Robust Video Region-of-interest Coding Based on Leaky Prediction

Qian Chen, Xiaokang Yang, *Senior Member, IEEE*, Li Song, and Wenjun Zhang, *Member, IEEE*

***Abstract*—Video region-of-interest (ROI) scalable coding scheme can ensure the priority of ROI. Error protection schemes can be used to guarantee the correct receipt of ROI stream when transporting ROI scalable video over error prone network. However, we find that the correct receipt of ROI bitstreams cannot ensure the correct decoding of ROI due to the unique issue of the cross error propagation between ROI and background in ROI scalable coding. In this paper, we propose a ROI scalable coding framework based on leaky prediction (LP) for robustly transporting video over error-prone network. Although several LP approaches have been proposed to improve layered coding, they cannot be applied to ROI scalable coding straightforwardly due to the cross error propagation issue. We deploy a leaky factor to weigh the two predictions, one from the constrained motion estimation within the ROI layer of the reference frame, and the other from the unrestricted motion estimation in the overall reference frame. Simulation results show that the proposed scheme enhances the robustness of ROI scalability while maintaining coding efficiency.**

***Index Terms*—leaky prediction, Region-of-interest, scalable coding, error resilience**

I. Introduction

FOR video over the Internet and wireless networks, it is very important to adapt the coder to time-varying networks, thus to obtain good visual quality with the limited available resources [1]. Region-of-interest (ROI) scalability is of great interest in application scenarios, *e.g.* video surveillance and handheld device, where some visual regions are more important or interesting than the other parts of video.

By applying strong protection to ROI packets, it is realistic to assume that ROI slice packets can be always correctly $received$ even over error-prone network, while the background layer can not. However, since the erroneous background layer is probably employed as reference in decoding ROI layer, there is no guarantee the ROI layer can be correctly $decoded$. As shown in Fig. 1, when the coding stream of background layer for $f(n-2)$ is truncated in transmission, error is introduced to the decoded background layer in $f(n-1)$ (marked by shaded area). Since the background layer of $f(n-1)$ might be referenced in decoding both ROI and background layers, the errors spread to the decoded $f(n)$. Note the ROI layer of $f(n)$ also is contaminated by the background layer of $f(n-1)$, which is referred to as *cross error propagation* in this paper. Even if all the following stream can be correctly received, the errors remain in frames $f(n+1), f(n+2)\cdots$ until the next intra-coded frame is received.

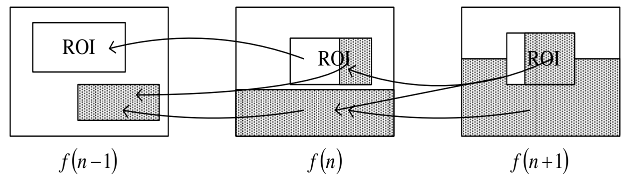


Fig. 1. Cross error propagation in ROI scalable coding (shaded area indicates error, and arrow indicates referring to previous frame)

To refrain background errors propagation, error concealment can be employed to reduce the visible artifacts. However, the error after concealment propagate to successive frames and remain visible for a longer period of time, making the resulting artifacts particularly annoying [2]. Another possible method to avoid background error propagation is to constrain motion estimation (ME) range within the ROI layer of the reference frame, to completely prevent inter-frame dependency between ROI and background layers. In H.263, the independent slice decoding (ISD) mode performs constrained ME [3]. However, because ROI regions usually take up only a small proportion of a frame, a block far from the ROI region can hardly find a good match inside the ROI reference, leading to severe coding inefficiency. Despite the robustness of the constrained ME, ISD is seldom used in ROI coding.

In order to circumvent the cross error propagation between ROI layer and background layers, we propose a novel approach for ROI scalable coding by introducing leakage in the motion compensation (MC) loop in this paper. It deploys a leaky factor to weigh the two predictions, one from the constrained motion estimation within the ROI layer of the reference frame, and the other from the unrestricted motion estimation in the overall reference frame. We analyze the proposed coding scheme, and prove it can reduce rate of the error propagation caused by the background layer destruction due to erroneous transmission. Simulation results confirm the error robustness of the proposed scheme and the maintained coding efficiency.

