^{-\alpha|e|}$, where A is a normalization factor, $e = x - y$ is the transmission error, and $\alpha = \sqrt{2/D_c(k, m)}$. Typically, e is independent to x and y , and $p(x|y)$ and $p(y|x)$ can be derived from $p(e)$, i.e., $p(x|y) = p(y + e|y)$ and $p(y|x) = p(x - e|x)$. $H(b_t^x | \mathcal{H})$ is the conditional entropy of the t th bit-plane of x , given available information \mathcal{H}

$$H(b_t^x | \mathcal{H}) = - \sum_{b_t^x \in \{0, 1\}} p(b_t^x | \mathcal{H}) \cdot \log_2 (p(b_t^x | \mathcal{H})) \quad (12)$$

where $p(b_t^x | \mathcal{H})$ is the conditional probability of b_t^x given \mathcal{H} .

To obtain $p(b_t^x | \mathcal{H})$, we adopt the *a priori* probability partition algorithm [23], where the range of conditional probability distribution $p(b_t^x | \mathcal{H})$ is equally divided into two parts whose areas represent the probability that the coming bit b_t^x equals to 0 and 1, respectively. As shown in Fig. 4, the area of the white and shadow region stands for the probability that b_t^x equals to 0 and 1, respectively, given information \mathcal{H} . Assuming that the first and the second bit-plane are 1 and 0, three subfigures in Fig. 4 from left to right correspond to the probability calculation for the first (the most significant), the second, and the third bit-planes. Specifically, to obtain $p(b_t^x | \mathcal{H})$ of the first bit-plane, the entire range of X is partitioned into two even parts. Thus, the probability of $p(b_t^x = 0 | \mathcal{H})$ can be derived from the area of $[0, 2^{N-1} - 1]$, and the area of the remaining part as the probability of $p(b_t^x = 1 | \mathcal{H})$. Since the first bit-plane is supposed to be 1, the value range of x after the first bit-plane turns to $[2^{N-1}, 2^N - 1]$; for the second bit-plane, the probability of $b_t^x = 0$ and $b_t^x = 1$ can be attained by the area of $[2^{N-1}, 2^{N-1} + 2^{N-2} - 1]$ and $[2^{N-1} + 2^{N-2}, 2^N - 1]$, respectively; for the third bit-plane, the range turns to be $[2^{N-1}, 2^{N-1} + 2^{N-3} - 1]$ and $[2^{N-1} + 2^{N-3}, 2^{N-1} + 2^{N-2} - 1]$.

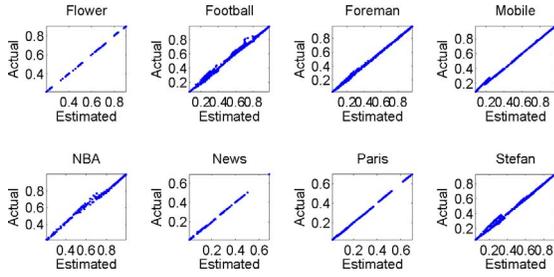
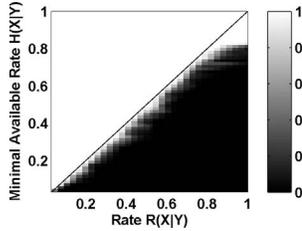
In summary, $p(b_t^x | \mathcal{H})$ for DC coefficients can be obtained by

$$p(b_t^x = 0 | \mathcal{H}) = \sum_{j=t+1}^N \sum_{i=0}^{(2^j-1)+\sum_{k=t+1}^j b_k^x 2^k} p(x|y) dx \quad (13)$$

$$p(b_t^x = 1 | \mathcal{H}) = \sum_{j=t+1}^N \sum_{i=2^j+\sum_{k=t+1}^j b_k^x 2^k}^{(2^{j+1}-1)+\sum_{k=t+1}^j b_k^x 2^k} p(x|y) dx. \quad (14)$$

Since there is no previous bit-plane for the most significant bit-plane, b_N^x is set to 0 in (13) and (14).

For AC coefficients, a uniform threshold quantizer is used [13]. With this quantizer, all the coefficient values $|x| \leq Q_{\text{step}}$ are quantized to 0, where Q_{step} is the quantization step. The *a priori* probability of the sign bit-plane b_s is calculated by $p(b_s = +1) = p(x > Q_{\text{step}} | y)$ and $p(b_s = -1) = p(x \leq Q_{\text{step}} | y)$,

Fig. 5. Estimation precision of conditional entropy $H(X|Y)$.Fig. 6. Correlation of a combination of $H(X|Y)$ and P_e versus the WZ bit-rate $R(X|Y)$.

respectively. The following bit-planes can be attained with (13) and (14) by setting N to 9 and replacing x and y by $|x|$ and $|y|$.

Fig. 5 shows the $H(X|Y)$ estimation results of several test sequences in CIF resolution. In simulation, the first 250 frames of each sequence and the five most significant bit-planes of each coefficient band are considered. In Fig. 5, x -axis is the actual $H(X|Y)$ obtained at the decoder side with the standardized WZ decoding, and y -axis is the estimated $H^E(X|Y)$ according to (11) at the encoder side. The results show that the estimation precision is very high. Although, the results are relatively less accurate for sequences with intensive motion, e.g., *NBA* and *Stefan*, an appropriate excessive bit-rate is generally allocated for applications without feedback channel to ensure successful decoding, as will be shown in Section III-C2. This excessive bit-rate can eliminate the effect of the $H(X|Y)$ estimation inaccuracy and ensures that the WZ bit-planes are decodable at the decoder side.

To better demonstrate the estimation results, we also compute the correlation coefficients between the actual and estimated $H(X|Y)$ for these sequences in Table I.

2) *Estimate the Encoding Rate*: Since WZ coding is involved with channel codes [22], [30], [31], e.g., LDPC codes, we are to investigate error correction capability of channel codes. For a given channel, although more redundant parity check bits increases the probability of correcting transmission errors, it also decreases the RD efficiency. The estimation of encoding rate is to infer the minimum bit-rate of WZ coding at which the decoder can achieve a predefined probability of the successful decoding. In this paper, we estimate the WZ encoding rate $R(X|Y)$ based on the relationship between $R(X|Y)$ and P_e for a given $H(X|Y)$.

