Recursive Mode Decision for Robust Video Transmission over Bursty Packet-Loss Wireless Networks

Feng Huang[1] Lifeng Sun[1] Yuzhuo Zhong[1]

[1] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Abstract

Packet losses over wireless networks often occur in bursts, which may cause substantial degradation to the transmitted video quality. In this paper, an adaptive RD-based mode decision scheme is proposed for robust video transmission over wireless networks. The proposed encoder captures the influence of encoder performance, packet loss, error concealment and error propagation to recursively estimate the distortion at the decoder. The estimation is then used for switching between intra-coding and inter-coding modes per macro-block. To reflect the sharp quality degradation caused by bursty packet losses and stop error propagation effectively, our scheme uses the worst deterioration of the reference frame to estimate the distortion. To avoid the excessive reduction of coding efficiency, the worst scenario is assumed only when the expected accumulated error of the reference frame is large. The error concealment technique is also tailored to the bursting characteristic of packet losses. So, our mode decision scheme finds a good tradeoff between error resilience and coding efficiency. The simulation results under different conditions show that our scheme has robust error resilient ability and good adaptability. It yields consistent and significant gains over the competing methods in a wide range of bitrate, transition probability and burst length. Moreover, our scheme has low computational complexity and does not require an accurate estimation of packet loss rate.

Keywords: Recursive Expected and Worst Pixel Estimate (REWPE), intra/inter mode decision, bursty packet losses, error propagation, H.264.

1. Introduction

Recently, there has a growing interest in real-time video transmission over wireless networks. Yet, the wireless channel is highly time-varying due to fading and interference effects. Therefore, packet video transmission over wireless networks is expected to experience packet losses due to either temporary link outages or fading-induced bit errors, both of which could cause bursty packet losses and thus cause substantial quality degradation to the transmitted video. As a result, the need exists for video coding and transmission schemes that not only provide efficient compression performance but also provide relatively robust transport performance in the presence of link error effects resulting in lost packets. Reliable transport protocols like TCP/IP recover from loss by using acknowledgements and retransmissions. However, the resulting latency is generally too large for real-time surveillant or interactive applications, where late packets are effectively lost. Thus, we only address the problem of robust video coding over bursty packet-loss networks without the support of retransmissions.

Block-based predictive video coding algorithms employ inter-frame prediction to reduce temporal dependencies. An error in the compressed video bit stream will propagate until some of the blocks are intra-coded without reference to previous frames. Coding a complete video frame in the intra-mode is an effective method to stop error propagation. Unfortunately, intra-coded frames are usually 4 to 8 times the size of inter-coded frames, yielding undesirably long delays when transmitting live video over a bandwidth-limited wireless channel. Moreover, the frequent transmission of these frames will potentially disrupt the transmissions of all the users in a bandwidth-limited environment such as wireless networks. Therefore, only some of the blocks within a frame are intra-coded to achieve both compression efficiency and error resilience.

The recursive optimal per-pixel estimate (ROPE) algorithm [1] estimates all the distortion of the decoder frame reconstruction due to quantization, error propagation and error concealment. The estimation is then used for switching between intra-coding and inter-coding modes per macro-block in a rate-distortion optimization framework. Weighted by the loss probability, ROPE combines the previous frame due to quantization and the previous frame due to concealment to produce the estimated frame.

However, the smoothed estimation does not effectively reflect the sharp quality degradation of the previous frame caused by bursty packet losses in wireless networks. This affects the accuracy of the estimation and the intra/inter mode decision of the current frame. Based on the statistical model of error propagation and the introduction of the concealment candidate image (CCI) [2], the REWPE (Recursive Expected and Worst Pixel Estimate) algorithm is proposed for mode decision, which has the following considerations:

- To reflect the sharp quality degradation and stop error propagation more effectively, we use the worst deterioration of the reference frame to estimate the distortion.
- To avoid the excessive reduction of coding efficiency, we introduce the worst deterioration of the reference frame to the MB mode decision of the current frame only when the expected accumulated error of the reference frame is large.
- To apply the ROPE-styled framework to wireless networks more effectively, the error concealment technique is tailored to the bursting characteristic of packet losses.

