RATE-DISTORTION BASED MODE SELECTION FOR VIDEO CODING OVER WIRELESS NETWORKS WITH BURST LOSSES

Yiting Liao and Jerry D. Gibson

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA, 93106
Email: {yiting, gibson}@ece.ucsb.edu

ABSTRACT

Video communications over wireless networks suffer various patterns of losses, including both random packet loss and burst losses. Previous error resilient techniques simply consider the average packet loss rate to enhance error robustness for video transmission. However, loss patterns, specifically burst losses, have great impact on video quality. In this paper, we propose a method that can take account of both random and burst losses to further improve the error resilience of video coding. Our method estimates the end-to-end distortion based on recursive optimal per-pixel estimate (ROPE) including both random and burst losses, and applies it for rate-distortion (RD)-based optimal mode selection. We apply our method in two cases: For single description video coding, we estimate the reconstructed pixel values for random packet loss and burst losses, and calculate the overall distortion. For multiple description video coding, we estimate the end-to-end distortion for multiple state video coding (MSVC) by considering the network conditions and multiple state recovery to reduce the error propagation due to packet loss in both descriptions for MSVC. Simulation results show that our proposed method achieves better performance than MSVC and original ROPE (only considering average packet loss rate) over wireless networks with burst losses.

Index Terms— rate-distortion optimization, end-to-end distortion, H.264, video coding, error resilience

1. INTRODUCTION

In wireless networks, video transmission may suffer from packet loss due to link errors, node failures, route changes, interference and fading in the wireless channel, etc. The packet loss can seriously degrade the received video quality, especially due to the propagated errors in the motion-compensated prediction loop. Therefore, it is challenging to provide error resilient video coding for reliable video communications over such lossy networks. A number of techniques have been proposed to increase the robustness of video communications to packet loss, such as intra/inter mode selection [1–8], reference picture selection [9] [10], and multiple description video coding [11].

Intra coding is an important technique for mitigating error propagation due to packet loss and makes the video stream more robust to errors. However, using more intra-coded macroblocks (MBs) can greatly reduce the coding efficiency since intra-coded MB generally requires more bits than inter-coded MB. Therefore, to select the optimal intra/inter mode that can achieve the best tradeoff between error robustness and coding efficiency has become a widely addressed problem. There are some simple intra updating methods such as refreshing contiguous intra blocks periodically [1], or intra-coding blocks randomly [2].

A more advanced category of intra refresh algorithms estimates the end-to-end distortion due to both compression and packet loss, and incorporates mode selection with rate-distortion (RD) optimization [2–8]. An early work of RD-based mode selection method is proposed in [3], in which the distortion is roughly estimated. In [2], the encoder considers the effects of error concealment and intra-codes the area that is severely affected by packet loss. However, the error propagation beyond one frame is ignored during the estimation procedure. In [4], the authors further incorporate the distortion due to error concealment of a current block with the distortion due to error propagation from concealed blocks to optimize mode selection. One drawback of the methods proposed in [2–4] is that the estimated distortion at the encoder is not very accurate. A more precise approach to estimate the end-to-end distortion is proposed in [5]. The authors generate K copies of the channel behavior at the encoder and calculate the decoder reconstruction to estimate the expected end-to-end distortion. This approach can very accurately estimate the distortion if K is large enough. However, it has extremely high computational complexity. In [6], an algorithm called "Recursive Optimal Per-pixel Estimate" (ROPE) is proposed to compute the distortion by recursively calculating the first and second moments of each pixel due to compression, error

This research has been supported by the California Micro Program, Applied Signal Technology, Cisco, Sony-Ericsson and Qualcomm, Inc, and by NSF Grant Nos. CCF-0429884, CNS-0435527, and CCF-0728646.

concealment, and error propagation. This algorithm provides an accurate estimation of end-to-end distortion at the cost of a modest increase in computational complexity. Since the ROPE algorithm achieves substantial gains over competing methods, numerous extensive work have been proposed based on the ROPE algorithm. For example, [7] estimates the variance of expected distortion by calculating the first four moments of each pixel and incorporates it to allocate channel resources. In [8], the overall distortion is divided into several separable distortion items to reduce the computing complexity. In [12], the authors estimate the expected end-to-end distortion to select between multiple description modes on a frame basis.

All these techniques only consider a simple network condition in which a average packet loss rate is assumed. However, [13] has shown that not only average packet loss rate but also the specific pattern of the loss affects the expected distortion; specifically, it proves that burst loss has a great impact on the distortion. Because of the likelihood of both random packet loss and burst losses in video communications over wireless networks, we propose a method which take account this more complicated network condition for optimal mode selection to enhance the error resilience of video.

The method estimates the end-to-end distortion based on the ROPE algorithm including the random and burst losses, and uses RD optimization for optimal mode selection. The method is applied in two cases. For single description video coding, we estimate the reconstructed pixel value due to random loss and burst losses, which results in a more precise estimation for end-to-end distortion. When applying to RD-based mode selection, this method helps to achieve the optimal tradeoff between error resilience and coding efficiency under different random and burst loss rates, and outperforms ROPE algorithm over lossy networks. Part of this work has been presented in [14]. In this paper, we extend the study in [14] by considering both bursty and random losses and performing more comprehensive simulations. For multiple description video coding, we estimate the expected end-to-end distortion of multiple state video coding (MSVC) [15] for optimal mode selection. MSVC transmits two descriptions over two different paths and it is effective to combat burst losses since the loss of consecutive frames in one description can be well concealed by the frames in the other description. However, it still suffers error propagation in both descriptions when random packet loss happens. Therefore, we estimate the reconstructed pixel value by considering the network condition, error propagation and multiple state recovery, and select the mode that enhances the error robustness of MSVC. The simulation results show that our proposed method can better combat random and burst losses over the network than original ROPE and MSVC.