Fig. 6 illustrates the relationship of a combination of $R(X|Y)$ and P_e versus $H(X|Y)$ for the Slepian–Wolf codec, where the gray level of each point represents the decoding failure

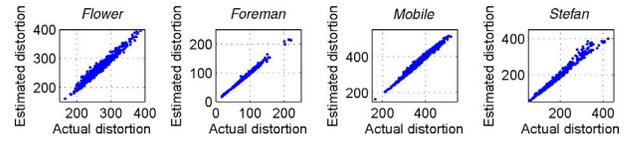


Fig. 7. Estimation precision of residual transmission distortion.

probability P_e . For a specific $H(X|Y)$, i.e., one row, P_e decreases with the increase of $R(X|Y)$. When P_e decreases to a predefined threshold, the corresponding $R(X|Y)$ is chosen as the desired encoding rate. The LDPCA approach [22], [32] with degree distribution in (15) is used in the WZ coding over a variety of the training test sequences, and can produce a set of coding rates: $\frac{2}{66}, \frac{3}{66}, \frac{4}{66}, \dots, \frac{66}{66}$

$$\lambda(x) = 0.316x^1 + 0.415x^2 + 0.128x^6 + 0.069x^7 + 0.020x^{18} + 0.052x^{20}. \quad (15)$$

D. Distortion Estimation for WZ Video Coding

Supposing that $P_e^t (t = 0, \dots, N-1)$ is the predefined decoding failure probability of the t th bit-plane, $D_{WZ}(k, m)$, the expected residual transmission distortion after WZ decoding, can be expressed by

$$D_{WZ}(k, m) = P_e^0 D_c(k, m) + \sum_{b=0}^{15} \left\{ \sum_{t=1}^{N-1} \alpha_t D_{WZ}^{b,t}(k, m) \right\} \quad (16)$$

where $D_c(k, m)$ is the transmission distortion, $D_{WZ}^{b,t}(k, m)$ is the residual transmission distortion after t WZ bit-planes of band b have been correctly decoded, k is the MB index, and m is the encoding mode. α_t is the probability that only the t th WZ bit-plane can be correctly decoded, and can be calculated by $\alpha_t = P_e^{t+1} \prod_{i=0}^t (1 - P_e^i)$, ($t = 0, \dots, N-2$) and $\alpha_{N-1} = \prod_{i=0}^{N-1} (1 - P_e^i)$.

In the state-of-the-art WZ video coding schemes, the MMSE inverse quantizer is used [24], where $D_{WZ}^{b,t}(k, m)$ in (16) can be estimated by

$$D_{WZ}^{b,t}(k, m) = \sum_{x \in [X_L^t, X_U^t]} \sum_{y \in Y} (x - \tilde{x})^2 \cdot p(y|x) \quad (17)$$

where $[X_L^t, X_U^t]$ is the quantization range of x which is related to the number of correctly decoded bit-planes. The more bit-planes are decoded, the smaller the quantization range is. \tilde{x} and $p(y|x)$ can be derived from Section III-C1.

Fig. 7 shows the estimated distortion from four test sequences where 20 different channel traces are tested. The distortion is measured with MSE and normalized by the size of MB. It can be seen that the proposed method is capable of accurately estimating the transmission distortion.

E. Summary of JSC-RDO-MS

In H.264/AVC standard, an MB can be divided into 16×16 , 16×8 , 8×16 , and 8×8 blocks, and each 8×8 block can be further divided into 8×8 , 8×4 , 4×8 , and 4×4 sub-blocks for improved motion prediction [33]. Considering there exist a plenty of coding modes in the H.264/AVC standardization, an extremely high computation complexity is required for the

TABLE I
CORRELATION COEFFICIENT BETWEEN ACTUAL AND ESTIMATED $H(X|Y)$

Test Sequences	<i>Flower</i>	<i>Football</i>	<i>Foreman</i>	<i>Mobile</i>	<i>NBA</i>	<i>News</i>	<i>Paris</i>	<i>Stefan</i>
Correlation coefficient	0.99989	0.99782	0.99964	0.99967	0.99826	0.99981	0.9991	0.9973

joint H.264/AVC and WZ mode selection. As follows, we adopt a sub-optimal mode selection algorithm which chooses the desired coding mode from the best Intra mode and the best Inter mode derived in the standardized H.264/AVC engine.

- 1) Compute the cost of Intra and Inter modes with the traditional video coding engine.
 - a) Compute the cost of Intra modes m_I for MB k , denoted as $J(k, m_I)$, and the Intra mode with the minimum cost as m_I^* .
 - b) Compute the cost of Inter modes m_P for MB k , denoted as $J(k, m_P)$, and the Inter mode with the minimum cost as m_P^* .
- 2) If $J(k, m_I^*) \leq J(k, m_P^*)$, MB k will be encoded with Intra mode m_I^* .
- 3) If $J(k, m_I^*) > J(k, m_P^*)$, compute the joint source-channel cost according to (9).
 - a) Estimate the WZ bit-rate $R_{WZ}^b(k)$ of each bit-plane for MB k , and the total WZ bit-rate of MB k is $R_{WZ}(k, m) = \sum_b R_{WZ}^b(k)$. The WZ rate is estimated according to the analysis in Section III-C.
 - b) Estimate the residual transmission distortion after WZ decoding $D_{WZ}(k, m)$, as Section III-D.
 - c) Compute the cost $J_{WZ}(k, m)$ for mode m_P^* and m_I^* according to (9).
- 4) If $J_{WZ}(k, m_I^*) \leq J_{WZ}(k, m_P^*)$, MB k will be coded with the Intra mode m_I^* .
- 5) Otherwise, MB k will be coded with the Inter mode m_P^* .

After all MBs in the *anchor frame* are coded, the total WZ bit-rate for each bit-plane can be obtained by $R_{WZ}^b = \sum_k R_{WZ}^b(k)$. The WZ encoding could be performed in the ‘‘WZ encoding’’ module in Fig. 2(a) with the bit-rate R_{WZ}^b .

IV. CORRELATION NOISE MODELING AT THE DECODER SIDE

The traditional error concealment schemes aim to restore the error contaminated frames based on the previously decoded frames and the correctly decoded regions of the current frame. However, they fail to estimate the residual distortion of the reconstructed frames after error concealment. In this paper, we propose a new error concealment algorithm not only to restore the error contaminated frames but also to estimate the residual distortion.

Once a packet is lost, both MVs and the prediction residue are all unavailable. Since the prediction residue is unpredictable from the previous frames, the inter-frame error concealment is to estimate the MVs of the lost blocks according to the spatial smoothness constraint and the correlation of MVs between neighboring blocks [33]. The transmission error e consists of two parts: 1) e_w , the errors caused by the loss of prediction residue, and 2) e_ρ , the errors caused by the

TABLE II
RELATIVE ERROR BETWEEN D_c AND $D_c^\rho + D_c^w$

Sequences	QP				
	16	22	28	34	40
<i>Foreman</i>	6.06%	5.11%	4.13%	3.46%	3.55%
<i>Paris</i>	7.30%	7.01%	6.40%	4.84%	3.45%
<i>Flower</i>	9.91%	9.44%	7.74%	5.00%	2.92%
<i>Mobile</i>	7.51%	7.25%	6.22%	3.45%	1.78%

MV estimation inaccuracy. The distribution of the transmission error e can be expressed by a mixture distribution as shown in (1). For clarity of the context here, it is rewritten as

$$p(e) = \sum_{e_\rho} p(e_w|e_\rho)p(e_\rho). \quad (18)$$

Since e_w follows Laplacian distribution [20], $p(e_w|e_\rho)$ in (18) also follows Laplacian distribution if e_w and e_ρ are uncorrelated. This assumption will be validated later. Thus, $p(e)$ can be modeled by a mixture of Laplacian distributions where each Laplacian distribution $p(e_w|e_\rho)$ is called a *mixture component*, while $p(e_\rho)$ is called the *mixing coefficient* as the weight of each *mixture component*.