The aim of this paper is to adapt such a ROPE-styled framework to a bandwidth-limited wireless network with bursty losses. The rest of the paper is organized as follows. In Section 2, we use a Gilbert model to describe bursty packet losses. In Section 3, we present a video distortion model, which captures the influence of encoder performance, packet loss, error concealment and error propagation. This is followed with a description of our proposed RD-optimized mode decision scheme, called REWPE (Recursive Expected and Worst Pixel Estimate). We present simulation results to demonstrate the superiority of the proposed methods over other conventional or state-of-the-art techniques in Section 5. Finally, conclusions are drawn.

2. Packet Loss Model

Most prior work, for simplicity, implicitly assumed that the burst length does not matter and used the average packet loss rate as the only parameter in considering the effect of packet loss upon the transmitted video quality. Recently, work reported in [3] demonstrated that the burst length is indeed important and its effects should be carefully considered.

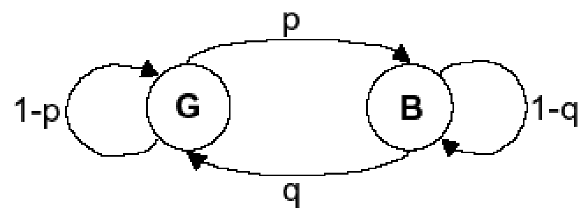

Fig. 1 two-state discrete Markov chain

We model bursty packet losses with the two-state discrete Markov chain, known as Gilbert's Model [4]. As shown in Fig. 1, the model has two states as *G* and *B*, representing good and bad conditions respectively. When the state is *G*, i.e., good state, a packet is transmitted successfully. The transmission fails if the channel is in bad state, i.e., *B*. Success or failure of the transmission only depends on the current state. p and q are the state transition probabilities. The transition between two states occurs at each packet instant. For $P(G)+P(B)=1$ being steady-state state probabilities, the state transition matrix is given by

$$P_T = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}. \quad (1)$$

Thus, average loss probability $P(B)$, and average burst length $L(B)$ can be calculated as

$$P(B) = \frac{p}{p+q}, L(B) = \frac{1}{q}. \quad (2)$$

3. Video Distortion Model

3.1. Preliminaries

In our coding system, we do not employ a feedback channel from the decoder to the encoder. Although very effective error recovery mechanisms can be implemented [5] with a feedback channel, the feedback channel may introduce additional delay and complexity. Moreover, it is not suitable for multipoint communication. Without a feedback channel, the encoder still employs network feedback, which is defined as information passed from the network to the application and includes packet loss rate, delay, jitter, etc.

In our coding system, we form a slice from a row of macroblocks, and assume that each slice is carried in a separate packet. In this setting, the loss rate of a MB (macroblock) equals $P(B)$. We assume that $P(B)$ is available at the encoder via network feedback.

Let f_n^i denote the original value of pixel i in frame n, and let $\hat{f}_n^i$ denote its *encoder* reconstruction. The constructed value at the *decoder*, possibly after error concealment, is denoted by $\tilde{f}_n^i$. Recall that for the encoder, $f_n^{\sim i}$ is a random variable. Using the SAD (the sum of absolute difference) as distortion metric, the overall expected distortion for this pixel is

$$D_n = E\{| f_n^i - f_n^{\sim i} |\} = D1_n + D2_n \text{ ,(3)}$$

where $D1_n$ is the distortion calculated assuming the corresponding MB of the current frame n, is correctly received, and $D2_n$ is the concealment distortion assuming the corresponding corrupted MB. As follows, we consider two cases depending on whether the pixel belongs to an intra-coded MB or an inter-coded MB.