The rest of the paper is organized as follows: Section 2 introduces some background information, including the packet loss model, multiple state video coding (MSVC), and the RD-based mode selection method. We present our proposed method for two cases in Section 3. Section 4 introduces the performance metrics to evaluate the video quality. The proposed method is compared with ROPE and MSVC under different loss patterns, and the simulation results are discussed in Section 5. Section 6 analyzes the computation complexity and the robustness to mismatch of our proposed method. Finally, conclusions are drawn in Section 7.

2. BACKGROUND

In this section, we first introduce a packet loss model that nicely characterizes both the random packet loss and burst losses over the wireless network. We also introduce the multiple state video coding (MSVC) method proposed in [15] and the RD-based mode selection method in H.264.

2.1. Packet Loss Model

In wireless networks, packet loss may occur due to numerous reasons, including link/node failures, route changes, and bit errors. These factors can cause both random packet loss and burst losses over the network. To investigate the video communications over such lossy networks, we first introduce a simple packet loss model that captures packet loss features in the network. As shown in Fig.1, this model considers both the random packet loss and burst losses during the transmission and can be used to generate different loss patterns over the wireless network.

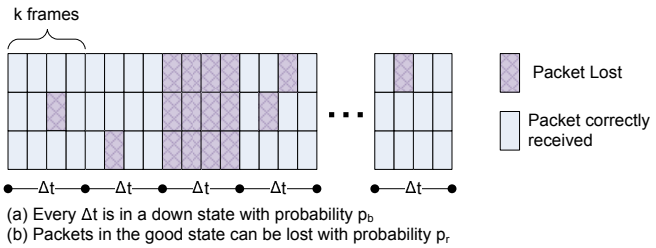


Fig. 1. Packet Loss Model

In this model, time is divided into Δt intervals and k frames are transmitted during an interval. Each interval may be either in a good state with probability $(1-p_b)$ or in a down state with probability p_b, which is independent and identically distributed. The packets transmitted in a down state are all lost while the packets transmitted in the good state may suffer from a random packet loss. Therefore, the packet loss model can be determined by three parameters: the burst loss rate p_b, the burst length k (frames), and the random packet loss rate p_r in a good state. And the total packet loss rate p in the networks can be calculated by,

$$p = p_b + (1-p_b)p_r = p_b + p_r - p_b p_r \qquad (1)$$

2.2. Multiple State Video Coding (MSVC)

MSVC proposed in [15] is an effective approach to enhance error resilience for video transmission. In MSVC, the system includes a multiple state video encoder/decoder and a path diversity transmission system as shown in Fig. 2.

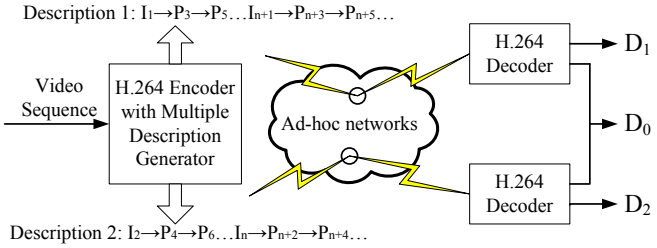


Fig. 2. MSVC system architecture

At the encoder, the video sequence is first temporally down-sampled into two sub-sequences, i.e. odd frames in the original sequence are extracted as one description and even frames as the other. The two descriptions are encoded separately using a H.264 video encoder [16] and transmitted over the networks in different paths. At the decoder, they are decoded and interleaved to get the reconstructed video sequence.

When one description experiences packet loss, the information in the other description can be used to improve the recovery of the corrupted video segment. This is referred as multiple state recovery [15]. In [17], the performance of MSVC is further improved by applying refined error concealment methods on a MB basis. For MSVC, the even and odd frames are transmitted in different paths, burst losses in one description can be well concealed by frames in the other description and cause less damage to the reconstructed video than single description coding in which burst losses may result in the loss of consecutive frames. However, random packet loss may cost error propagation in both descriptions and we try to alleviate the error propagation by applying optimal mode selection for MSVC.

2.3. RD-based Mode Selection

Video standards such as H.264 provide different intra and inter modes to encode a MB. In order to decide the best mode for each MB, Lagrangian optimization technique is used to minimize the distortion subject to a rate constraint [18]. In other words, the coding mode that minimizes the Lagrangian cost in the following equation is chosen to code the MB,

$$\min_{mode}(J_{\mathrm{MB}}) = \min_{mode}(D_{MB} + \lambda_{\mathrm{mode}} R_{\mathrm{MB}}) \tag{2}$$

where λ_{mode} is the Lagrangian multiplier for the mode decision given by Eq. (3) in H.264. R_{MB} denotes the bits needed for coding the MB in the specific mode, which includes the bits for the MB header, the motion vector, the reference frame, and the transformed coefficients. D_{MB} represents the distortion of the MB.

$$\lambda_{\mathrm{mode}} = 0.85 \times 2^{(QP-12)/3} \tag{3}$$

In the next section, we propose a method to estimate the end-to-end distortion of each pixel at the encoder by considering both random and burst losses. The method is applied for two cases. For single description coding, we estimate the concealed pixel value under random packet loss and burst losses to calculate the overall distortion more accurately. For multiple description video coding with path diversity, we estimate the end-to-end distortion for MSVC to improve its robustness to packet loss.