Let D_c^ρ be the distortion caused by the error of MV estimation and D_c^w the distortion caused by the loss of prediction residue, the total distortion D_c could be obtained by

$$D_c = E\{(e_\rho + e_w)^2\} = D_c^\rho + D_c^w + 2E\{e_\rho e_w\}. \quad (19)$$

If e_w and e_ρ are uncorrelated, we will get (20) since the prediction residue is zero mean [20]

$$E\{e_\rho e_w\} = E\{e_\rho\}E\{e_w\} = 0. \quad (20)$$

To validate (20), we code a set of test sequences with random transmission errors over 20 random erasure channel trails and the packet loss ratio 8%. The results are shown in Table II, where the relative estimation error between D_c and $D_c^\rho + D_c^w$ is defined as

$$e_r = \frac{1}{T} \sum_{n=1}^T \frac{|(D_c^\rho + D_c^w) - D_c|}{D_c} \times 100 (\%) \quad (21)$$

where T is the total number of frames. e_w is obtained from the error-free H.264/AVC bit-streams. When transmission error occurs, the total transmission distortion e is available at the decoder side. Hence, e_ρ can be obtained by $e_\rho = e - e_w$. With e_ρ , e_w and e , D_c^ρ , D_c^w and D_c can accordingly be obtained. It can be seen that the relative error is small.

Thus, it is reasonable to suppose that e_ρ and e_w are uncorrelated. Accordingly, (18) can be rewritten as

$$p(e) = \sum_{e_\rho} p(e_w)p(e_\rho). \quad (22)$$

For the Laplacian mixture model in (22), it aims to model the *mixture component* $p(e_w)$ and *mixing coefficient* $p(e_\rho)$. Since the prediction residue for the state-of-the-art video coding schemes (e.g., H.264/AVC) is almost a stationary i.i.d. random noise, it can be easily modeled based on the neighboring correctly decoded blocks of either the current or the previous frames. For the packet loss occurring at either *anchor frames* or *non-anchor frames*, we will focus on the modeling of the *mixing coefficient* $p(e_\rho)$ in the pixel domain and the DCT domain.

A. Correlation Noise Modeling in the Pixel Domain

Supposing $x_n(i)$ is the i th pixel of the n th frame and $v_n(i)$ is the corresponding MV of $x_n(i)$, the sub-pixel motion compensation for P-frames is expressed by

$$\begin{aligned} x_n(i) &= x_n^\rho(i) + w_n(i) = f_{MC}(x_{n-1}(i), v_n(i)) + w_n(i) \\ &= \sum_{l=1}^L a_l x_{n-1}(u_l(i + v_n(i))) + w_n(i) \end{aligned} \quad (23)$$

where $x_n^\rho(i)$ is the MC value from the previous frame X_{n-1} , $w_n(i)$ is the prediction residue, and $u_l(i + v_n(i))$ refers to the spatial index of the pixel in the frame X_{n-1} to predict $x_n(i)$. The MC interpolation coefficient a_l satisfies $\sum_l a_l = 1$. When transmission erasure occurs, the error concealment estimates lost MVs and approximates $x_n(i)$ by

$$x_n^\rho(i) = f_{MC}(x_{n-1}(i), v_n'(i)) = \sum_{l=1}^L a_l x_{n-1}(u_l(i + v_n'(i))) \quad (24)$$

where $v_n'(i)$ is an estimate of $v_n(i)$ from error concealment, and the mismatch between $x_n^\rho(i)$ and $x_n(i)$ 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 motion compensation $x_n^\rho(i)$ with $f_{MC}(x_{n-1}(i), v_n'(i) + v_e(i))$. By this means, the distribution of $x_n^\rho(i)$ can be derived from the distribution of corresponding $v_e(i)$, that is

$$p(x_n^\rho(i) = a) = \sum_{f_{MC}(x_{n-1}(i), v_n'(i) + v_e(i)) = a} p(v_e(i)). \quad (25)$$

That is, the probability of $x_n^\rho(i) = a$ equals to the sum of the probability of $v_e(i)$ which could generate a motion-compensated result a , i.e., $f_{MC}(x_{n-1}(i), v_n'(i) + v_e(i)) = a$.

Fig. 8 shows the distribution of v_e for a set of test sequences. The results show that the distribution of v_e for the successive frames is highly correlated. The probability distribution $p(v_e)$ is modeled by a Laplacian distribution $p(v_e) = \frac{\alpha_{v_e}}{2} \exp(-\alpha_{v_e}|v_e|)$, whose parameter α_{v_e} is estimated by error concealment on the preceding frames. With the estimated MV v_n' and actual MV v_n , the decoder can obtain the MV estimation error $v_e = v_n - v_n'$ of the current frame, and then estimate α_{v_e} with $\alpha_{v_e} = \sqrt{2/\sigma_{v_e}^2}$.

After $p(v_e)$ is obtained, we can estimate the probability distribution of $x_n^\rho(i)$ according to (25). Fig. 9 shows a portion of the estimated possible value of $x_n^\rho(i)$ for the *Mobile* sequence, where the blue “o” denotes the error free pixel value $x_n(i)$, the black “x” stands for the error concealment result $x_n^\rho(i)$, the red “+” is the estimated possible value of $x_n^\rho(i)$ at the decoder

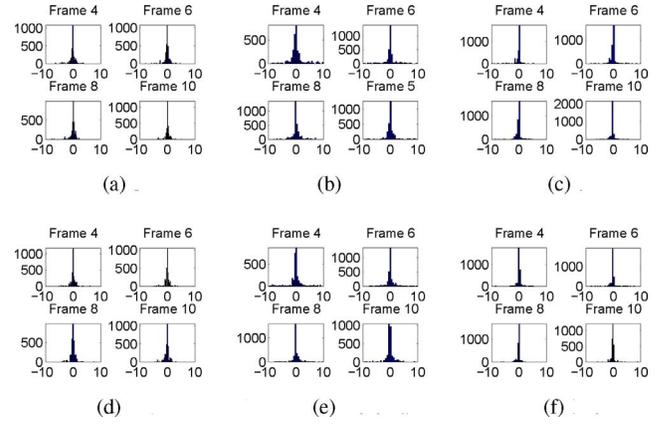


Fig. 8. Distribution of the mismatch v_e for test sequences. (a) *Flower* MVX. (b) *Foreman* MVX. (c) *Mobile* MVX. (d) *Flower* MVY. (e) *Foreman* MVY. (f) *Mobile* MVY.

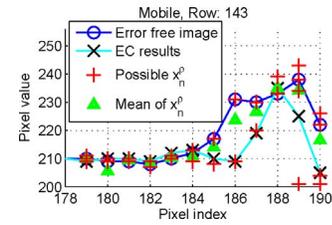


Fig. 9. Estimated possible value of $x_n^\rho(i)$.

side, and the green “▲” is the expectation of $x_n^\rho(i)$. It can be seen that the proposed algorithm is able to accurately predict the error-free results.