Leaky factor is a general tool to balance two switched subsystems in many areas. For video coding, the leaky factor has been used to improve Fine Granularity Scalability(FGS) [4] [5] and layered video coding [6] [7]. However, the previous leaky approaches are not applicable for ROI coding. Usually in scalable and layered coding, base layer is supposed to be successfully decoded at the decoder, making it error free for the prediction from the base layer prediction source. While in ROI scalable coding, due to the unique issue of the cross error

Part of this paper was published in IEEE SiPS2007.

Qian Chen, Xiaokang Yang, Li Song and Wenjun Zhang are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, PRC, 200240 e-mail: {qianchen , song_li, xkyang, zhangwenjun}@sjtu,edu.cn

propagation, ROI layer cannot be simply analogous to base layer. To our best knowledge, no prior work in the literature has studied these issues related to ROI. Thus, the novelties of our paper are two folds: 1) We addressed the cross error propagation between ROI layer and background layers; 2) We developed robust video ROI coding system based on LP.

In Section II, we present the basic idea of the LP-based ROI scalable coding along with its framework. We prove that LP-based ROI alleviates the error propagation caused by background loss compared to the conventional framework. Section III gives the simulation results of the proposed scheme. It demonstrates that LP-based ROI coding enhances the robustness of ROI scalable coding while preserving coding efficiency. The effect of leaky factor on the proposed scheme is also analyzed in this section. Section IV concludes the paper.

II. Leaky Prediction Based ROI Scalable Coding

A. Basic Ideas to Build the Leaky Prediction Based ROI Scalability

We deploy a leaky factor α ($0 \leq \alpha \leq 1$) to dampen the effect of the background error that spreads to the following frames. Each ROI MB has two prediction sources : 1) a globally optimal prediction obtained by unrestricted ME in the overall reference frame; 2) a prediction obtained by confining ME within the ROI reference. The two predictions are scaled by gain factors α and $1-\alpha$ respectively to generate a mixed prediction for the coding MB. For each ROI MB in $f(n)$, its prediction $f_{MB}^{(p)}(n)$ is formulated as:

$$f_{MB}^{(p)}(n) = \alpha MC_1(f^{(r)}) + (1-\alpha) MC_2(f_{ROI}^{(r)}) \quad (1)$$

where $f^{(r)}$ is the MB prediction from the overall frame reference and $f_{ROI}^{(r)}$ is the MB prediction from the ROI layer reference. $MC(\cdot)$ denotes motion compensation. Note that MC_1 and MC_2 deploy two distinct motion vectors, $MV1$ and $MV2$ respectively, and both motion vectors should be sent to the decoder.

If $\alpha = 1$, constrained ME is completely excluded from MC loop. Hence $f_{MB}^{(p)}(n)$ is identical to that of the conventional coder without leaky prediction, which is optimal in error-free case. Such a structure has the best coding efficiency but the worst error robustness, since the ROI decoding may refer to previous background layer, which could be unavailable. When $\alpha = 0$, only the ROI layer is employed as reference. Since the loss rate is supposed to be much lower in ROI than in background, thus by simply employing the prediction from ROI layer, it reduces error at the cost of coding efficiency. If $0 < \alpha < 1$, a mix of the two predictions is obtained, to achieve tradeoff between coding efficiency and error robustness.

B. Leaky Prediction Based ROI Coding Framework

Let $f_{MB}(n)$ denote one MB in $f(n)$, and $f_{MB}^{(p)}(n)$ denote its prediction as in (1). Then the residual signal is

$$e_{MB}(n) = f_{MB}(n) - f_{MB}^{(p)}(n) \quad (2)$$

And the quantized residual is

$$\hat{e}_{MB}(n) = Quant\{e_{MB}(n)\} \quad (3)$$

where $Quant\{\cdot\}$ denotes the quantization operation. Hence the reconstruction of the MB, denoted as $f_{MB}^{(r)}(n)$, is

$$f_{MB}^{(r)}(n) = f_{MB}^{(p)}(n) + \hat{e}_{MB}(n) \quad (4)$$

The coding framework of the proposed scheme is illustrated in Fig. 2. For each decoded MB in ROI, it employs the overall reference (in Frame Buffer) and the ROI layer reference (in ROI Layer Buffer) to get two predictions, and then scales them by gain factors α and $1-\alpha$ respectively to generate a weighted sum of the two predictions. $e(n)$ is the corresponding predictive residual, with $\hat{e}(n)$ its quantized version. $MV1$ is the motion vector generated in the unrestricted ME in the overall reference, and $MV2$ is the motion vector from the constrained ME within the ROI layer reference. Both $\hat{e}(n)$ and two sets of MV should be sent to the decoder. For each background MB, its coding process is simplified to that in conventional coding structure, where α equals to 1 in (1).