3. PROPOSED METHOD

3.1. Preliminaries

Table 1 defines the notations used in the derivation of the distortion. The distortion of each MB is the sum of the distortion

Table 1. Notations

	Definitions
d_n^i	Distortion of pixel i in frame n
f_n^i	Original value of pixel i in frame n
$\hat{f}_n^i$	Encoder-reconstructed value of pixel i in frame n
$\tilde{f}_n^i$	Decoder-reconstructed value of pixel i in frame n (after error concealment)
$\hat{r}_n^i$	Quantized residue of pixel i in frame n (Inter mode)

of the pixels in the MB,

$$D_{MB} = \sum_{i \in MB} d_n^i \tag{4}$$

The expected end-to-end distortion for the pixel f_n^i is given by

$$\begin{aligned} d_n^i &= E[(f_n^i - \tilde{f}_n^i)^2] \\ &= (f_n^i)^2 - 2f_n^i E[\tilde{f}_n^i] + E[(\tilde{f}_n^i)^2] \end{aligned} \tag{5}$$

Notice that the value of $\tilde{f}_n^i$ is a random variable at the encoder. In order to estimate the expected distortion d_n^i at the encoder, we need to calculate the first and second moments of $\tilde{f}_n^i$ for an intra or inter MB separately.

3.2. Extended ROPE with Burst Losses

In [6], the authors develop the ROPE algorithm to recursively compute the first and second moments of $\tilde{f}_n^i$ based on the packet loss rate p and error concealment method. We notice that the ROPE algorithm only considers a simple loss model,

in which each packet may be lost with a packet loss rate p. While in wireless networks, the loss pattern is usually more complicated. The video packets may suffer from burst losses as well as random loss. Reference [13] shows that the loss pattern has a significant impact on the distortion and burst losses generally cause a larger distortion than isolated losses. Therefore, we extend the ROPE algorithm with burst losses to better estimate the decoder-reconstructed pixel value for single description video coding.

When burst losses happen, the concealed pixel is further away from the last correctly received frame and it generally has a greater distortion. Therefore, we distinguish it from the concealed pixel value due to random loss. By separately estimating the concealed pixel value due to random loss and burst losses, we can more accurately calculate the end-to-end distortion at the encoder for optimal mode decision.

We assume that the temporal-copy error concealment is used to recover the lost video segment. That is, a lost MB is concealed by copying the previous correctly received MB in the corresponding position. The packet loss model in Section 2.1 is applied, in which three parameters need to be considered for the extended ROPE algorithm: burst loss rate p_b, burst length k (frames), and random packet loss rate p_r. Using the notations in Table 1, we calculate the first and second moments of $\tilde{f}_n^i$ in intra and inter modes respectively.

3.2.1. Pixel in an intra-coded MB

According to the packet loss model, each packet may experience three network conditions:

1. The packet is correctly received with probability $(1-p_b)(1-p_r)$. We thus have $\tilde{f}_n^i = \hat{f}_n^i$.
2. The packet suffers burst losses with probability p_b. This means that k consecutive frames are lost during the time interval Δt. The lost MB is then concealed by the co-located MB in the last correctly received frame. That is $\tilde{f}_n^i = \tilde{f}_{n-(n \bmod k)}^i$.
3. The packet encounters random loss with probability $(1-p_b)p_r$. Then the lost MB is recovered by copying the co-located MB in the previous frame. Therefore, we have $\tilde{f}_n^i = \tilde{f}_{n-1}^i$.

Based on the three cases, the first and second moments of $\tilde{f}_n^i$ in an intra-coded MB are calculated by,

$$\begin{aligned} E[\tilde{f}_n^i] &= (1-p_r)(1-p_b)(\hat{f}_n^i) + (1-p_b)p_r E[\tilde{f}_{n-1}^i] \\ &\quad + p_b E[\tilde{f}_{n-(n \bmod k)}^i] \end{aligned} \tag{6}$$

$$\begin{aligned} E[(\tilde{f}_n^i)^2] &= (1-p_r)(1-p_b)(\hat{f}_n^i)^2 + (1-p_b)p_r E[(\tilde{f}_{n-1}^i)^2] \\ &\quad + p_b E[(\tilde{f}_{n-(n \bmod k)}^i)^2] \end{aligned} \tag{7}$$

3.2.2. Pixel in an inter-coded MB

When the pixel is inter-coded, there are also three cases to estimate the decoder-reconstructed pixel value:

1. The packet is correctly received with probability $(1-p_b)(1-p_r)$. For an inter-coded pixel, we assume that pixel i is predicted from pixel j in the previous frame and the quantized residue is $\hat{r}_n^i$. Then the encoder reconstruction $\hat{f}_n^i$ is computed by adding the quantized residue to the prediction, that is, $\hat{f}_n^i = \hat{r}_n^i + \hat{f}_{n-1}^j$. Thus, the decoder-reconstructed pixel value is given by, $\tilde{f}_n^i = \hat{r}_n^i + \tilde{f}_{n-1}^j$.
2. The packet suffers burst losses with probability p_b. Similar to the intra-coded pixel, the pixel is concealed from the last correctly received frame and we have $\tilde{f}_n^i = \tilde{f}_{n-(n \bmod k)}^i$.
3. The packet encounters random loss with probability $(1-p_b)p_r$ and the pixel is concealed by the pixel in the previous frame: $\tilde{f}_n^i = \tilde{f_{n-1}^i}$.

Finally, the first and second moments of $\tilde{f}_n^i$ in an inter-coded MB are given by:

$$\begin{aligned} E[\tilde{f}_n^i] &= (1-p_r)(1-p_b)(\hat{r}_n^i + E[(\tilde{f}_{n-1}^j)]) \\ &\quad + (1-p_b)p_r E[\tilde{f}_{n-1}^i] + p_b E[\tilde{f}_{n-(n \bmod k)}^i] \end{aligned} \tag{8}$$

$$\begin{aligned} E[(\tilde{f}_n^i)^2] &= (1-p_r)(1-p_b)E[(\hat{r}_n^i + \tilde{f}_{n-1}^j)^2] + (1-p_b)p_r \\ &\quad E[(\tilde{f}_{n-1}^i)^2] + p_b E[(\tilde{f}_{n-(n \bmod k)}^i)^2] \end{aligned} \tag{9}$$

Using Eqns. (6)-(9), we can recursively estimate the first and second moments of $\tilde{f}_n^i$ and calculate the overall end-to-end distortion for each MB. By applying the RD-based mode selection method in Section 2.3, the optimal mode that provides a good trade-off between coding efficiency and error resilience for the specific random and burst loss rates is chosen.