By taking the expectation of $x_n^\rho(i)$ as the concealed result, the algorithm is validated for error concealment. In the experiments, Chessboard FMO is turned on, one frame is divided into four slice groups and each slice group is encapsulated into a packet, and the transmission channel is modeled by a random erasure channel where 20 different random erasure channel trials with packet loss ratio 8% are tested. As Table III, it can be seen that the proposed algorithm achieves about 2–3 dB PSNR gain compared with the “Frame copy” and “Motion Copy” methods in H.264/AVC.

1) *Effect of Error Propagation*: Once the current frame occurring transmission errors is a *non-anchor frame*, the effect of motion-compensated prediction and sub-pixel displacement vectors should be iteratively analyzed to model the error propagation process. Under the context, $p(x_n^\rho(i))$ can be given by

$$\begin{aligned} p(x_n^\rho(i) = b) &= \sum_{f_{MC}(x_{n-1}(i), v_n(i)) = b} \left(\prod_{l=1}^L p(x_{n-1}^\rho(u_l(i + v_n(i)))) \right). \end{aligned} \quad (26)$$

If transmission error occurs in the current frame, the corresponding transmission error needs to be estimated according to (25) and integrated with errors propagated from the previous frames obtained by (26).

B. Correlation Noise Modeling in DCT Domain

Considering the WZ coding is typically implemented in the DCT domain, we also model the correlation noise on the DCT coefficients with the Laplacian mixture model in (22).

TABLE III
PSNR COMPARISON OF DIFFERENT ERROR CONCEALMENT ALGORITHMS

Sequences		Flower			Foreman			Mobile			Stefan		
QP		22	28	34	22	28	34	22	28	34	22	28	34
PSNR (dB)	Frame Copy	20.58	20.57	20.53	26.09	26.04	25.94	19.00	19.00	19.05	19.82	18.83	18.84
	Motion Copy	26.91	26.47	25.43	31.05	30.70	29.76	24.76	24.80	24.09	26.47	26.25	25.73
	The proposed	31.87	30.87	28.44	34.35	33.65	31.98	30.04	29.51	27.59	28.87	28.76	27.78

1) *Estimate the Mixing Coefficient*: As we know, DCT transform is a linear transform and the DCT coefficient c_ρ can be expressed as a linear combination of pixels in the corresponding M-by-M block by

$$c_\rho = \sum_{i \in \mathcal{B}} b_l x^\rho(i) \quad (27)$$

where $l(i)$ is the index of pixels in the M-by-M block \mathcal{B} , $0 \leq l \leq M^2 - 1$, and b_l is the l th element of DCT basis function. The probability distribution of DCT coefficient c_ρ is then given by

$$p(c_\rho = c) = \sum_{i \in \mathcal{B}} \sum_{b_l x^\rho(i) = c} \left(\prod_{i \in \mathcal{B}} p(x_n^\rho(i)) \right) \quad (28)$$

where $p(x_n^\rho(i))$ is obtained from (25) and (26), and $c_\rho = \sum_{i \in \mathcal{B}} b_l x_n^\rho(i)$ is the DCT coefficient value after motion compensation with $v'_n(i) + v_e(i)$. Thus, the mixing coefficient in (22) can be attained from (28).

2) *Estimate the Mixture Components*: In general, the prediction residual can be modeled by a zero-mean additive Laplacian distribution $p(w) = \frac{\alpha_w}{2} \exp(-\alpha_w |w|)$, where

$$\alpha_w = \sqrt{2/\sigma_w^2} \quad (29)$$

and σ_w^2 is the variance of the prediction residue [15]. σ_w^2 of the lost blocks can be estimated on average from $\{\sigma_w^2(n)\}$ of the correctly decoded neighboring blocks and that of the co-located blocks in previous frames.

Fig. 10 shows the estimated PDF for each DCT coefficient of a 4×4 block, where the blue solid curve and the black dash-and-dot curve denote the estimated $p(c|c_s)$ with both the proposed mixture Laplace model and the Laplacian model, respectively. The red line indicates the error-free value c , the green lines label the quantization interval Δ and the quantization step size $|\Delta|$ is 16. It can be seen that the probability $p(c \in \Delta)$ of the proposed model is greater than the traditional Laplacian model. That means, fewer bits are required to correct the transmission-induced error, and leads to a smaller WZ bit-rate. Furthermore, since an MMSE inverse quantizer reconstructs the coefficient c using the centroid of $p(c)$, $c \in \Delta$, the reconstructed frame with the proposed model would be better than the traditional Laplacian model. These arguments will be validated with experiments.

V. SUMMARY OF THE PROPOSED SCHEME

In this section, we summarize the major steps of the encoding/decoding process in the proposed scheme.

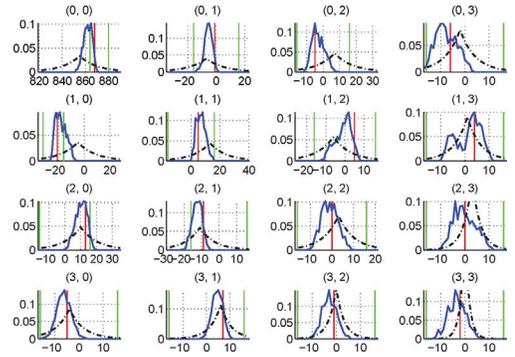


Fig. 10. Estimated PDF $p(c|c_s)$ for DCT coefficients of a 4×4 block.

A. Encoder Side

Case 1 If the current frame is not an *anchor frame*, encode it with the conventional H.264/AVC encoder. Simulate the error concealment of the decoder side to obtain the transmission distortion according to the targeted packet loss ratio in transmission.

Case 2 If the current frame is an *anchor frame*, the JSC-RDO-MS algorithm and the estimation of WZ bit-rate are implemented according to Section III-E. After all MBs in the *anchor frame* are encoded, perform WZ coding for bit-plane b with the encoding rate $R_{WZ}^b = \sum_k R_{WZ}^b(k)$.

B. Decoder Side

Step 1 Decode the received bit-stream with the traditional video decoder, and operate the virtual error concealment on the correctly decoded slices to obtain the distribution of MV estimation error $p(v_e)$.

Step 2 If transmission errors occur in the current frame, estimate transmission errors with (25) in Section IV-A. If the previous frames are contaminated by transmission errors, analyze the effect of motion compensation by (26) in Section IV-A, and integrate with Step 1.

Step 3 If the current frame is an *anchor frame*, compute the probability distribution of DCT coefficients with a Laplacian mixture model whose parameters are estimated by (28) and (29) in Section IV-B.

Step 4 According to the estimated transmission error, the soft information required by WZ decoding can be calculated, and then the WZ decoding can be implemented to correct the transmission error.

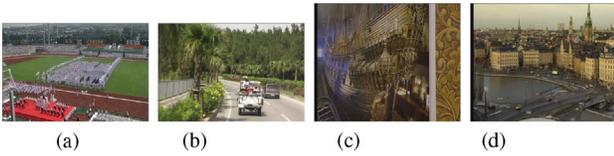


Fig. 11. Sampled frames of test sequences with a large resolution (the tenth frame). (a) *Open Ceremony*. (b) *Driving*. (c) *MobCal*. (d) *Stockholm*.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we sufficiently evaluate the performance of the proposed scheme. The memoryless random erasure channel is considered since this type of channels is common in digital communications. For example, the Internet can be modeled as a random erasure channel where packet loss is a random erasure error. Another example is the digital channel protected with FEC codes since a decoding failure in FEC will result in an erasure.

