DISTORTION ESTIMATION FOR REFERENCE FRAME MODIFICATION METHODS

Amir Naghdinezhad, Fabrice Labeau

Department of Electrical & Computer Engineering, McGill University, Montreal, Canada
emails: amir.naghdinezhad@mail.mcgill.ca, fabrice.labeau@mcgill.ca

ABSTRACT

Due to the transmission of encoded video over error prone channels, using error resilient techniques at the encoder has become an essential issue. These techniques try to decrease the impact of transmission errors by using different approaches such as inserting Intra MacroBlocks (MBs), changing the prediction structure, or considering the channel state in selecting the best MB modes. In this work, we make use of the channel aware mode decision scheme used in the Loss Aware Rate Distortion Optimization (LARDO) method while simultaneously using the prediction structure of the Improved Generalized Source Channel Prediction (IGSCP) technique. In order to combine these two schemes, we estimate the end-to-end distortion for the IGSCP prediction structure in the H.264/AVC encoder. Simulation results, using the JSVM software, demonstrate the effectiveness of our technique for different sequences.

1. INTRODUCTION

In several multimedia applications, compressed video is usually transmitted over channels that are not necessarily error free. Due to the predictive structure used in video compression, transmission errors may propagate temporally and spatially to succeeding frames. This error propagation may result in drastic quality degradation at the decoder side which is disturbing for the users. Various techniques have been proposed to address this problem.

One approach is to add redundancy at the encoder by inserting more Intra MacroBlocks (MB). Intra MBs prevent the propagation of errors from previous frames. The inserted Intra MBs may be selected randomly or at specific places based on optimum techniques [1–3]. The main problem of these methods is that they do not consider the trade of between rate and distortion. As a result, they do not achieve the optimum coding efficiency. Moreover, conventional mode decision methods are designed to achieve maximum performance in error free channels. Several techniques have been proposed to modify this part of the encoder in order to satisfy both coding efficiency and error resilience. These approaches estimate the end-to-end distortion in error prone channels and utilize it in order to consider the state of channel [4–6]. Another approach is to change the prediction structure by modifying the reconstructed frame into a new one which is less vulnerable to transmission errors. The modified reconstructed frame is used as the reference in the prediction of succeeding frames. These techniques are called Reference Frame Modification (RFM) techniques [7–9].

In order to achieve more resilience in coded streams, these methods can be used together. In this work, we combine error resilience mode decision methods with the Improved Generalized Source Channel Prediction (IGSCP) technique [9], which is one of the most efficient RFM techniques, to improve the error robustness. However, modifying the reference frame will change the end-to-end distortion, and to the best of our knowledge none of the above techniques have considered estimating the distortion in combination with reference frame modification methods. In order to estimate the distortion, we modify the LARDO method [6]. This method is reported to be as accurate as other methods [4, 5] in estimation, while it needs less computation resources.

The rest of this paper is organized as follows. In section 2, error resilience mode decision methods and reference frame modification techniques are briefly discussed. The proposed end-to-end distortion estimation for the IGSCP method is introduced in Section 3. In Section 4, simulation results are presented and, finally, Section 5 concludes the paper.

2. BACKGROUND

2.1 Error resilient mode decision methods

In most video coding standards, each MacroBlock can be encoded as Inter, Intra or Skip mode. Furthermore, each MB can be divided into smaller blocks which results in having a choice among 17 different modes in the H.264/AVC standard [10]. The best mode should be selected in a way that satisfies the main goal of an encoder, which is minimizing the total distortion D subject to a bit rate constraint $Rate$, for each MB:

$$\text{minimize } D \text{ subject to } R < \text{Rate}.$$

This problem is typically solved using Lagrangian optimization [11] where the best mode is selected in such a way that the Lagrangian cost function is minimized. This cost function is defined as:

$$J(mode) = D(mode) + \lambda R(mode). \tag{1}$$

$D(mode)$ and $R(mode)$ respectively denote the distortion between the original and the reconstructed MB and the number of bits for coding the prediction residue, selected motion vectors, and MB header corresponding to the mode. λ is computed as:

$$\lambda = 0.85 \times 2^{QP/3},$$

where QP represents the quantization parameter [11]. In this rate-distortion optimization method, the source distortion is only considered for selecting the best mode. Therefore, it achieves the best performance for error free channels. Several researchers modified this method in order to satisfy both coding efficiency and error resilience. The proposed methods are known as error resilience mode decision methods.