3.3. Optimal Mode Selection for MSVC

In Section 2.2, we know that MSVC transmits two independently decodable descriptions over two different paths to reduce the loss of consecutive frames. Burst losses in one description only cause the loss of consecutive odd (even) frames, which can be well concealed by the even (odd) frames in the other description. This property makes MSVC robust to burst losses. However, the distortion due to random loss still propagates to future frames and multiple state recovery may cause the error to propagate in both descriptions. Therefore, MSVC is quite vulnerable to random loss. In order to enhance the error resilience of MSVC under both random loss and burst losses, we propose the optimal mode selection for MSVC.

The idea is similar to the ROPE method, except that MSVC uses multiple state recovery to conceal the error and it needs to be considered during the estimation process. We assume that the refined error concealment methods on a MB basis are applied [17]. We estimate the first and second moments of $\tilde{f}_n^i$ by considering the packet loss rate p, and the multiple state recovery and calculate the expected end-to-end distortion for each MB. When applying RD-based mode selection, the proposed method can better recover from random loss.

3.3.1. Pixel in an intra-coded MB

To compute the first and second moments of $\tilde{f}_n^i$ for an Intra MB, we need to consider the following scenarios:

1. The packet for f_n^i is correctly received with probability $1-p$ and thus we have $\tilde{f}_n^i = \hat{f}_n^i$.

2. The packet for f_n^i is lost and the neighbor group of blocks (GOB) is received with probability $p(1-p)$. In this case, we estimate the motion vector of lost pixel from one of the available neighbor MBs and use motion-compensated concealment to recover the lost pixel. We choose one frame as the reference from each description and get two reconstructed values $\tilde{f}_{n-1}^{j_1}$ and $\tilde{f}_{n-2}^{j_2}$. Then pixel $\tilde{f}_n^i$ is recovered from $\tilde{f}_{n-1}^{j_1}$ or $\tilde{f}_{n-2}^{j_2}$ depending on which reconstructed value is closer to $\hat{f}_n^i$, i.e. $\tilde{f}_n^i = \tilde{f}_{n-m}^{j_m}$, where $m = \arg\min_{x\in\{1,2\}}(\tilde{f}_{n-x}^{j_x} - \hat{f}_n^i)^2$.

3. The packet for f_n^i and the neighbor GOB are both lost with probability p^2. Then either $\tilde{f}_{n-1}^i$ or $\tilde{f}_{n-2}^i$ is used to conceal $\tilde{f}_n^i$. Thus, $\tilde{f}_n^i = \tilde{f}_{n-k}^i$, where $k = \arg\min_{x\in\{1,2\}}(\tilde{f}_{n-x}^i - \hat{f}_n^i)^2$.

Based on the above cases, we can calculate the first and second moments of $\tilde{f}_n^i$ in an intra MB by Eqns. (10) and (11).

$$\begin{aligned} E[\tilde{f}_n^i] &= (1-p)(\hat{f}_n^i) + p(1-p)E[\tilde{f}_{n-m}^{j_m}] \\ &\quad + p^2 E[\tilde{f}_{n-k}^i] \qquad (10) \\ E[(\tilde{f}_n^i)^2] &= (1-p)(\hat{f}_n^i)^2 + p(1-p)E[(\tilde{f}_{n-m}^{j_m})^2] \\ &\quad + p^2 E[(\tilde{f}_{n-k}^i)^2] \qquad (11) \\ \text{where} \quad & m = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^{j_x}] - \hat{f}_n^i)^2, \\ & k = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^i] - \hat{f}_n^i)^2 \end{aligned}$$

3.3.2. Pixel in an inter-coded MB

For MSVC, the odd frame is predicted from previous odd frames and the even frame is predicted from previous even frames. Therefore, the quantized residue $\hat{r}_n^i = \hat{f}_n^i - \hat{f}_{n-2}^j$ for MSVC, where pixel i in frame n is predicted from pixel j in frame $n-2$. Assume that $j_m (m = 1, 2)$ is the pixel corresponding to the estimated concealment motion vector for pixel i in frame $n-m$. Then we can calculate the first and second moments of $\tilde{f}_n^i$ according to the three cases similar to those in Section 3.3.1,

$$\begin{aligned} E[\tilde{f}_n^i] &= (1-p)(\hat{r}_n^i + E[\tilde{f}_{n-2}^j]) \\ &\quad + p(1-p)E[\tilde{f}_{n-m}^{j_m}] + p^2 E[\tilde{f}_{n-k}^i] \qquad (12) \\ E[(\tilde{f}_n^i)^2] &= (1-p)E[(\hat{r}_n^i + \tilde{f}_{n-2}^j)^2] \\ &\quad + p(1-p)E[(\tilde{f}_{n-m}^{j_m})^2] + p^2 E[(\tilde{f}_{n-k}^i)^2] \qquad (13) \\ \text{where} \quad & m = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^{j_x}] - \hat{f}_n^i)^2, \\ & k = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^i] - \hat{f}_n^i)^2 \end{aligned}$$

4. PERFORMANCE METRICS

In order to analyze the performance of the decoded video sequences, we use the average PSNR of all frames over all realizations to evaluate the objective video quality. However, due to non-linear behavior of human visual system, video sequences with close average PSNR may reveal different perceptual video quality for human viewers. Therefore, we also introduce $PSNR_{r,f}$ proposed in [19] to evaluate the perceptual video quality.