3.2. Distortion in Intra-coded MB

Let us first assume that an intra MB is received correctly at the decoder. We thus have $f_n^{\sim i} = f_n^{\wedge i}$, and the probability of this event is *1-P(B)*. Then we derive $D1_n$ for the intra-coded MB

$$D1_n(i) = (1 - P(B)) | f_n^i - f_n^{\wedge i} | \text{ .(4)}$$

If the slice (packet) containing the MB is lost, the loss probability of its previous (above) slice is *1-1/q* deduced from (1)(2), which reflects the bursty nature of the packet losses. Therefore, the motion vectors of the nearest MBs in the previous slice are usually not available for the error concealment of the lost MB. We set the estimated motion vector of the lost MB to zero, which associates pixel *i* in the current frame with pixel *i* in the reference (previous) frame. We thus have $f_n^{\sim i} = E\{f_{n-1}^{\sim i}\}$ and the $D2_n$ for the intra-coded MB is computed as

$$D2_n(i) = P(B) | f_n^i - E\{f_{n-1}^{\sim i}\} | \text{ .(5)}$$

3.3. Distortion in Inter-coded MB

In inter-mode, the encoder predicts the current pixel *i* from the pixel *j* in the previous frame by the true motion vector of the MB. Thus, the encoder prediction of this pixel is $f_{n-1}^{\wedge j}$. The motion prediction residue is then compressed and we denote the quantized residue by $e_n^{\wedge i}$. Thus we have $e_n^{\wedge i} = f_n^{\wedge i} - f_{n-1}^{\wedge j}$. If the current MB is received correctly, the decoder has access to both $e_n^{\wedge i}$ and the motion vector. We thus have $f_n^{\sim i} = e_n^{\wedge i} + E\{f_{n-1}^{\sim j}\}$. Then we derive $D1_n$ for the inter-coded MB

$$D1_n(i) = (1 - P(B)) | f_n^i - (e_n^{\wedge i} + E\{f_{n-1}^{\sim j}\}) | \text{ .(6)}$$

If the packet containing the inter-coded MB is lost, the decoder performs error concealment in a manner identical to that of an intra-coded MB. So the concealment distortion $D2_n$ is identical for both intra-mode and inter-mode. Note that these recursions are performed at the encoder in order to calculate the expected distortion at the decoder. The encoder can exploit this result directly for mode switching.

4. REWPE Algorithm for Mode Decision

4.1. RD-Optimized Mode Decision

Mode decision within a rate-distortion framework is a useful tool for video compression in error-free channels [6]. The overall expected distortion computed by our video distortion model is incorporated within the rate-distortion framework in order to automatically choose the number and the location of the intra-coded MB's. The conventional rate-distortion problem in video coding is described as an unconstrained Lagrange minimization, $J=D+\lambda R$. Since individual MB contributions to this cost are additive, the optimal encoding mode for each MB is chosen by

$$\min_{mode}(J_{MB}) = \min_{mode}(D_{MB} + \lambda R_{MB}) = \min_{mode}(D1_{MB} + D2_{MB} + \lambda R_{MB}) \text{ .(7)}$$

Considering $D2_n$ is identical regardless of the MB mode, we simplify the computation of the cost as

$$\min_{mode}(J_{MB}) = \min_{mode}(D1_{MB} + \lambda R_{MB}) \text{ ,(8)}$$

where the distortion of the MB is the sum of the distortion contributions of the individual pixels

$$D1_{MB} = \sum_{i \in MB} D1_n(i) \text{ .(9)}$$

Note that R_{MB} denotes the bit rate, and λ is Lagrange multiplier.

H.264 is the ITU-T's new video compression recommendation that significantly outperforms all previous standards. In terms of the MB mode, H.264 supports not only the multiple inter-modes (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4) with different block types, but also skip-mode and intra-modes (4x4, 16x16). A loss-aware RD-optimized MB mode selection algorithm (LA-RDO), has been selected into H.264 reference implementation [7]. In the encoder, *K* copies of the random channel behavior and the hypothetical decoder are manipulated. The expected distortion at the decoder can be estimated rather accurately if *K* is chosen large enough. However, the added complexity in the encoder is obviously at least *K* times the decoder complexity. Therefore, LA-RDO is limited in practical implementations. Furthermore, the simulations in Section 5 show that the performance of LA-RDO is not as good as the one of our proposed REWPE.