By considering the end-to-end distortion instead of source distortion, both coding efficiency and error resilience are addressed. Various approaches have been proposed to estimate the end-to-end distortion in error prone channels. Error robust rate distortion optimization (ER-RDO) [4] addresses the estimation of end-to-end distortion. This method estimates the MB distortion as the average distortions of that MB over K different independent packet loss patterns. Setting K to a large value (>100) leads to an accurate estimation, but requires high computation complexity and massive storage requirement which is not practical in all applications. This method has been adopted in the H.264/AVC test model [12].

Recursive optimal per-pixel estimate (ROPE) [5] is another algorithm that computes the end-to-end distortion at the pixel level. Assuming the reconstructed pixel at the decoder ($\tilde{f}_n^i$) is a random variable, the first and second moments of each pixel are calculated. Comparing to ER-RDO with ($K = 100$), ROPE requires less computational resources, but it is still complex in terms of computation and storage [6].

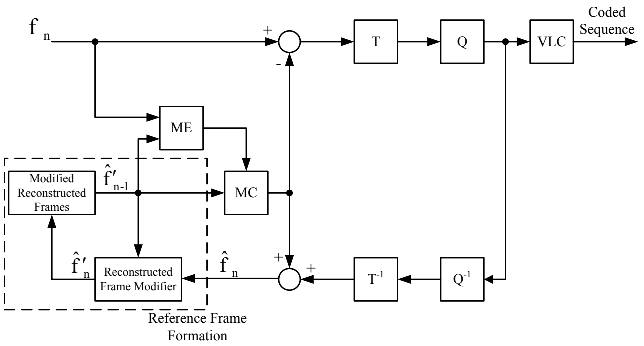


Figure 1: Block diagram of a video encoder with reference frame modification. (ME, MC, T and Q represent Motion Estimation, Motion Compensation, Transform and Quantization respectively).

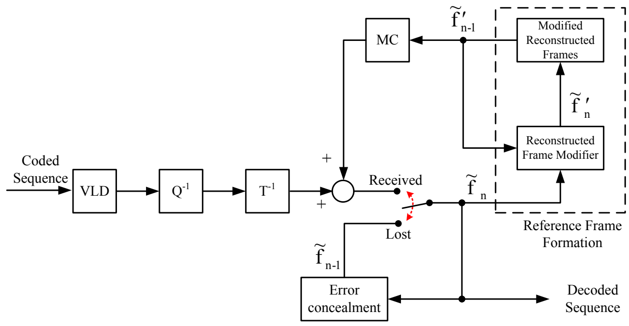


Figure 2: Block diagram of a video decoder with reference frame modification.

The Loss Aware Rate Distortion Optimization (LARDO) [13] method calculates the end-to-end distortion with lower complexity. In this method, the distortion is categorized as source, error propagation, and error concealment distortions. When the MB is received properly at the decoder, the distortion is the summation of the source distortion and the error propagation distortion from the reference block. In case of Intra coding, there is no error propagation from previous frames. Each MB might be lost with probability p. In this case, the only source of distortion is error concealment. The end-to-end distortion of pixel f_n^i, i.e. pixel i in frame n, is computed as:

$$
\begin{aligned}
D(n,i) &= E\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} \\
&= (1-p)\Big(D_s(n,i) + D_{ep}(ref,j)\Big) + p\, D_{ec}(n,i).
\end{aligned}
$$

Source, error propagation and error concealment distortions are represented by $D_s(n,i)$, $D_{ep}(n,i)$ and $D_{ec}(n,i)$ respectively. In order to reduce the storage complexity, the error propagation distortion is estimated recursively per pixel but stored per block for each frame. The stored values are used to calculate the error propagation to future frames. The distortion estimation of LARDO was reported as accurate as ROPE and ER-RDO with $K = 100$ for H.264/AVC but with much less computational complexity [6]. Furthermore, LARDO finds optimum modes by using Lagrangian method (1) by considering the accurately estimated end-to-end distortion. LARDO has been extended to multi-layer coding for SVC [14] and implemented in the SVC reference software, Joint Scalable Video Model (JSVM) [15].