In our experiments, H.264/AVC (JM 12.2/High Profile) codec is adopted to generate the primary bit-stream. The maximum search range is set to 32, and the entropy coding is CABAC. The format of test sequences is YUV 4:2:0 with CIF (352×288) at 15 frames/s. To validate the performance of the proposed scheme on sequences with larger resolution, we also perform experiments on test sequences with 720×480 (*Driving* and *Open Ceremony*) and 720×576 (*MobCal* and *Stockholm*) resolution which are shown in Fig. 11. The sequences are coded with “I-P-P-P-A-P-...” where “A” denotes an *anchor frame*. The RUL, i.e., the interval between two successive *anchor frames*, is set to 5. A chessboard FMO pattern is turned on. There are four slice groups in one frame, and the data of one slice group is encapsulated into one data packet for transmission. With this setup, a packet loss in transmission will lead to the loss of one slice group. If not specified, the quantization parameter (QP) is set to 28 for Intra and Inter frame, the packet (slice) loss rate is set to 8%, and 20 channel trials are tested in experiments. The side-information for WZ decoding is obtained by the error concealment approach referring to Section IV-A.

The quantization step (Qstep) is derived from a quantization parameter for WZ coding (QPW). The QPW range is 4–51, and QPW value 4 corresponds to Qstep 1. The Qstep doubles in size for every increment of 6 in QPW. At the decoder side, an MMSE reconstruction approach is used [24]. Slepian–Wolf codec adopts the LDPCA [22] approach with a block length of 6336 bits and degree distribution in (15).

A. Performance Gain of Each Module

To evaluate the efficiency of the components in the proposed scheme, we provide the individual performance on averaged *anchor frames*.

1) *Performance Gain by the Non-Stationary Correlation Noise Modeling*: Fig. 12 compares the RD performance of *anchor frames* of four CIF test sequences with the proposed Laplacian mixture model and the traditional Laplacian model. It can be seen that the mixture of Laplacian distributions significantly improves the RD performance up to 2.2 dB, especially at lower bit rates. To be concrete, more accurate

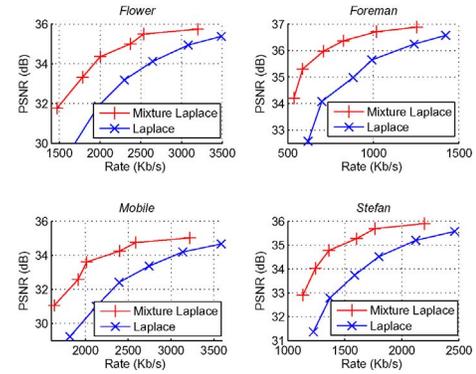


Fig. 12. Performance gain from the proposed non-stationary correlation noise model.

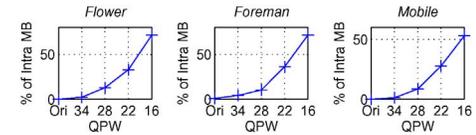


Fig. 13. Intra MB percentage in *anchor frames*.

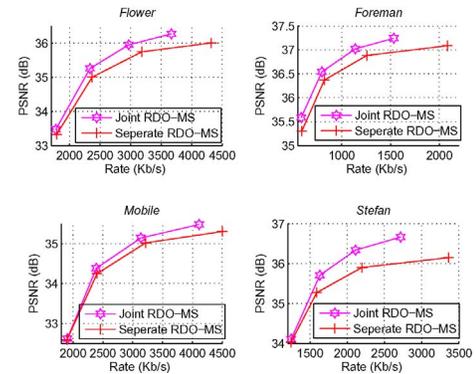


Fig. 14. Performance gain of joint source-channel RDO-MS.

model could correct the same amount of transmission error with fewer bits and reconstruct the frame with lower distortion. The more accurate correlation noise model will also favor the MMSE inverse quantizer at the lower bit-rate region. At the high bit-rate region, the decoded bit-planes have mitigated the effect of inaccurate correlation noise model.

2) *Effect of Joint RDO-MS*: Fig. 13 shows the percentage of Intra MBs in *anchor frames* with different WZ quantization parameter (QPW) when joint RDO-MS is adopted. It reflects that the proportion of Intra MB in *anchor frames* increases with the decreasing of QPW.

Fig. 14 shows the RD curve of *anchor frames* with and without the proposed joint RDO-MS algorithm which takes both the bit-rate and distortion of WZ coding into consideration. For the “Separate RDO-MS,” *anchor frames* are coded and protected with WZ coding. Its mode selection does not count the involved WZ bit-rate and the end-to-end distortion. It can be seen that the proposed joint RDO-MS algorithm achieves a better RD performance than the “Separate RDO-MS,” and the performance gain increases with the number of bit-rates.

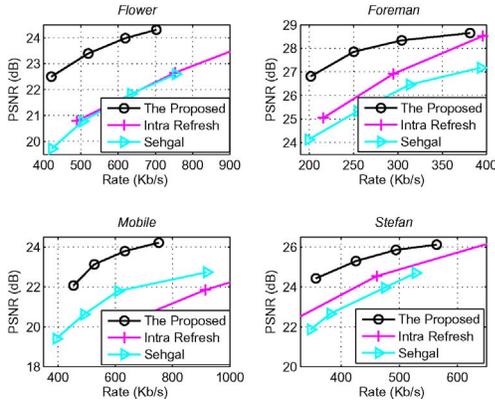


Fig. 17. RD performance of *anchor frames* with packet loss ratio 8% (low bit-rate environments).

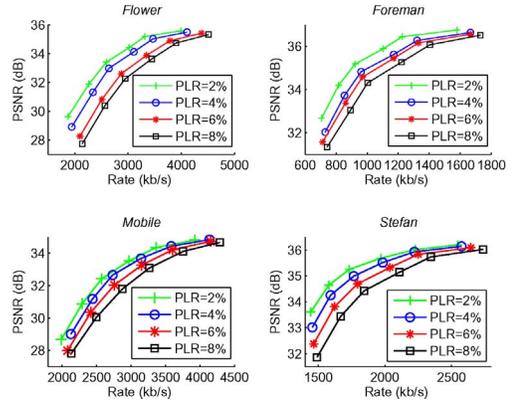


Fig. 20. RD performance of *anchor frames* with different packet loss ratio (PLR).

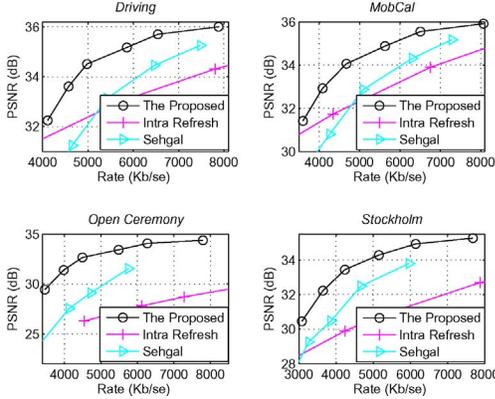


Fig. 18. RD performance of *anchor frames* with packet loss ratio 8% (high resolution).