$PSNR_{r,f}$ is defined as the PSNR achieved by $f\%$ of frames for the $r\%$ of realizations, which shows the video quality guaranteed for $r\%$ of realizations among $f\%$ frames. The definition of $PSNR_{r,f}$ can be written as

$$PSNR_{r,f} = \arg_x P_{real}(P_{frame}(PSNR > x) \geq f) \geq r) \qquad (14)$$

Here, $P_{frame}(PSNR > x)$ is the percentage of frames that have PSNR higher than x in one realization and $P_{real}(\Omega)$ is the percentage of realizations that satisfy the condition Ω. For example, $PSNR_{r=80\%,f=90\%} = 35dB$ means that there are 80% of the realizations having 90% of frames with PSNR higher than 35 dB. We use $PSNR_{r,f}$ to evaluate the perceptual video quality because of two findings [19, 20]: (1) The bad-quality frames dominate users' experience with the video; (2) For PSNRs higher than a certain threshold, increasing PSNR does not help to enhance the perceptual video quality. We know that average PSNR treats every frame equally and does not perfectly correlate with the perceptual video quality because of the non-linear behavior of human vision system. While $PSNR_{r,f}$ can capture the performance

loss due to damaged frames in a video sequence ($f\%$). Furthermore, $PSNR_{r,f}$ captures the performance experienced by a user for multiple uses ($r\%$) of the channel, or alternatively, it can be interpreted as a performance indicator for multiple users ($r\%$) of the channel.

5. PERFORMANCE EVALUATION

5.1. Simulation Settings

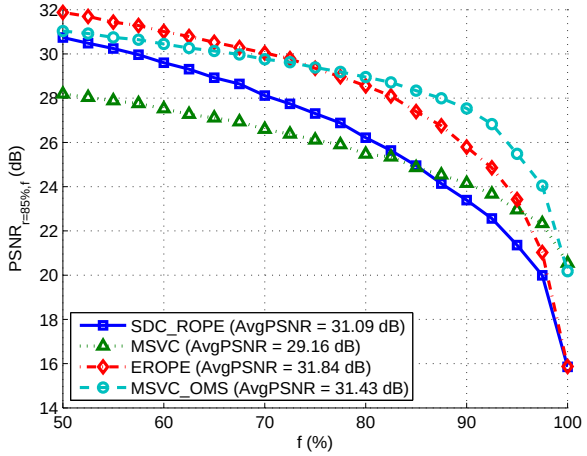
We implement our proposed method by modifying H.264 reference software JM13.2. Currently, we use the temporal copy method in the implementation. That is, the lost MB is concealed by copying the co-located MB in the last correctly received frame. For MSVC, to conceal a missing MB in frame n, we examine the corresponding MB from the nearest frame in each description (frame $n-1$ and frame $n-2$). The pixels that minimizes the side match distortion are used for concealment. The following four approaches are implemented for comparison:

- **SDC_ROPE**: Single description coding with ROPE proposed in reference [6], in which the total packet loss rate is applied to estimate the distortion
- **MSVC**: Multiple state video coding introduced in Section 2.2 with refined error concealment methods proposed in [17]
- **EROPE**: Single description coding with extended ROPE that accounts for both random and burst losses proposed in Section 3.2
- **MSVC_OMS**: The optimal mode selection approach for MSVC proposed in Section 3.3, in which the packet loss rate and multiple state recovery are considered to estimate the distortion

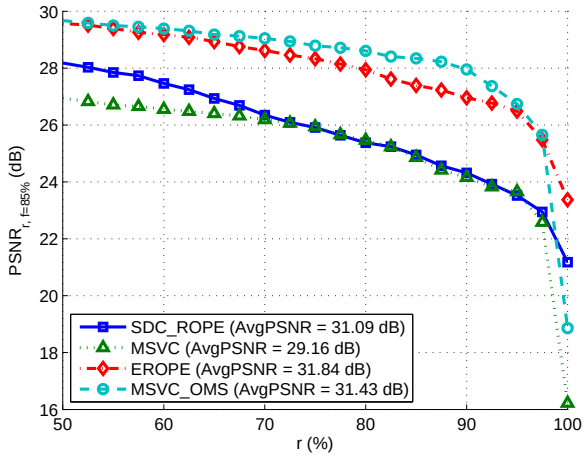
Video sequences of 300 frames with QCIF format are used in the simulation. The sequences are encoded at 30 fps and packetized to RTP format. The packet loss model in Section 2.1 is used and we simulate each video sequence over 500 different realizations under the same network settings.

5.2. Performance Comparison

We first compare the performance of four approaches under certain network condition ($p_r = 3\%$, $p_b = 3\%$, $k = 5$). The average PSNRs for SDC_ROPE, MSVC, EROPE, and MSVC_OMS are 31.09 dB, 29.16 dB, 31.84 dB, and 31.43 dB respectively (Also shown in Fig. 3). We see that MSVC has the worst average PSNR, which is 1.9 dB-2.7 dB lower than other methods. And EROPE achieves a PSNR about 0.8 dB higher than SDC_ROPE and 0.4 dB higher than MSVC_OMS.



(a) $PSNR_{r=85\%,f}$



(b) $PSNR_{r,f=85\%}$

Fig. 3. Comparing $PSNR_{r,f}$ of SDC_ROPE, MSVC, EROPE, and MSVC_OMS, Foreman sequence at bitrate 300 kbps, $p_r = 3\%, p_b = 3\%, burstlength = 5$

Based on the average PSNR, we see that EROPE and MSVC_OMS achieve better objective video quality. However, we see that average PSNRs of SDC_ROPE, EROPE, and MSVC_OMS are all above 31 dB, which indicate good video quality. In this case, the bad-quality frames may dominate viewer's experience and they are critical to decide the perceptual video quality. Therefore, we further examine the $PSNR_{r,f}$ of the four approaches to analyze their perceptual performance.