2.2 Reference frame modification methods

Reference frame modification (RFM) methods are techniques that decrease the error propagation by modifying the reconstructed frame which is used as the reference for the succeeding frame. The modified reconstructed frame, is usually formed as a weighted summation of the current reconstructed frame and other values. The block diagrams of a video encoder and decoder with reference frame modification are shown in Figure 1 and Figure 2 respectively. $\hat{f}_n$ and $\tilde{f}_n$ denote the reconstructed frame at the encoder and the decoder. The modified reconstructed value at the encoder and decoder are represented by $\hat{f}'_n$ and $\tilde{f}'_n$ respectively, and the original frame is denoted by f_n. It should be mentioned that the modified reconstructed frames ($\hat{f}'_{n-1}$ and $\tilde{f}'_{n-1}$) are utilized as the reference frames for motion compensation of frame n. It is assumed that when a block is lost, the previous frame is used for error concealment

In leaky prediction [7], the weighted sum of the current reconstructed frame ($\hat{f}_n$) and a constant (K) forms the reference for successive frame. Using this constant leads to exponential decay of the effect of errors propagated from preceding frames, but it also decreases the prediction efficiency. As a result, there would be a trade off between efficient prediction and error robustness. This trade off is controlled by a leaky factor (α).

In Generalized Source Channel Prediction (GSCP) [8], the previous modified reconstructed frame ($\hat{f}'_{n-1}$) is used instead of the constant (K). Since $\hat{f}'_{n-1}$ is more correlated with current frame, using GSCP results in better prediction. This method is defined as:

$$\hat{f}'_n = \alpha \hat{f}_n + (1-\alpha)\hat{f}'_{n-1}.$$

Furthermore, due to the fact that Intra coded pixels do not propagate transmission errors from previous frames, the improved GSCP (IGSCP) technique [9] puts more emphasis on Intra blocks. In this method, if the co-located MB in the previous frame is Intra coded, it will be copied to the modified reconstructed frame. If not, it acts in the same way as the GSCP method. This technique is defined as:

$$\hat{f}'^i_n = \begin{cases} \hat{f}^i_{n-1} & \text{if } f^i_{n-1} \text{ is Intra coded,} \\ \alpha \hat{f}^i_n + (1-\alpha)\hat{f}'^i_{n-1} & \text{otherwise.} \end{cases} \quad (2)$$

3. END-TO-END DISTORTION ESTIMATION FOR IGSCP

In order to improve the performance of error resilience mode decision methods, they can be used in cooperation with reference modification techniques. However, utilizing RFM techniques will change the end-to-end distortion estimation, and to the best of our knowledge, this issue has not been addressed before. In this work, we combined error resilience mode decision with IGSCP. In order to estimate the distortion, we modify the LARDO method, which is as accurate as other distortion estimation methods but with lower computational complexity. It should be mentioned that in RFM methods, if the i^{th} pixel in frame n is Inter coded, the reconstructed frames at the encoder will be:

$$\hat{f}^i_n = \hat{f}'^j_{ref} + \hat{r}^i_n \quad (3)$$

It is assumed that the reference of i^{th} pixel in the n^{th} frame is pixel j in the ref^{th} frame (typically $ref = n-1$). The quantized prediction error is denoted by $\hat{r}^i_n$. At the decoder side, when a pixel is Inter coded, the reconstructed frame will be:

$$\tilde{f}^i_n = \begin{cases} \tilde{f}'^j_{ref} + \hat{r}^i_n & w.p. \quad 1-p, \\ \tilde{f}^i_{n-1} & w.p. \quad p. \end{cases} \quad (4)$$

For an Intra coded pixel, we will have:

$$\tilde{f}^i_n = \begin{cases} \hat{f}^i_n & w.p. \quad 1-p, \\ \tilde{f}^i_{n-1} & w.p. \quad p. \end{cases} \quad (5)$$

where p is the probability of packet loss. It is assumed that simple picture copy is used as the error concealment method at the decoder. In other words, if pixel i is lost, it will be concealed by copying pixel i from frame $n-1$. Assuming the reconstructed pixel at the decoder ($\tilde{f}_n^i$) is a random variable, the end-to-end distortion for Inter coded pixel is calculated by considering whether or not $\tilde{f}_n^i$ is lost, as:

$$\begin{aligned}
D(n,i) &= E\left\{\left(f_n^i-\tilde{f}_n^i\right)^2\right\}\\
&=(1-p)E\left\{\left(f_n^i-(\tilde{f}'^j_{ref}+\hat{r}_n^i)\right)^2\right\}+pE\left\{\left(f_n^i-\tilde{f}_{n-1}^i\right)^2\right\}\\
&=(1-p)E\left\{\left(f_n^i-(\tilde{f}'^j_{ref}+\hat{f}_n^i-\hat{f}'^j_{ref})\right)^2\right\}+pE\left\{\left(f_n^i-\tilde{f}_{n-1}^i\right)^2\right\}\\
&=(1-p)\left(E\left\{\left(f_n^i-\hat{f}_n^i\right)^2\right\}+E\left\{\left(\hat{f}'^j_{ref}-\tilde{f}'^j_{ref}\right)^2\right\}\right)+pE\left\{\left(f_n^i-\tilde{f}_{n-1}^i\right)^2\right\} \qquad (6)\\
&=(1-p)\left(D_s(n,i)+D'_{ep}(ref,j)\right)+pD_{ec}(n,i), \qquad (7)
\end{aligned}$$

where $\hat{f}_n^i$ and $\tilde{f}_n^i$ denote the reconstructed value of pixel i at the encoder and at the decoder, respectively. The modified reconstructed value at the encoder and decoder are represented by $\hat{f}'^i_n$ and $\tilde{f}'^i_n$. (6) is based on the assumption that effects of source distortion at the encoder and error propagation at the decoder are additive. Source, error concealment distortions are represented by $D_s(n,i)$ and $D_{ec}(n,i)$ respectively. $D'_{ep}(n,i)$ is the error propagation distortion in modified reconstructed frame and is calculated differently based on the RFM methods. It should be noted that when the i^{th} pixel in the n^{th} frame is Intra coded, there is no error propagation from previous frames. Therefore, end-to-end distortion is calculated as:

$$\begin{aligned}
D(n,i)&=(1-p)E\left\{\left(f_n^i-\hat{f}_n^i\right)^2\right\}+pE\left\{\left(f_n^i-\tilde{f}_{n-1}^i\right)^2\right\}\\
&=(1-p)D_s(n,i)+pD_{ec}(n,i). \qquad (8)
\end{aligned}$$

Using the estimated distortion, (1) is modified to:

$$\begin{aligned}
J(mode)&=D(mode)+\lambda' R(mode)\\
&=(1-p)\left(D_s(mode)+D'_{ep}(mode)\right)+pD_{ec}(mode)\\
&\quad+\lambda' R(mode),
\end{aligned}$$

where $D(mode)$ is the sum of (7) or (8) over all pixels of the MB, depending on whether the MB is Inter or Intra coded. Since error concealment is independent of selected coding mode, there is no need to calculate D_{ec} for the optimization. Furthermore, based on [6], $\lambda'=(1-p)\lambda$. So, we obtain an equivalent:

$$J'(mode)=D_s(mode)+D'_{ep}(mode)+\lambda R(mode). \qquad (9)$$

Since source distortion can easily be calculated during encoding, the main issue will be the calculation of the error propagation distortion. This distortion is calculated recursively and stored for future use. In IGSCP, the modified reconstructed frame at the decoder ($\tilde{f}'^i_n$) is formed based on the mode of the co-located MB in previous frame and the loss probabilities of current and previous frames. In Table 1, different values of $\tilde{f}'^i_n$ are demonstrated. These cases are calculated by using (2) and by considering the available information at the decoder. It should be noted that when one block is lost, the decoder does not have any information about the mode of that block. Since the number of Intra MBs is usually less than the number of Inter coded MBs, the decoder assumes that the lost MB was coded as Inter. In the following, $D'_{ep}(n,i)$ is calculated for each of four different cases.

Table 1: Different values of current modified reconstructed frame at the decoder ($\tilde{f}'^i_n$).

f_{n-1}^i	$\tilde{f}_n^i$	$\tilde{f}_{n-1}^i$	$\tilde{f}'^i_n$	Case #	Prob.
Inter	not Lost	not Lost	$\alpha\tilde{f}_n^i+(1-\alpha)\tilde{f}'^i_{n-1}$	I	$(1-p)^2$
	not Lost	Lost	$\alpha\tilde{f}_n^i+(1-\alpha)\tilde{f}'^i_{n-1}$	I	$(1-p)p$
	Lost	not Lost	$\tilde{f}'^i_{n-1}$	II	$(1-p)p$
	Lost	Lost	$\tilde{f}'^i_{n-1}$	II	p^2
Intra	not Lost	not Lost	$\tilde{f}_{n-1}^i$	III	$(1-p)^2$
	not Lost	Lost	$\alpha\tilde{f}_n^i+(1-\alpha)\tilde{f}'^i_{n-1}$	IV	$(1-p)p$
	Lost	not Lost	$\tilde{f}_{n-1}^i$	III	$(1-p)p$
	Lost	Lost	$\tilde{f}'^i_{n-1}$	II	p^2