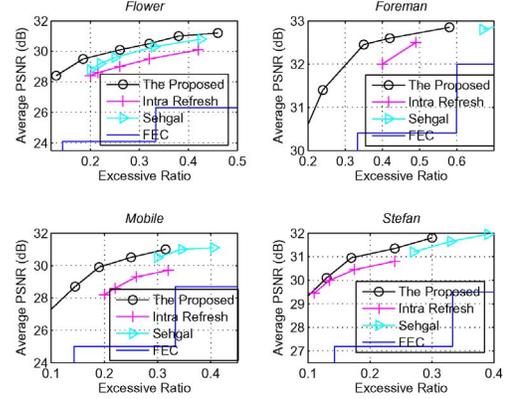


Fig. 21. Averaged PSNR of the decoded sequences as a function of the excessive ratio.

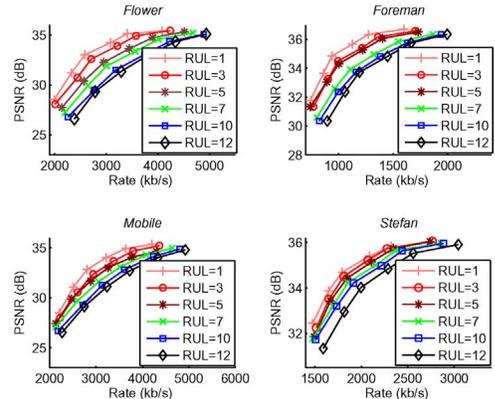


Fig. 19. RD performance of *anchor frames* with different recovery unit length (RUL).

2) *The RD Performance of Anchor Frames with Different Recovery Unit Length:* With different RUL, Fig. 19 shows the RD performance of *anchor frames*. Longer RUL would induce larger transmission distortion at *anchor frames* and thus worse RD performance of *anchor frames*.

3) *The RD Performance of Anchor Frames with Different Packet Loss Ratio:* With different packet loss ratio (PLR), it can be seen in Fig. 20 that smaller PLR results in better RD performance of *anchor frames* because of less transmission

erasure and smaller transmission distortion, in other words, less WZ bit-rate.

4) *The Average PSNR of the Decoded Sequences:* Fig. 21 shows the averaged PSNR of the decoded sequences for different error resilient schemes as a function of the excessive ratio which is defined as

$$r = \frac{R_{\text{total}} - R_{\text{orig}}}{R_{\text{orig}}} \quad (30)$$

where R_{total} is the total rate of the sequence and R_{orig} is the standardized H.264/AVC bit-rate. It can be seen that the proposed scheme outperforms other schemes, especially at the low excessive ratio region.

5) *PSNR of Each Frame:* Fig. 22 shows the PSNR of the 250 frames for different error resilient schemes. The results show that the proposed scheme which does not involve with a serious error accumulation of the residual distortion, could achieve a better performance.

6) *Subjective Quality Comparison:* From a perspective of subjective quality, the 100th frame reconstructed by different error resilient schemes are shown in Fig. 23. As compared to the Sehgal's scheme and IR scheme, the proposed scheme could eliminate most of the transmission distortion and maintain a clear visual quality improvement without annoying artifacts.

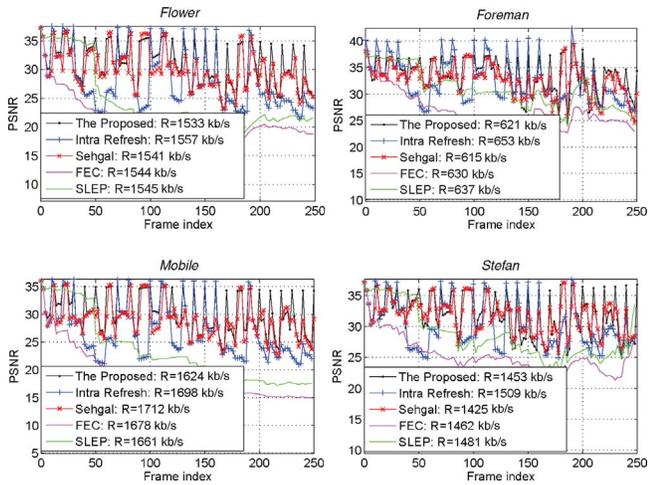


Fig. 22. PSNR comparison of different error resilient schemes.

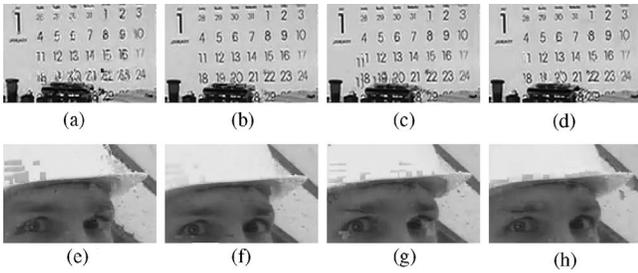


Fig. 23. Subjective quality comparison for the 100th decoded frame of *Mobile* and *Foreman* sequences. (a) EC only. (b) Proposed (1559 kb/s). (c) Sehgal's (1559 kb/s). (d) IR (1520 kb/s). (e) EC only. (f) Proposed (487 kb/s). (g) Sehgal's (488 kb/s). (h) IR (475 kb/s).

VII. CONCLUSION

This paper proposed a generic WZ-ERVC scheme which refers to two major contributions. At the encoder side, a joint source-channel R-D optimized mode selection (JSC-RDO-MS) algorithm with WZ coding is tuned to optimize the overall R-D performance for a total bit-rate of the primary description (H.264/AVC bit-stream) and the redundant description (embedded WZ bit-stream) of *anchor frames*. It is motivated by an analysis of the RD impact on the WZ bit-rate. The expected rate and distortion of WZ coding for JSC-RDO-MS depend on the side information that is only available at the decoder side. A symbol-level correlation noise model is adopted to estimate the minimal achievable rate, i.e., the conditional entropy $H(X|Y)$ for WZ video coding, and the minimum WZ coding bit-rate which guarantees a maximum decoding failure probability P_e is determined. In sequence, we statistically estimate the number of the parity bits from a combination of P_e and $H(X|Y)$ with a training set over a variety of video sequences. At the decoder side, we propose an online correlation noise model in both the pixel and the DCT domains to capture the spatially non-stationary characteristics of source correlation in WZ coding. The correlation noise between the source and side-information is modeled by a mixture of Laplacians whose parameters are obtained by analyzing the coherence of the motion field

of successive frames and the energy of prediction residual. Without using the source data, it leads to a more realistic WZ-ERVC solution when no feedback channel is available or a stringent end-to-end delay constraint exists.

APPENDIX

Let us consider a zero-mean stationary first-order Markov source $\{X_n\}_{n \in \mathbb{Z}}$ modeled by a linear predictor $X_n = \rho \hat{X}_{n-1} + W_n$, where $|\rho| < 1$ is the correlation coefficient between successive frames, W_n is the prediction residual which follows an i.i.d. zero mean Gaussian distribution, and is independent of X_n . In this model, $\rho \hat{X}_{n-1}$ is the best unbiased estimate of X_n given \hat{X}_{n-1} . In a predictive encoder, X_n is predicted from \hat{X}_{n-1} , instead of X_{n-1} to achieve a perfect synchronization between the encoder and the decoder. The interval between two successive *anchor frames* is T .