4.2. Concealment Candidate Image (CCI)

After the previous frame *n* is encoded, we calculate $E\{f_n^{\sim i}\}$ of intra-coded MB as

$$E\{f_n^{\sim i}\} = (1 - P(B)) f_n^{\wedge i} + P(B) E\{f_{n-1}^{\sim i}\},(10)$$

for the mode decision of frame $n+1$. Similarly, we can calculate $E\{f_n^{\sim i}\}$ of inter-coded MB as

$$E\{f_n^{\sim i}\} = (1 - P(B))(e_n^{\wedge i} + E\{f_{n-1}^{\sim j}\}) + P(B) E\{f_{n-1}^{\sim i}\}.(11)$$

As most of link layer techniques target at reducing packet loss rate, it is reasonable to assume that $P(B)$ is less than 20% in video transmission. So, the smoothed estimation $E\{f_n^{\sim i}\}$ cannot effectively reflect the sharp quality degradation to the frame n due to bursty packet losses in wireless networks. This affects the accuracy of the estimation and the intra/inter mode decision of the frame $n+1$.

To reflect the sharp quality degradation and stop error propagation more effectively, we take into account the worst deterioration of the frame n when estimating the overall distortion of the frame $n+1$. It is advisable to displace $E\{f_n^{\sim i}\}$ in the RD-Optimized mode decision of frame $n+1$. Thus, we introduce the concealment candidate image (CCI) [2] of frame n. More specifically, each pixel in CCI is computed as

$$C\{f_n^{\sim i}\} = E\{f_{n-1}^{\sim i}\},(12)$$

whether the pixel is in intra-coded MB or not. In essence, each pixel value in CCI-n is the same as the concealed one due to packet loss. Since $C\{f_n^{\sim i}\}$ reflects the worst deterioration of frame n, we refine $D1_{n+1}$ of inter-coded MB (6) as

$$D1_{n+1}(i) = (1 - P(B)) | f_{n+1}^{i} - (e_{n+1}^{\wedge i} + C\{f_n^{\sim j}\}) |,(13)$$

while $D1_{n+1}$ of intra-coded MB follows Eq. (4).

4.3. Proposed REWPE Algorithm

Since $D1_{n+1}$ usually increases after the refinement from Eqs. (6) to (13), the encoder will select more intra MB mode for frame $n+1$ to stop error propagation. To avoid the excessive reduction of coding efficiency due to selecting more intra MBs, we develop an adaptive rule to decide when to displace the averaged estimation $E\{f_n^{\sim i}\}$ with the worst estimation $C\{f_n^{\sim i}\}$.

We compute the mean accumulated error (MAE) of frame n as

$$MAE_n = \sum_{i \in Frame-n} | f_n^{\wedge i} - E\{f_n^{\sim i}\} | /(XY),(14)$$

where X, Y is the width and height of the video frames. MAE_n means how reliable $E\{f_n^{\sim i}\}$ estimates the decoder reconstruction. The larger MAE_n is, the less reliable $E\{f_n^{\sim i}\}$ reflects the decoded video quality via networks. Thus, If MAE_n is larger than a typical threshold T_n, we compute $D1_{n+1}$ via Eq. (13); otherwise, we compute $D1_{n+1}$ via Eq. (6). The Threshold T_n is represented as a linear equation:

$$T_n = \alpha + \beta n / n_{GOP},(15)$$

where n_{GOP} denotes the length of a group of pictures (GOP) and α, β are determined empirically. We will investigate the effect of the threshold T_n on the performance gain through simulations.