CASE I: In this case, the co-located MB in the previous frame was Inter coded and the current pixel was received correctly. So, $\tilde{f}'^i_n$ is formed as the weighted summation of current reconstructed frame ($\tilde{f}_n^i$) and previous modified reconstructed one ($\tilde{f}'^i_{n-1}$).

$$\begin{aligned}
D'_{ep_I}(n,i)&=E\left\{\left(\hat{f}'^i_n-\tilde{f}'^i_n\right)^2\right\}\\
&=E\left\{\left(\alpha\hat{f}_n^i+(1-\alpha)\hat{f}'^i_{n-1}-\left(\alpha\tilde{f}_n^i+(1-\alpha)\tilde{f}'^i_{n-1}\right)\right)^2\right\}\\
&=E\left\{\left(\alpha(\hat{f}_n^i-\tilde{f}_n^i)+(1-\alpha)(\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1})\right)^2\right\}\\
&=E\left\{\left(\alpha(\hat{f}'^j_{ref}+\hat{r}_n^i-\tilde{f}'^j_{ref}-\hat{r}_n^i)+(1-\alpha)(\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1})\right)^2\right\}\\
&=\alpha^2E\left\{\left(\hat{f}'^j_{ref}-\tilde{f}'^j_{ref}\right)^2\right\}+(1-\alpha)^2E\left\{\left(\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1}\right)^2\right\}\\
&\quad+2\alpha(1-\alpha)E\left\{\hat{f}'^j_{ref}-\tilde{f}'^j_{ref}\right\}E\left\{\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1}\right\} \qquad (10)\\
&=\alpha^2D'_{ep}(ref,j)+(1-\alpha)^2D'_{ep}(n-1,i)\\
&\quad+2\alpha(1-\alpha)Dif_{ep}(ref,j)Dif_{ep}(n-1,i) \qquad (11)
\end{aligned}$$

(10) is based on the assumption that the mean modified reconstructed frame error at pixel i and j are independent. In the case that $i=j$ and $ref=n-1$, (10) is easily modified to $D'_{ep}(n-1,i)$. $Dif_{ep}(n,i)$ is also derived recursively for each case. With similar calculation, we obtain:

$$Dif_{ep\_I}(n,i)=E\left\{\hat{f}'^i_n-\tilde{f}'^i_n\right\}=\alpha Dif_{ep}(ref,j)+(1-\alpha)Dif_{ep}(n-1,i) \qquad (12)$$

CASE II: In this case, the current frame is lost and the decoder knows that the co-located pixel in previous frame was either lost or Inter coded. As a result, the decoder do the concealment by copying the previous modified reconstructed frame.

$$\begin{aligned}
D'_{ep_II}(n,i)&=E\left\{\left(\hat{f}'^i_n-\tilde{f}'^i_{n-1}\right)^2\right\}\\
&=E\left\{\left((\hat{f}'^i_n-\hat{f}'^i_{n-1})+(\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1})\right)^2\right\}\\
&=(\hat{f}'^i_n-\hat{f}'^i_{n-1})^2+E\left\{(\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1})^2\right\}+2(\hat{f}'^i_n-\hat{f}'^i_{n-1})E\left\{\hat{f}'^i_{n-1}-\tilde{f}'^i_{n-1}\right\}\\
&=(\hat{f}'^i_n-\hat{f}'^i_{n-1})^2+D'_{ep}(n-1,i)+2(\hat{f}'^i_n-\hat{f}'^i_{n-1})Dif_{ep}(n-1,i) \qquad (13)
\end{aligned}$$

In the similar way, we will have:

$$Dif_{ep\_II}(n,i)=E\left\{\hat{f}'^i_n-\tilde{f}'^i_{n-1}\right\}=(\hat{f}'^i_n-\hat{f}'^i_{n-1})+Dif_{ep}(n-1,i) \qquad (14)$$