At the decoder end, the received $MV1$ locates the prediction from the overall reference, and $MV2$ locates the prediction from ROI layer reference. The two predictions are attenuated by α and $1-\alpha$ identical to that at the encoder, and summed to form the mixed prediction. Finally we compensate the prediction by $\hat{e}^{(dec)}(n)$ as in (4) to generate the reconstructed MB. Note that if no $MV2$ is sent for the current MB, the decoder does not need to get the prediction from the ROI layer reference.

C. Error Resilience Performance

We compute the error in the reconstructed ROI layer at decoder when background error occurs. To verify the enhanced error resilience of the proposed scheme, the error propagation performance in two schemes are compared, the conventional ROI coding and the proposed LP-based ROI coding. We consider a video-over-network system as follows:

(1) The leakage introduced by spatial filtering in a motion-compensated prediction [2] is not taken into account. Only the leaky effect from the proposed scheme is analyzed.

(2) The background layer of $f(n-1)$ is damaged in transmission, and Δ is the error. Considering (1), the error of background layer would remain Δ in the following frames.

(3) For a given ROI MB, let p denote the probability that it can find prediction in the background layer of the reference, and $1-p$ the probability that it refers to only the ROI layer. Note that the meaning of p differs in two schemes: 1) In conventional coding, where only the unrestricted ME is conducted, the proportion of p ROI layer MB get prediction from background layer reference. 2) In the proposed scheme, the proportion of p ROI layer MB get mixed predictions from both background and ROI layers of the reference.