In conclusion, the proposed algorithm, called REWPE (Recursive Expected and Worst Pixel Estimate), is described more formally as follow.

Step 1: For intra mode, calculate $D1_{MB}$ for the MB of frame $n+1$ via Eqs. (4), (9).

Step 2: For inter mode, if $MAE_n > T_n$, calculate $D1_{MB}$ for the MB of frame $n+1$ via Eqs. (13), (9). Otherwise, calculate $D1_{MB}$ for the MB of frame $n+1$ via Eqs. (6), (9).

Step 3: Select the best mode for the MB of frame $n+1$ via Eq. (8).

Step 4: If any MB of frame $n+1$ has not been encoded via Step 1-3, then go to Step 1.

Step 5: Update $E\{f_{n+1}^{\sim i}\}$, $C\{f_{n+1}^{\sim i}\}$ for each pixel of frame $n+1$. Increase n, and then go to Step 1.

5. Simulations

We implemented the REWPE algorithm for mode decision by appropriately modifying the JVT Software Version M8.6. We assume the RTP payload format for packetizing the H.264 video stream [8]. The color QCIF sequence is encoded at the specified bit rate or at the specified quantizing parameter by the H.264 baseline encoder. The peak signal-to-noise ratio (PSNR) of the decoder luminance reconstruction is computed for each frame and averaged over the whole sequence. We average the PSNR over 20 different loss realizations with the same model parameters to get a convictive PSNR.

We compare the performance of four RD frameworks for MB mode decisions. They are listed as follows.

- The encoder doesn't consider packet losses in the H.264 draft (RDO) [7].
- The encoder simulates the error-prone channel by multiple decoders [7] and makes the mode decision. This is the loss-aware RD optimized mode selection algorithm (LA-RDO).
- The H.264 encoder adopts the recursive optimal per-pixel estimate (ROPE) algorithm for mode switching.
- Our proposed recursive expected and worst pixel estimate (REWPE) algorithm.

Fig. 2 shows the performance of the three RD-optimized mode selection schemes versus transition probability p for the *foreman* sequence (the performance of LA-RDO is shown later together with its complexity). The average packet loss rate varies with the transition probability p when the transition probability q is fixed. The initial state of the channel is good. In Fig. 3, we present the results on the *carphone* sequence. We use 75 even frames of the first 150 frames from the two sequences. This means the frame rate reduces from *30f/s* to *15f/s*. The number of previous frames used for inter motion search is set to *3*. The result supports the claim that the proposed REWPE algorithm yields consistent and significant gains over the two methods: ROPE and RDO. In Fig. 2 and Fig. 3, we also plot three different REWPE implementations with respect to the threshold T_n. The best performance is achieved when $\alpha=3.0$ and $\beta=1.0$, that is, $T_n^* = 3.0 + n / n_{GOP}$. The optimum threshold T_n^* is used for the residual experiments.

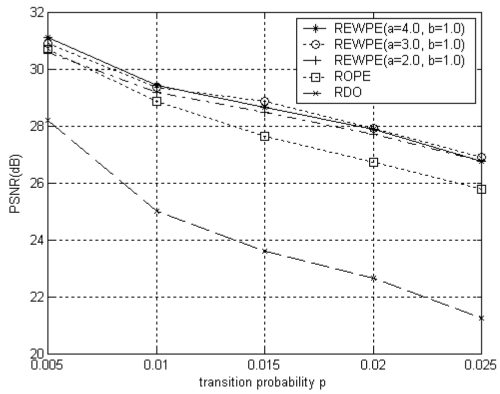

Fig. 2 *PSNR* versus transition probability *p* of three mode decision schemes. *Foreman* QCIF sequence, transition probability *q=0.1*, *bitrate=90kbps*