CASE III: The co-located pixel in previous frame was Intra coded and it was correctly received at the decoder. So, the decoder acts the same as the encoder:

$$D'_{ep\_III}(n,i) = E\left\{\left(\hat{f}'^{i}_{n} - \tilde{f}^{i}_{n-1}\right)^2\right\} = E\left\{\left(\hat{f}^{i}_{n-1} - \tilde{f}^{i}_{n-1}\right)^2\right\} = 0 \quad (15)$$

$$Dif_{ep\_III}(n,i) = E\left\{\hat{f}'^{i}_{n} - \tilde{f}^{i}_{n-1}\right\} = E\left\{\hat{f}^{i}_{n-1} - \tilde{f}^{i}_{n-1}\right\} = 0 \quad (16)$$

CASE IV: In this case, f^{i}_{n-1} was Intra coded, but since it was lost during the transmission, the decoder considers that as an Inter coded pixel. We will have:

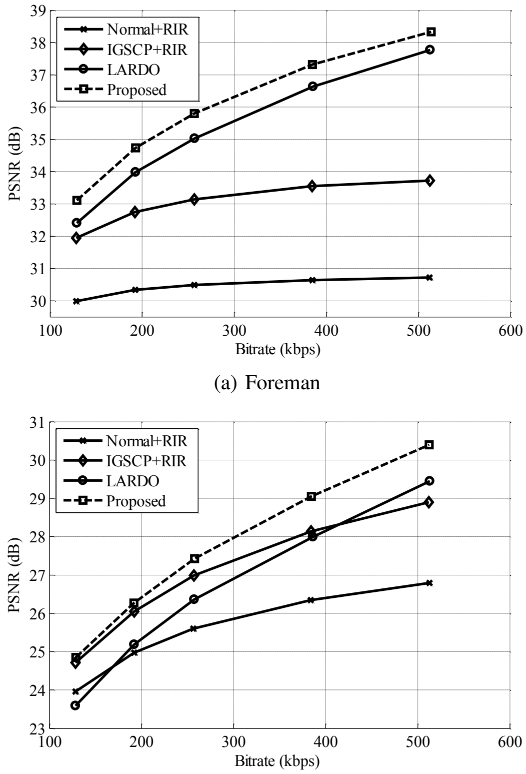
$$\begin{aligned}
D'_{ep_IV}(n,i) &= E\left\{\left(\hat{f}'^{i}_{n} - \left(\alpha \tilde{f}^{i}_{n} + (1-\alpha)\tilde{f}'^{i}_{n-1}\right)\right)^2\right\} \\
&= E\left\{\left(\hat{f}^{i}_{n-1} - \left(\alpha \tilde{f}^{i}_{n} + (1-\alpha)\tilde{f}'^{i}_{n-1}\right) \right.\right. \\
&\qquad \left.\left. + \alpha\left(\hat{f}^{i}_{n} - \hat{f}^{i}_{n}\right) + (1-\alpha)\left(\hat{f}'^{i}_{n-1} - \hat{f}'^{i}_{n-1}\right)\right)^2\right\} \\
&= E\left\{\left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right) \right.\right. \\
&\qquad \left.\left. + \alpha\left(\hat{f}^{i}_{n} - \tilde{f}^{i}_{n}\right) + (1-\alpha)\left(\hat{f}'^{i}_{n-1} - \tilde{f}'^{i}_{n-1}\right)\right)^2\right\} \\
&= E\left\{\left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right) \right.\right. \\
&\qquad \left.\left. + \alpha\left(\hat{f}'^{j}_{ref} + \hat{r}^{i}_{n} - \tilde{f}'^{j}_{ref} - \hat{r}^{i}_{n}\right) + (1-\alpha)\left(\hat{f}'^{i}_{n-1} - \tilde{f}'^{i}_{n-1}\right)\right)^2\right\} \\
&= E\left\{\left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right) \right.\right. \\
&\qquad \left.\left. + \alpha\left(\hat{f}'^{j}_{ref} - \tilde{f}'^{j}_{ref}\right) + (1-\alpha)\left(\hat{f}'^{i}_{n-1} - \tilde{f}'^{i}_{n-1}\right)\right)^2\right\} \\
&= \left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right)\right)^2 + \alpha^2 D'_{ep}(ref,j) + (1-\alpha)^2 D'_{ep}(n-1,i) \\
&\quad + 2\left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right)\right)\left((1-\alpha)Dif_{ep}(n-1,i) + \alpha Dif_{ep}(ref,j)\right) \\
&\quad + 2\alpha(1-\alpha)Dif_{ep}(n-1,i)Dif_{ep}(ref,j) \qquad (17)
\end{aligned}$$