Figure 3(a) shows $PSNR_{r,f}$ of SDC_ROPE, MSVC, EROPE, and MSVC_OMS for Foreman sequence with fixed $r = 85\%$. In Fig. 3(a), we see that SDC_ROPE and EROPE have more number of low-quality frames than MSVC_OMS in 85% of the realizations. For example, as shown in Fig. 3(a), about 15% of frames in 85% of the realizations for

Table 2. Comparing performance of four approaches for Foreman sequence, $p_r = 3\%, p_b = 3\%, k = 5$

Objective video quality	EROPE > MSVC_OMS > SDC_ROPE > MSVC
Number of bad-quality frames	SDC_ROPE > MSVC > EROPE > MSVC_OMS
Quality guaranteed for multiple users	MSVC_OMS > EROPE > SDC_ROPE ≈ MSVC

Table 3. Average PSNR and $PSNR_{r=85\%,f=85\%}$ for different video sequences at bitrate 100 kbps, $p_r = 3\%, p_b = 3\%, k = 5$

Video sequence	Average PSNR (dB)				$PSNR_{r=85\%,f=85\%}$			
	SDC_ROPE	MSVC	EROPE	MSVC_OMS	SDC_ROPE	MSVC	EROPE	MSVC_OMS
Akiyo	37.77	36.00	38.02	37.26	33.12	33.12	33.66	34.25
Claire	37.11	35.89	37.57	37.27	30.91	33.05	32.13	35.08
News	31.09	29.97	31.33	30.71	26.43	26.65	27.33	27.72
Mother&Daughter	35.18	33.80	35.48	34.60	31.23	30.92	32.18	32.25
Salesman	33.58	32.67	33.64	33.15	30.29	30.63	30.97	31.49

SDC_ROPE have a PSNR lower than 25 dB, while fewer than 5% of frames in 85% of the realizations for MSVC_OMS achieve a PSNR lower than 25 dB. Figure 3(a) demonstrates that MSVC_OMS achieve the best perceptual video quality among the four approaches because it has fewest number of bad-quality frames that dominate viewers' experience.

Figure 3(b) plots $PSNR_{r,f}$ of SDC_ROPE, MSVC, EROPE and MSVC_OMS with fixed f=85%. In the figure, we see that EROPE and MSVC_OMS achieve higher $PSNR_{r,f=85\%}$ than SDC_ROPE and MSVC under most values of r. This means that EROPE and MSVC_OMS can guarantee a higher PSNR than SDC_ROPE and MSVC for 85% frames in almost all of the realizations. For example, the PSNRs guaranteed for 85% of the frames in 85% of the realizations for SDC_ROPE, MSVC, EROPE, and MSVC_OMS are 24.94 dB, 24.86 dB, 27.39 dB, and 28.34 dB respectively. It indicates that EROPE and MSVC_OMS provide better video quality than SDC_ROPE and MSVC for most of the users over the network.

Table 2 summarizes the performance of four approaches for Foreman sequence. Since EROPE better estimates the end-to-end distortion by considering random packet loss and burst losses, and provides a good trade-off between coding efficiency and error resilience, it achieves the best objective video quality among the four approaches. And MSVC_OMS combines the benefits of multiple description coding and optimal mode selection. Although the decreased correlation between the adjacent frames in each description reduces the coding efficiency of MSVC_OMS, the usage of multiple descriptions and path diversity enhances its robustness to burst losses while optimal mode selection helps MSVC_OMS to better combat random loss. Therefore, MSVC_OMS guarantees the smallest number of bad-quality frames for most of the users, which indicates the best perceptual video quality among the four approaches.

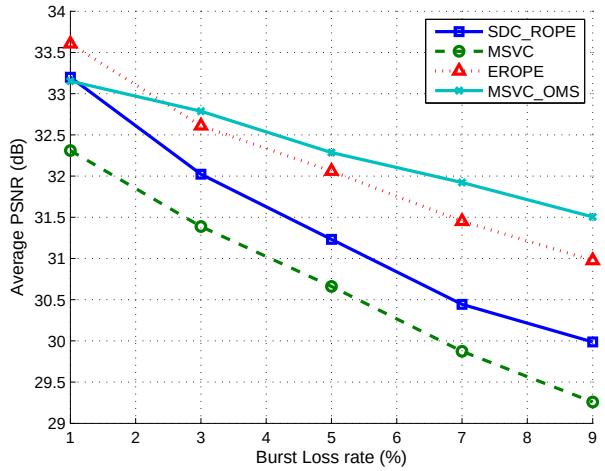
Table 3 shows the average PSNR and $PSNR_{r=85\%,f=85\%}$ for different video sequences at bitrate 100 kbps under packet loss parameters $p_r = 3\%, p_b = 3\%, k = 5$. When compared the two single description coding approaches, we see that our proposed EROPE approach outperforms SDC_ROPE in both average PSNR and $PSNR_{r=85\%,f=85\%}$. For the two multiple description coding approaches, we also see that our MSVC_OMS achieves higher average PSNR and $PSNR_{r=85\%,f=85\%}$ than MSVC. These results show that our two proposed approaches enhance the error resilience of video and provide better objective and subjective video quality than the original approaches, respectively.

When compared between our EROPE approach and our MSVC_OMS approach, we see that EROPE has higher average PSNR in the range of 0.3 - 0.8 dB and MSVC_OMS achieves the higher $PSNR_{r=85\%,f=85\%}$ from 0.1 dB and 2.9 dB. The results indicate that the two proposed approaches both provide good error resilience to packet loss. The EROPE approach is applied for single description video coding and suitable for the single-path network. While MSVC_OMS is used for the multiple description video coding with path diversity.