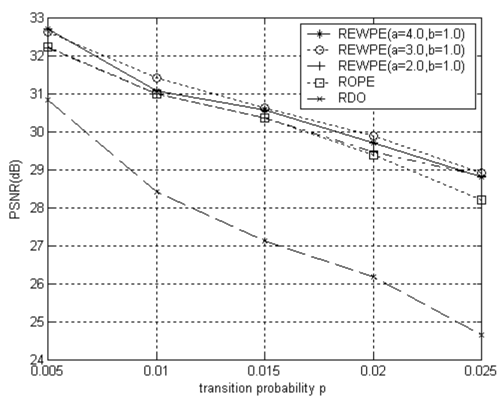

Fig. 3 *PSNR* versus transition probability *p* of three mode decision schemes. *Carphone* QCIF sequence, transition probability *q=0.1*, *bitrate=60kbps*

We then compare the proposed REWPE mode switching algorithm to the other methods in terms of *PSNR* versus average burst length *1/q*. To investigate the effect of the average burst length on the performance more effectively, the average packet loss rate (*P(B)=p/(p+q))* is a constant *0.1*. Fig. 4 shows that the performance of the three methods slowly decreases as the average burst length grows. Fig. 4 also shows that REWPE increases its performance gains over ROPE and RDO as the average burst length grows. The result means that the REWPE algorithm is more effective when packet losses are bursty. Thus, video transmission over wireless networks could benefit from REWPE to tackle bursty losses.

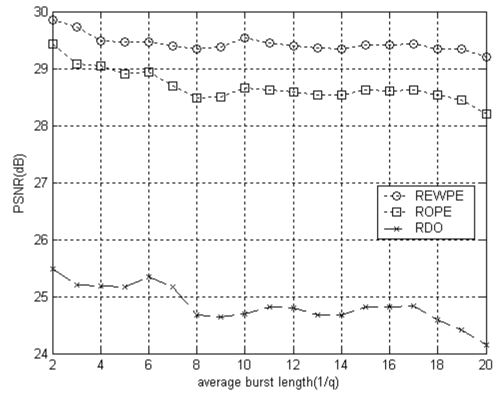

Fig. 4 *PSNR* versus *average burst length1/q* of three mode decision schemes. *Foreman* QCIF sequence, transition probability *p=q/9*, *bitrate=90kbps*

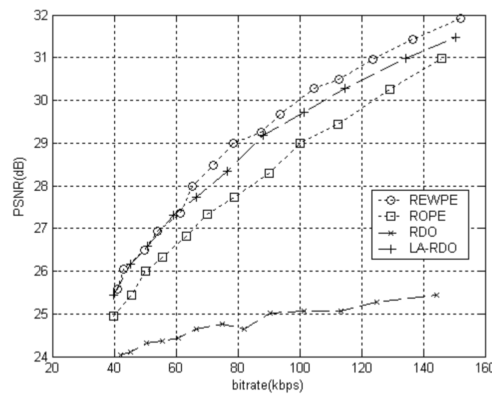

Fig. 5 *PSNR* versus *bitrate* of four mode decision schemes. *Foreman* QCIF sequence, transition probability *p=0.*01, *q=0.1*

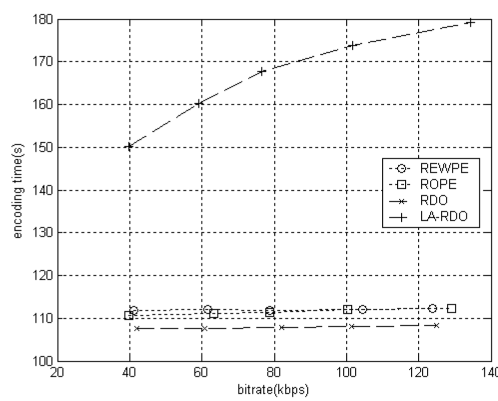

Fig. 6 *encoding time* versus *bitrate* of four mode decision schemes. *Foreman* QCIF sequence, *75* even frames, transition probability *p=0.*01, *q=0.1*

Fig. 5 shows the performance versus bitrate on the *foreman* sequence. We change the bitrate by specifying the quantizing parameter. The performance of LA-RDO is also evaluated through this simulation. The number of hypothetical decoders, *K*, is set to *10*.