Using similar derivations, we obtain:

$$\begin{aligned}
Dif_{ep_IV}(n,i) &= E\left\{\hat{f}'^{i}_{n} - \left(\alpha \tilde{f}^{i}_{n} + (1-\alpha)\tilde{f}'^{i}_{n-1}\right)\right\} \qquad (18) \\
&= \left(\hat{f}^{i}_{n-1} - \left(\alpha \hat{f}^{i}_{n} + (1-\alpha)\hat{f}'^{i}_{n-1}\right)\right) + (1-\alpha)Dif_{ep}(n-1,i) + \alpha Dif_{ep}(ref,j)
\end{aligned}$$

It should be mentioned that if the current pixel is coded as Intra, there is no error propagation from previous frames. Consequently, $Dif_{ep}(ref,j)$ and $D'_{ep}(ref,j)$ are equal to zero in (11)-(18). By considering the four different cases ((11)-(18)), $D'_{ep}(n,i)$ and $Dif_{ep}(n,i)$ are calculated respectively as:

$$\begin{aligned}
D'_{ep}(n,i) &= E\left\{\left(\hat{f}'^{i}_{n} - \tilde{f}'^{i}_{n}\right)^2\right\} \\
&= \begin{cases} p^2 D'_{ep_II}(n,i) + p(1-p)D'_{ep_IV}(n,i) & \text{if } f^{i}_{n-1} \text{ is Intra,} \\ (1-p)D'_{ep_I}(n,i) + pD'_{ep_II}(n,i) & \text{otherwise.} \end{cases} \\
Dif_{ep}(n,i) &= E\left\{\hat{f}'^{i}_{n} - \tilde{f}'^{i}_{n}\right\} \\
&= \begin{cases} p^2 Dif_{ep_II}(n,i) + p(1-p)Dif_{ep_IV}(n,i) & \text{if } f^{i}_{n-1} \text{ is Intra,} \\ (1-p)Dif_{ep_I}(n,i) + pDif_{ep_II}(n,i) & \text{otherwise.} \end{cases}
\end{aligned}$$



(a) Foreman

(b) Mobile

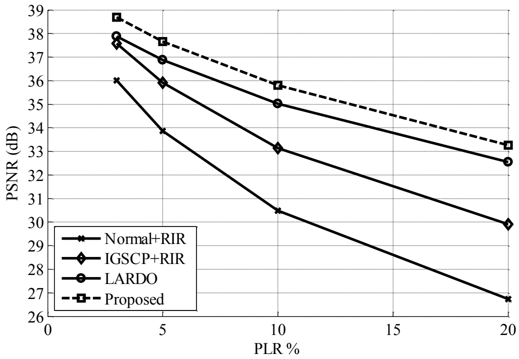
Figure 3: Rate distortion curves for different methods for (a) Foreman sequence (b) Mobile sequence with QCIF size and 15 fps with packet loss rate of 10%.

4. SIMULATION RESULTS

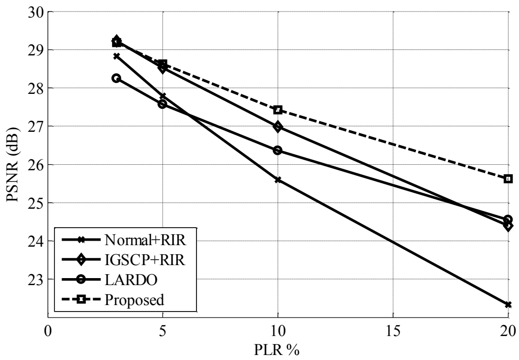
In order to compare the performance of different techniques, a set of simulations were conducted with the following conditions, which are mostly based on the common conditions for error resilience simulations [16]:

- JSVM 9.15 was used as the encoder and the decoder.
- IPPP prediction structure with GOP size of 1 (single layer).
- 128 encoded pictures were repeated 40 times (5120 pictures) and transmitted through a packet loss channel.
- The transmitted video is decoded and the average peak signal to noise ratio (PSNR) is calculated over all 5120 pictures.
- Tested packet loss rates (PLR) are %3, %5, %10, %20.
- Standard sequences such as "Mobile", "Foreman", "Mother&Daughter", "Stefan", "News", "Akiyo" and "Flower" were encoded with QCIF @ 15fps and bit rates of 128, 192, 256, 384 and 512 kbps.
- Each slice contains one row of MBs.
- At the decoder side, the picture copy method concealed the lost picture or slices by using the previous frame.