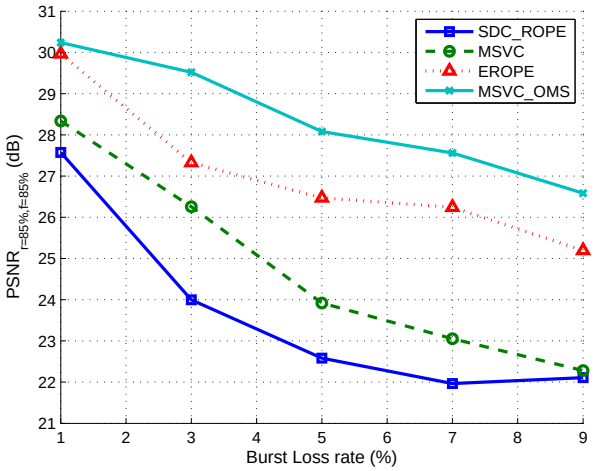
5.3. Impact of Burst Loss Rate

In this section, we investigate the impact of burst loss rate on different approaches when the random loss rate and burst length are fixed. Figure 4 shows the performance of SDC_ROPE, MSVC, EROPE, and MSVC_OMS with fixed random loss rate 1% and burstlength 5 under different burst loss rates.

In Fig. 4(a), we see that the average PSNRs of SDC_ROPE

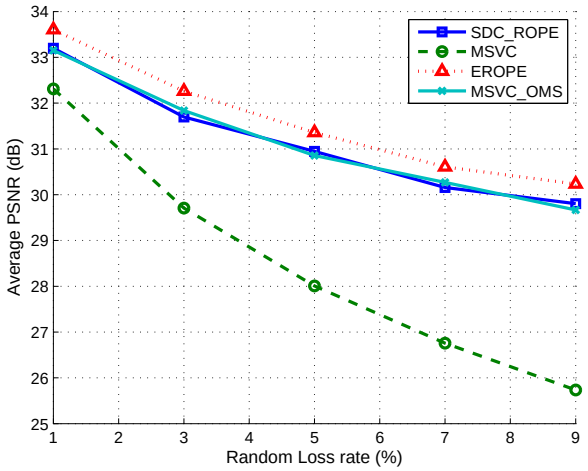


(a) Average PSNR vs. burst loss rate

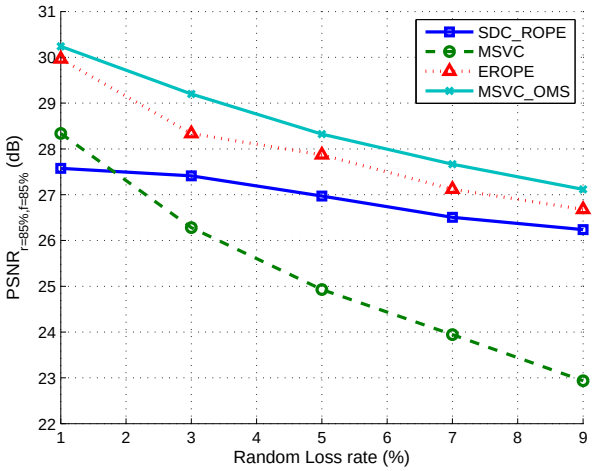


(b) $PSNR_{r=85\%,f=85\%}$ vs. burst loss rate

Fig. 4. Comparing performance of SDC_ROPE, MSVC, EROPE, and MSVC_OMS under different burst loss rates, Foreman sequence at bitrate 300 kbps, $p_r = 1\%, burstlength = 5$



(a) Average PSNR vs. random loss rate



(b) $PSNR_{r=85\%,f=85\%}$ vs. random loss rate

Fig. 5. Comparing SDC_ROPE, MSVC, EROPE, and MSVC_OMS under different random loss rates, Foreman sequence at bitrate 300 kbps, $p_b = 1\%, burstlength = 5$

and MSVC constantly drop as the burst loss rate increases while EROPE and MSVC_OMS are more efficient to combat burst losses. MSVC_OMS works best when the burst loss rate is high. In Fig. 4(b), we see that MSVC_OMS achieves a $PSNR_{r=85\%,f=85\%}$ by up to 5.5 dB higher than ROPE, 4.5 dB higher than MSVC, and 2.2 dB higher than EROPE, which shows that MSVC_OMS provides the best perceptual video quality among the four approaches for high burst loss rate. According to Fig. 4, we see that our proposed approaches achieve higher gains over previous approaches when the burst loss rate is high. And MSVC_OMS is the most suitable approach for high burst loss rate case.

5.4. Impact of Random Loss Rate

Figure 5 shows the performance of SDC_ROPE, MSVC, EROPE, and MSVC_OMS under different random loss rates with fixed burst loss rate 1% and burstlength 5. In Fig. 5(a), we see that SDC_ROPE, EROPE, and MSVC_OMS achieve close average PSNR for different random loss rates and the average PSNR of MSVC drops badly as random loss rate increases. This is because SDC_ROPE, EROPE, and

MSVC_OMS all include the distortion caused by random loss in mode selection, which makes them robust to random loss. While the error propagation due to random loss in both descriptions of MSVC can greatly degrade its video quality. In Fig. 5, we see that the performance gains achieved by EROPE and MSVC_OMS compared to ROPE do not vary much under different random loss rates. These results are as expected since the proposed methods are mainly designed to better combat burst losses while maintain similar performance for random loss.

6. DISCUSSION

6.1. Complexity Considerations

In Section 5, we show that our proposed approaches achieve better performance than SDC_ROPE and MSVC. In this section, we compare the cost of computational complexity and storage of our proposed approaches with ROPE. We know that ROPE costs a modest increase in computational complexity, which mostly introduces by calculating the two moments of $\tilde{f}_n^i$ for the cases of intra mode and inter mode, for each pixel. When compared EROPE to ROPE, we know that the concealed pixel value caused by burst losses is estimated separately and it introduces 4 more addition/multiplication operations for each pixel in an intra-coded/inter-coded MB. Since the error concealment is the same regardless of the coding mode of the MB, the total number of extra addition/multiplication operations is 6 for each pixel. For MSVC_OMS, the extra operations come from the selection of multiple state recovery and it is the same for both intra-coded MB and inter-coded MB. Therefore, MSVC_OMS requires 8 more extra addition/multiplication operations for each pixel than ROPE. Based on the above analysis, we see that the computational complexity of EROPE and MSVC_OMS is in the same order of ROPE. Furthermore, all the additional complexity occurs only at the encoder.