If a frame is correctly decoded, decoded picture \tilde{X}_n would be $\rho \tilde{X}_{n-1} + W_n$, otherwise, it would be the expectation of the current frame based on its previously decoded frames $E\{X_n | \tilde{X}_{n-1}, \tilde{X}_{n-2}, \dots\}$, i.e., $E\{X_n | \tilde{X}_{n-1}\}$ with the first-order Markov model. Using \tilde{X}_n as the side-information for WZ decoding, we have

$$Y_n = \tilde{X}_n = \begin{cases} \rho \tilde{X}_{n-1} + W_n, & \text{with prob. } 1 - p \\ E\{X_n | \tilde{X}_{n-1}\}, & \text{with prob. } p \end{cases} \quad (31)$$

where p is the packet loss ratio, and

$$\begin{aligned} E\{X_n | \tilde{X}_{n-1}\} &= E\{(\rho \tilde{X}_{n-1} + W_n) | \tilde{X}_{n-1}\} \\ &= E\{\rho \tilde{X}_{n-1} | \tilde{X}_{n-1}\} + E\{W_n\} \\ &= \tilde{\rho} \tilde{X}_{n-1} \end{aligned} \quad (32)$$

where $\tilde{\rho}$ is the estimated correlation coefficient between successive frames.

If a MB in *anchor frame* X_n is encoded in Inter mode, the expected distortion between \tilde{X}_n and \hat{X}_n is

$$\begin{aligned} D_c^P(n) &= E\{(\hat{X}_n - \tilde{X}_n)^2\} \\ &= pE\{[\rho \hat{X}_{n-1} + \hat{W}_n - (\tilde{\rho} \tilde{X}_{n-1})]^2\} + \\ &\quad (1-p)E\{[\rho \hat{X}_{n-1} + \hat{W}_n - (\rho \tilde{X}_{n-1} + \hat{W}_n)]^2\} \\ &= pE\{(\rho \hat{X}_{n-1} - \tilde{\rho} \tilde{X}_{n-1})^2\} + pE\{\hat{W}_n^2\} + \\ &\quad (1-p)E\{[\rho(\hat{X}_{n-1} - \tilde{X}_{n-1})]^2\} \\ &= pE\{[\rho \hat{X}_{n-1} - (\rho - \rho_e) \tilde{X}_{n-1}]^2\} + p\sigma_w^2 + \\ &\quad (1-p)\rho^2 D_c^P(n-1) \\ &= pE\{[\rho(\hat{X}_{n-1} - \tilde{X}_{n-1}) + \rho_e \tilde{X}_{n-1}]^2\} + p\sigma_w^2 + \\ &\quad (1-p)\rho^2 D_c^P(n-1) \\ &= p\rho^2 D_c^P(n-1) + pE\{(\rho_e \tilde{X}_{n-1})^2\} + p\sigma_w^2 + \\ &\quad (1-p)\rho^2 D_c^P(n-1) \\ &= \rho^2 D_c^P(n-1) + p\sigma_{\rho_e}^2 E\{\tilde{X}_{n-1}^2\} + p\sigma_w^2 \\ &= \dots \\ &= \rho^{2T} D_c^P(n-T) + p \frac{1-\rho^{2T}}{1-\rho^2} (\sigma_{\rho_e}^2 E\{\tilde{X}^2\} + \sigma_w^2). \end{aligned} \quad (33)$$

If it is encoded in Intra mode, the expected distortion between \tilde{X}_n and \hat{X}_n is

$$\begin{aligned}
D_c^I(n) &= E\{(\hat{X}_n - \tilde{X}_n)^2\} \\
&= pE\{[\rho\hat{X}_{n-1} + \hat{W}_n - (\tilde{\rho}\tilde{X}_{n-1})]^2\} \\
&= pE\{(\rho\hat{X}_{n-1} - \tilde{\rho}\tilde{X}_{n-1})^2\} + pE\{(\hat{W}_n)^2\} \\
&= p\rho^2E\{(\hat{X}_{n-1} - \tilde{X}_{n-1})^2\} + pE\{(\rho_e\tilde{X}_{n-1})^2\} + p\sigma_w^2 \\
&= p\rho^2D_c^P(n-1) + p\sigma_{\rho_e}^2E\{\tilde{X}_{n-1}^2\} + p\sigma_w^2 \\
&= \dots \\
&= p\rho^{2T}D_c^P(n-T) + (p + p^2\rho^2\frac{1-\rho^{2T-2}}{1-\rho^2}) \\
&\quad (\sigma_{\rho_e}^2E\{\tilde{X}^2\} + \sigma_w^2). \tag{34}
\end{aligned}$$

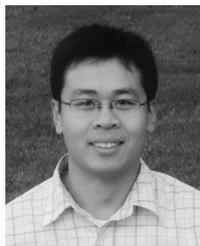
Supposing transmission errors occurring before *anchor frame* X_{n-T} could be corrected by the *anchor frame* X_{n-T} , so $D_c^P(n-T) = 0$. Equations (33) and (34) can be rewritten as follows:

$$D_c^P(n) = p\frac{1-\rho^{2T}}{1-\rho^2} (\sigma_{\rho_e}^2E\{\tilde{X}^2\} + \sigma_w^2) \tag{35}$$

$$D_c^I(n) = (p + p^2\rho^2\frac{1-\rho^{2T-2}}{1-\rho^2}) (\sigma_{\rho_e}^2E\{\tilde{X}^2\} + \sigma_w^2). \tag{36}$$