Figure 3 shows the rate distortion curves of each method for "Foreman" and "Mobile" sequences with QCIF size and frame rate of 15 fps. The packet loss rate is set to 10% and α for IGSCP is set to 0.8. In order to improve the error robustness, the Normal and IGSCP methods are combined with random Intra refresh (RIR) with 10% Intra rate. As it was reported in [9], the simulation results show that using IGSCP improves the robustness of normal coding. Also, LARDO performs better than random Intra refresh method by using the estimated end-to-end distortion in mode decision, which shows the effect of source-channel optimized mode decision in increasing the robustness of the video stream. It can be noticed that by increasing the bit rates, the LARDO and the proposed methods, outperform



(a) Foreman



(b) Mobile

Figure 4: PSNR vs. Packet loss rate for different methods for (a) Foreman sequence and (b) Mobile sequence with QCIF size and 15 fps at 256 kbps.

other two techniques. The reason is that these two methods make use of optimal mode decision and consequently, they add more Intra MBs in higher bit rates.

The performance of each technique in different packet loss scenarios is illustrated in Figure 4. "Foreman" and "Mobile" sequences with QCIF size are used as the input videos. The sequences are encoded at 15 fps and bit rate of 256 kbps. It can be observed that the proposed method, which is the combination of IGSCP and LARDO, performs better than other techniques. Considering all packet loss scenarios and different packet loss rates, gains of up to 1.1 dB and 4.3 dB over LARDO and IGSCP can be achieved. It should be mentioned that for "Stefan" sequence, the improvement of the proposed method over LARDO was marginal (about 0.4 dB).

5. CONCLUSION

In order to lessen the impact of transmission errors on the quality of the decoded video, various techniques have been proposed. In this work, we combine reference frame modification techniques with error resilience mode decision. Since RFM methods change the prediction structure, new equations for estimation of the end-to-end distortion were derived. The simulation results show that the proposed technique achieve better performance comparing to both LARDO and IGSCP techniques.

REFERENCES

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

[2] Q. Chen, Z. Chen, X. Gu, and C. Wang, "Attention-based adaptive intra refresh for error-prone video transmission," *IEEE Communications Magazine*, vol. 45, no. 1, pp. 52–60, 2007.

[3] H. J. Ma, F. Zhou, R. X. Jiang, and Y. W. Chen, "A network-aware error-resilient method using prioritized intra refresh for wireless video communications," *Journal of Zhejiang University*, vol. 10, no. 8, pp. 1169–1176, 2009.

[4] T. Stockhammer, T. Wiegand, and D. Kontopodis, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," in *Proceeding of IEEE International Packet Video Workshop*, Pittsburgh, USA, Apr. 2002.

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

[6] 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.

[7] H. Huang, C. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 372–385, 2002.

[8] H. Yang and K. Rose, "Generalized source-channel prediction for error resilient video coding," in *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, Oct. 2006, pp. 533–536.

[9] R. Satyan, S. Nyamweno, and F. Labeau, "Novel prediction schemes for error resilient video coding," *Signal Processing: Image Communication*, vol. 25, no. 9, pp. 648 – 659, 2010.

[10] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[11] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," in *Proceeding of IEEE International Conference on Image Processing (ICIP)*, Greece, Oct. 2001, pp. 542–545.

[12] "H.264 reference software (ver JM 13.2)," *[Available Online] http://iphome.hhi.de/suehring/tml/*.

[13] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proceedings of the International Conference on Image Processing (ICIP)*, Singapore, Oct. 2004, pp. 163–166.

[14] Y. Guo, Y.-K. Wang, and H. Li, "Error resilient mode decision in scalable video coding," in *Proceedings of the International Conference on Image Processing (ICIP)*, Atlanta, USA, Oct. 2006, pp. 2225–2228.

[15] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model 11 (JSVM 11)," JVT-X202, Jul. 2007.

[16] Y. Wang, S. Wenger, and M. Hannuksela, "Common conditions for SVC error resilience testing," ISO/IEC JTC 1/SC29/WG 11, JVT-P206d1, Jul. 2005.