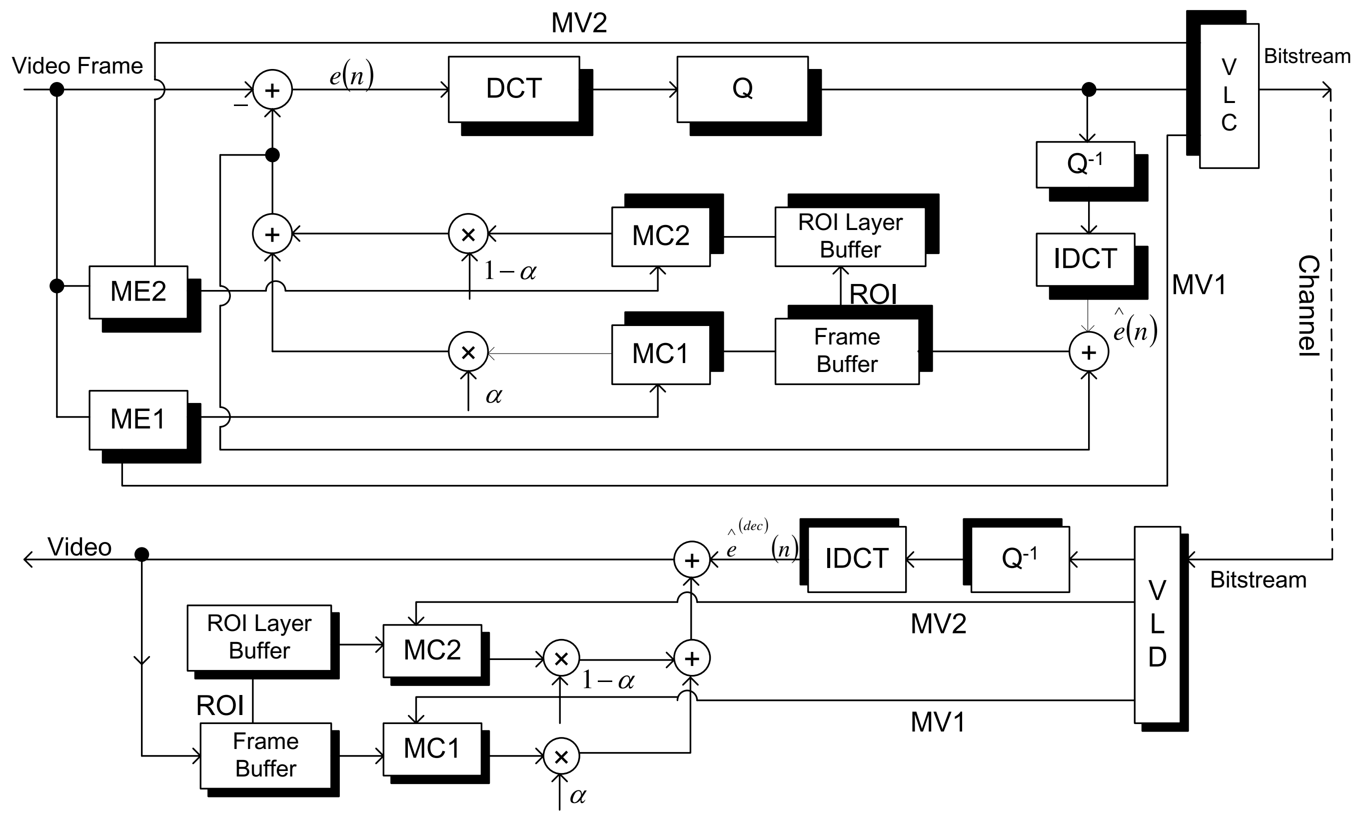


Fig. 2. Coding framework of the leaky prediction based ROI coding scheme.

We first derive the error propagation in the conventional coding scheme.

$$\begin{aligned} f^{(p,ROI)}(n) &= pMC(f^{(r,BG)}(n-1)+\Delta) \\ &\quad +(1-p)MC(f^{(r,ROI)}(n-1)) \\ &= pMC(f^{(r,BG)}(n-1)) \\ &\quad +(1-p)MC(f^{(r,ROI)}(n-1))+p\Delta \end{aligned} \tag{5}$$

For clarity,

$$I(n) = f^{(p,ROI)}(n) \tag{6}$$

$$O(n) = MC(f^{(r,BG)}(n-1)) \tag{7}$$

We rewrite (5) in recursive form of $I(n)$ as

$$I(n) = pO(n) + (1-p)I(n-1) + e(n) \tag{8}$$

where $e(n)$ is the ROI layer error in $f(n)$, here $e(n) = p\Delta$.

For simplicity of formulation, we assume no more background error occurs in the following frames [5]. The error that propagates from background in $f(n)$ to ROI in $f(n+1)$ would be

$$e(n+1) \quad = p\Delta + (1-p)p\Delta$$

Therefore

$$\begin{aligned} e(n+t) &= p\Delta + (1-p)p\Delta + \cdots \\ &\quad +(1-p)^t p\Delta \\ &= \Delta[1-(1-p)^{(t+1)}] \end{aligned} \tag{9}$$

Suppose p is constant, and then the background error spreads across the following ROI layers in a fixed way defined in (9). After infinite time steps, the error propagating to the ROI layer equals to Δ, the error in the background layer.

Likewise, we compute the error propagation in the LP-based ROI coding scheme. The prediction of ROI layer in $f(n)$ is

$$\begin{aligned} f^{(p,ROI)}(n) &= p\{\alpha MC_1(f^{(r,BG)}(n-1)+\Delta) \\ &\quad +(1-\alpha)MC_2(f^{(r,ROI)}(n-1))\} \\ &\quad +(1-p)MC_2(f^{(r,ROI)}(n-1) \end{aligned} \tag{10}$$

We rewrite (10) as

$$I(n) = p(\alpha O(n) + (1-\alpha)I(n-1)) + (1-p)I(n-1) + e(n) \tag{11}$$

where

$$I(n) = f^{(p,ROI)}(n) \tag{12}$$

$$O(n) = MC_1(f^{(r,BG)}(n-1)) \tag{13}$$

and $e(n)$ is the ROI layer error in $f(n)$, here $e(n) = p\alpha\Delta$.