Although the performance of LA-RDO approaches the one of REWPE, its computation complexity is significantly higher than others', which is presented in Fig. 6. Our computer configuration is as follows. CPU: PM 1.6G, DDR RAM: 512 MB, OS: WindowsXP Professional.

In practical, it is possible to encounter mismatch between the loss rate assumed at the encoder and the actual loss rate during the transmission. Fig. 7 describes the performance when the encoder assumes *p=q/9=0.011*, that is *P(B) = 0.1*, while the actual *p* is *0.005*, *0.01*, *0.015*, *0.02* or *0.025*. The proposed REWPE also achieves performance gains over the other competing methods, especially when the actual loss rate is not less than the hypothetical loss rate. Since REWPE takes into account the worst deterioration of the reference frame, it could put up with more unexpected packet losses.

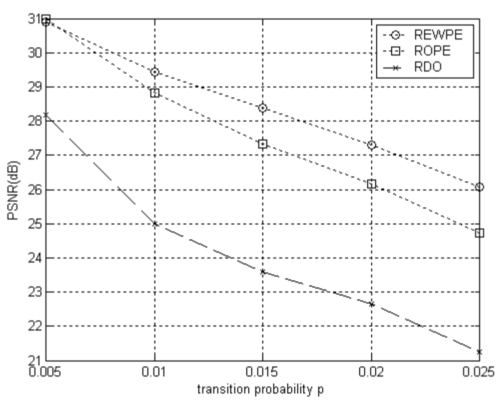

Fig. 7 *PSNR* versus actual transition probability *p*. *Foreman* QCIF sequence, transition probability *q=0.1*, average loss rate *P(B)=0.1* assumed at the encoder, *bitrate=90kbps*

6. Conclusion

In this paper, an adaptive RD-based mode decision scheme is proposed for robust video transmission over wireless networks with bursty packet losses. Similar to the recursive optimal pixel estimate (ROPE) algorithm, we intend to make our H.264 error-resilient encoder 'farseeing' enough to reconstruct the distortion at the decoder.

We highlight several features of our scheme to distinguish from ROPE. Firstly, to reflect the sharp quality degradation caused by bursty packet losses and stop error propagation more effectively, it uses the worst deterioration of the reference frame to estimate the distortion. Secondly, to avoid the excessive reduction of coding efficiency, the first feature is applied only when the expected accumulated error of the reference frame is large. Thirdly, the error concealment technique is tailored to the bursting characteristic of packet losses.

As the key algorithm of our scheme, REWPE finds a good tradeoff between error resilience and coding efficiency. The simulation results under different simulation conditions show that the proposed algorithm has robust error resilient ability and good adaptability. It yields consistent and significant gains over the competing methods in a wide range of bitrate, transition probability and burst length. Moreover, our scheme has low computational complexity and does not require an accurate estimation of packet loss rate.

7. References

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

[2] W. J. Chen, J. N. Hwang, "The CBERC: A Content-Based Error-Resilient Coding Technique for Packet Video Communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, August 2001.

[3] Y. J. Liang, J. G. Apostolopoulos and B. Girod, "Analysis of Packet Loss for Compressed Video: Does Burst-Length Matter?", in the *Proceeding of ICASSP2003*, Apr. 2003.

[4] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communications channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 163–171, Feb. 1995.

[5] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," in *Proc. of IEEE Special Issue Video Mobile Multimedia*, 1999.

[6] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74–90, Nov. 1998.

[7] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC Draft Text, Joint Video Team document JVT-E146d37, Oct. 2002.

[8] S. Wenger, T. Stockhammer, and M. M. Hannuksela, "RTP payload format for H.264 video," in Internet Draft, Work in Progress, Mar. 2003, Draft-wenger-avt-rtp-h264-01.txt.