REFERENCES

- [1] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
- [2] S. H. G. Chan, X. Zheng, Q. Zhang, W. W. Zhu, and Y. Q. Zhang, "Video loss recovery with FEC and stream replication," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 370–381, Apr. 2006.
- [3] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 316–329, Apr. 2005.
- [4] H. Yang and K. Rose, "Rate-distortion optimized motion estimation for error resilient video coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 2, Mar. 2005, pp. 173–176.
- [5] R. M. Schreier and A. Rothermel, "Motion adaptive intra refresh for the H.264 video coding standard," *IEEE Trans. Consumer Electron.*, vol. 52, no. 1, pp. 249–253, Feb. 2006.
- [6] B. Katz, S. Greenberg, N. Yarkoni, N. Blaustien, and R. Giladi, "New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 308–319, Mar. 2007.
- [7] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
- [8] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod, "Systematic lossy forward error protection for video waveforms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, pp. 609–612.
- [9] S. Rane, P. Baccichet, and B. Girod, "Systematic lossy error protection of video signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1347–1360, Oct. 2008.
- [10] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner–Ziv coding of video: An error-resilient compression framework," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 249–258, Apr. 2004.
- [11] Y. Zhang, C. Zhu, and K.-H. Yap, "A joint source-channel video coding scheme based on distributed source coding," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1648–1656, Dec. 2008.
- [12] R. Bernardini, M. Fumagalli, M. Naccari, R. Rinaldo, M. Tagliasacchi, S. Tubaro, and P. Zontone, "Error concealment using a DVC approach for video streaming applications," in *Proc. EURASIP Eur. Signal Process. Conf.*, Sep. 2007, pp. 1–4.
- [13] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner–Ziv video coding performance," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 2, May 2006, pp. 525–528.
- [14] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. I. L. Lagendijk, "Analyzing symbol and bit plane-based LDPC in distributed video coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, pp. II-17–II-20.
- [15] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1177–1190, Sep. 2008.
- [16] Y. Zhang, H. Xiong, L. Song, and S. Yu, "Spatial non-stationary correlation noise modeling for Wyner–Ziv error resilience video coding," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2893–2896.
- [17] M. Morbee, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. 521–524.
- [18] M. Morbee, J. Prades-Nebot, A. Roca, A. Pizurica, and W. Philips, "Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel," in *Proc. Int. Conf. ACIVS*, vol. 4678, Aug. 2007, pp. 663–674.
- [19] X. Li, "On the importance of source classification in Wyner–Ziv video coding," in *Proc. SPIE Video Commun. Image Process.*, vol. 6822, Jan. 2008, pp. 68221Z.1–68221Z.11.
- [20] I.-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.
- [21] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 716–732, Jun. 2006.
- [22] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process.*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.
- [23] J. Chen, A. Khisti, D. M. Malioutov, and J. S. Yedidia, "Distributed source coding using serially concatenated-accumulate codes," in *Proc. IEEE Inform. Theory Workshop*, Oct. 2004, pp. 209–214.
- [24] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner–Ziv video coding with multiple side information," in *Proc. IEEE 9th Workshop Multimedia Signal Process.*, Oct. 2007, pp. 183–186.
- [25] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.
- [26] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
- [27] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
- [28] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks," *IEEE Trans. Multimedia*, vol. 9, no. 3, pp. 445–454, Apr. 2007.
- [29] Y.-K. Tu, J.-F. Yang, and M.-T. Sun, "Efficient rate-distortion estimation for H.264/AVC coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 600–611, May 2006.
- [30] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, Oct. 2001.
- [31] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conf.*, Apr. 2002, pp. 252–261.
- [32] D. Varodayan. (2006). *Rate-Adaptive LDPC Accumulate Codes for Distributed Source Coding* [Online]. Available: <http://www.stanford.edu/%7Edivad/ldpca.html>
- [33] I. E. G. Richardson, "H.264/MPEG4 part 10," in *H.264 and MPEG-4 Video Compression*. Chichester, U.K.: Wiley, Sep. 2003, pp. 170–177.
- [34] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–673, Jul. 2003.



Yongsheng Zhang received the B.S. and M.S. degrees in electronic engineering from Shandong University, Jinan, Shandong, China, in 2002 and 2005, respectively. He is currently working toward the Ph.D. degree in electrical engineering from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China.

His current research interests include distributed video coding and image processing.



Hongkai Xiong (M'01–SM'10) received the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003.

Since 2003, he has been with the Department of Electronic Engineering, Shanghai Jiao Tong University, where he is currently an Associate Professor. From December 2007 to December 2008, he was with the Department of Electrical and Computer Engineering, Carnegie Mellon University, PA, as a Research Scholar. He has published over 80 international journal/conference papers. In SJTU, he directs the Intelligent Video Modeling Laboratory and “multimedia communication” area in the Key Laboratory of Ministry of Education of China—“Intelligent Computing and Intelligent System” which is also co-granted by Microsoft Research, Beijing, China. His current research interests include source coding/network information theory, signal processing, computer vision and graphics, and statistical machine learning.

Dr. Xiong was the recipient of New Century Excellent Talents in University in 2009. In 2008, he received the Young Scholar Award of Shanghai Jiao Tong University. He has served on various IEEE conferences as a technical program committee member. He acts as a member of the Technical Committee on Signal Processing of the Shanghai Institute of Electronics.



Zhihai He (S'98–M'01–SM'06) received the B.S. degree in mathematics from Beijing Normal University, Beijing, China, in 1994, the M.S. degree in mathematics from the Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China, in 1997, and the Ph.D. degree in electrical engineering from the University of California, Santa Barbara, in 2001.

In 2001, he joined Sarnoff Corporation, Princeton, NJ, as a Technical Staff Member. In 2003, he joined the Department of Electrical and Computer Engineering, University of Missouri, Columbia, as an Assistant Professor. His current research interests include image/video processing and compression, network transmission, wireless communication, computer vision analysis, sensor networks, and embedded system design.

Dr. He received the 2002 IEEE Transactions on Circuits and Systems for Video Technology Best Paper Award, and the SPIE VCIP Young Investigator Award in 2004. Currently, he is an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *Journal of Visual Communication and Image Representation*. He is also a Guest Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Special Issue on Video Surveillance. He is a Member of the Visual Signal Processing and Communication Technical Committee of the IEEE Circuits and Systems Society, and is a technical program committee member or the session chair for a number of international conferences.



Songyu Yu graduated from Shanghai Jiao Tong University, Shanghai, China, in 1963.

He has been a Professor of electrical engineering with the Department of Electronic Engineering, Shanghai Jiao Tong University, since 1992. He has published more than 50 research papers and two monographs. His current research interests include image processing, video coding, video communications, and digital TV.

Mr. Yu has received a number of awards from the state, the Department of Education, and the

Department of Electrical Technology. He serves as the Secretary-General of the Shanghai Images and Graphics Association.



Chang Wen Chen (F'04) received the B.S. degree from the University of Science and Technology of China, Hefei, Anhui, China, in 1983, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Urbana, in 1992.

He has been a Professor of computer science and engineering with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, since 2008. Previously,

he was an Allen S. Henry Distinguished Professor with the Department of Electrical and Computer Engineering, Florida Institute of Technology, Melbourne, from 2003 to 2007. He was with the Faculty of Electrical and Computer Engineering, University of Missouri, Columbia, from 1996 to 2003, and with the University of Rochester, Rochester, NY, from 1992 to 1996. From 2000 to 2002, he was the Head of the Interactive Media Group, David Sarnoff Research Laboratories, Princeton, NJ. He has consulted with Kodak Research Laboratories, Rochester, NY, Microsoft Research, Microsoft Research, Beijing, China, Mitsubishi Electric Research Laboratories, Cambridge MA, NASA Goddard Space Flight Center, Greenbelt, MD, and the U.S. Air Force Rome Laboratories, Rome, NY.

Dr. Chen was the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from January 2006 to December 2009. He has served as an Editor for the PROCEEDINGS OF THE IEEE, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE MULTIMEDIA, the *Journal of Wireless Communication and Mobile Computing*, the *EUROSIP Journal of Signal Processing: Image Communications*, and the *Journal of Visual Communication and Image Representation*. He has chaired and served on numerous technical program committees for the IEEE and other international conferences. He was elected a fellow of the IEEE for his contributions to digital image and video processing, analysis, and communications, and a fellow of the SPIE for his contributions to electronic imaging and visual communications.