For storage complexity, we see that ROPE only needs to store the two moments of each pixel in the previous frame, while EROPE stores the two moments of previous k frames and MSVC_OMS stores the two moments of the previous two frames. This extra storage cost introduced by EROPE and MSVC_OMS is negligible in most applications.

6.2. Mismatch of Network Conditions

Our approaches assume that the network conditions are known at the encoder and are used as the coding parameters. We need to analyze the situation that mismatch happens between the assumed network condition and actual condition in the network. There are two cases to be considered: either the assumed packet loss rate is lower or higher than the actual loss rate in the network. For the first case, the distortion caused by packet loss for ROPE, EROPE, and MSVC_OMS is all underestimated; nevertheless, EROPE and MSVC_OMS are still more robust to packet loss than ROPE. For the second case, the distortion is overestimated and it may introduce unnecessary redundancy for error resilience. The worst mismatch in this case is that the network is error-free and the decoder reconstruction is equal to the encoder reconstruction. For example, when the assumed network condition is $p_r = 3\%, p_b = 3\%, k = 5$, the average PSNRs at the encoder for ROPE, MSVC, EROPE, and MSVC_OMS are 34.93 dB, 35.10 dB, 34.28 dB, and 34.38 dB, respectively. When no packet loss happens in the network, the average PSNRs at the decoder are the same as above PSNRs, which all represent quite good video quality. As we have known, for PSNRs higher than a certain threshold, increasing PSNR does not help to enhance the perceptual quality [19], thus the cost of coding efficiency does not affect the perceptual video quality much. And our proposed approaches start to achieve gains when the video transmission suffers packet loss over the network.

7. CONCLUSIONS

In this paper, we propose the error resilient video coding method that enhances the robustness of video to both random loss and burst losses over wireless networks. The method estimates the end-to-end distortion under the specific random and burst loss rates, and applies RD-based mode selection to select the optimal coding mode. The proposed method is applied for single description video coding and multiple description video coding respectively. For single description video coding, we calculate the reconstructed pixel value caused by random loss and burst losses, which results in a more accurately estimation of distortion. The accuracy of the estimation enhances its error robustness over lossy networks. The simulation results show that it achieves better objective and subjective video quality than ROPE for various loss patterns. For multiple description video coding, we estimate the distortion for MSVC and optimally select the coding mode in both descriptions. Compared to MSVC, our approach alleviates the error propagation due to random loss in the two descriptions of MSVC and achieves better performance than MSVC under both random packet loss and burst losses. Note that the complexity of our approaches, which is only incurred at the encoder, is comparable to the ROPE approach.

We have shown that our EROPE approach outperforms the original ROPE approach and our MSVC_OMS approach outperforms the MSVC approach under various loss patterns. Compared between our proposed approaches, we see that EROPE outperforms in objective video quality and MSVC_OMS achieves gains in perceptual video quality. Still, the performance of the two approaches is very close. And they are used for the single and multiple description coding cases respectively.

8. REFERENCES

[1] Q. F. Zhu and L. Kerofsky, "Joint source coding, transport processing, and error concealment for H. 323-based packet video," *Proceedings of SPIE*, vol. 3653, pp. 52, 1998.

[2] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the internet," *Signal Processing: Image Communication*, vol. 15, no. 1, pp. 25–34, 1999.

[3] R. O. Hinds, T. N. Pappas, and J. S. Lim, "Joint block-based video source/channel coding for packet-switched networks," *Proceedings of SPIE*, vol. 3309, pp. 124, 1998.

[4] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust videocommunications over error-prone networks," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 6, pp. 952–965, 2000.

[5] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L video coding in packet loss environment," *Proc. of Int. Packet Video Workshop*, 2002.

[6] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packetloss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, 2000.

[7] Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A.K. Katsaggelos, "VAPOR: Variance-aware per-pixel optimal resource allocation," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 289–299, 2006.

[8] 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 Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, 2007.

[9] T. Wiegand, N. Farber, K. Stuhlmuller, and B. Girod, "Error-resilient video transmission using long-term memorymotion-compensated prediction," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 6, pp. 1050–1062, 2000.

[10] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 252–265, 2001.

[11] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, pp. 57–70, 2005.

[12] Brian A. Heng, John G. Apostolopoulos, and Jae S. Lim, "End-to-end rate-distortion optimized md mode selection for multiple description video coding," *EURASIP Journal on Applied Signal Processing*, pp. Article ID 32592, 12 pages, 2006.

[13] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 7, pp. 861–874, 2008.

[14] Y. Liao and J. D. Gibson, "Enhanced error resilience of video communications for burst losses using an extended rope algorithm," *to appear, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, 2009.

[15] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Visual Communications and Image Processing*, vol. 1, 2001.

[16] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, Jul. 2003.

[17] Y Liao and J. D. Gibson, "Refined error concealment for multiple state video coding over ad hoc networks," *the 42nd Annual Asilomar Conference on Signals, Systems, and Computers*, October 26-29, 2008.

[18] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H. 263 standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 6, no. 2, pp. 182–190, 1996.

[19] J. Hu, S. Choudhury, and J. D. Gibson, "PSNR$_{r,f}$: Assessment of delivered AVC/H. 264 video quality over 802.11a WLANs with multipath fading," *First Multimedia Communications Workshop: State of the Art and Future Directions, Jun*, 2006.

[20] Z. Wang, H.R. Sheikh, and A.C. Bovik, "Objective video quality assessment," *The Handbook of Video Databases: Design and Applications*, pp. 1041–1078, 2003.