The error that propagates from background in $f(n)$ to ROI in $f(n+1)$ would be

$$\begin{aligned} e(n+1) &= p[\alpha\Delta + (1-\alpha)e(n)] + (1-p)e(n) \\ &= p\alpha\Delta + (1-p\alpha)e(n) \\ &= \beta\Delta + (1-\beta)e(n) \end{aligned} \tag{14}$$

Clearly, (14) is an iterative expression of the ROI layer error. Consequently, for $f(n+t)$

$$\begin{aligned} e(n+t) &= \beta\Delta + (1-\beta)e(n+t-1) \\ &= \beta\Delta + (1-\beta)(\beta\Delta + \cdots (\beta\Delta + (1-\beta)e(n))) \\ &= \Delta[1-(1-\beta)^{(t+1)}] \\ &= \Delta[1-(1-p\alpha)^{(t+1)}] \end{aligned} \quad (15)$$

Comparing (15) with (9), it can be easily seen that LP-based leaky prediction introduces a gain factor α to flexibly control the cross error propagation rate. The smaller of α, the slower the background error would propagate to ROI.

III. Simulation Results and Analysis

Extensive experiments have been carried out to validate the robustness and efficiency of the proposed ROI scalable coding scheme. The proposed scheme was developed by modifying H.264/AVC software JM9.8 [8]. First, packet loss simulation is conducted to show the enhanced error resilience of the proposed scheme, and the rate-distortion (RD) performance is also given to illustrate its efficiency. Finally, we investigate the effect of leaky factor on the performance of the scheme. In simulation, ROI is defined by the method in [9]. And γ is the size ratio of ROI area to its overall frame.

A. Error Resilience Performance

First, we design a simple simulation to verify the error resilience analysis in Section II-C, where background layer of only the first P frame is discarded during transmission. We code the test sequence *foreman* (QCIF, 12.5fps, 100 frames, QP=28, IPPP mode, $\gamma = 36\%$, bit rate = 79.26kbps). We replace the corrupted image content by corresponding pixels from previous frame as a simple approach for error concealment, which yields good results for sequences with little motion [10].

As can be observed in Fig. 3(a) and Fig. 3(b), the PSNR of both ROI layer and the whole frame recovers from error at a faster rate in the proposed framework than in JM9.8. The proposed scheme provides better ROI quality, whereas preserving the background quality similar to that in the conventional coder. Hence, it upgrades the quality of the overall picture.

However, in true applications, continuous frame errors are more likely to occur than single frame error. To better evaluate the proposed scheme, we conduct two groups of simulation with higher bitstream loss rate, namely periodical frame error and continuous frame error.We use two test sequences: 200 frames QCIF *foreman*, coded at 12.5fps, and average ROI ratio $\gamma = 35\%$; 160 frames CIF *stefan*, coded at 12.5fps, and average ROI ratio $\gamma = 17\%$.

In the simulation group of periodical frame error, background slices of 2 consecutive frames in *foreman* are completely lost every 50 frames, while background slices of 2 consecutive frames in *stefan* are lost every 40 frames. A PSNR gain up to 2dB can be observed in the ROI layer with the proposed scheme in *foreman*in Fig.4(a). And the error decay performance of the ROI layer obviously precedes that without leaky in Fig.5(a). Meanwhile, we notice better error recovery curves of the overall frame in both two sequences with the proposed framework .

In the simulation group of continuous frame error, we set loss rate of background slice in both sequences to 5% per frame. The results are shown in Fig. 6 and Fig. 7. Since errors appear in successive frames, it does not allow for enough time to show the evident error decay process after each erroneous frame. However, better PSNR recovery curve in both ROI layer and overall frame can still be observed with the proposed scheme.

Based on the simulations above, the enhanced error robustness of the proposed LP-based ROI coding scheme has been well validated.The errors dampen with time at a much faster rate in both ROI layer and overall frame with the proposed framework than the conventional scheme, especially in slow motion sequence *foreman*. This is because in slow motion sequence, it is easier for the ROI layer MB to find good prediction in previous ROI reference, and better coding efficiency can be achieved accordingly.

B. Coding Efficiency

Leaky prediction is a well-known technique to enhance error robustness at the cost of coding efficiency. And it would inevitably introduce bit redundancy in *error-free* case. We now investigate the coding efficiency of the proposed scheme.

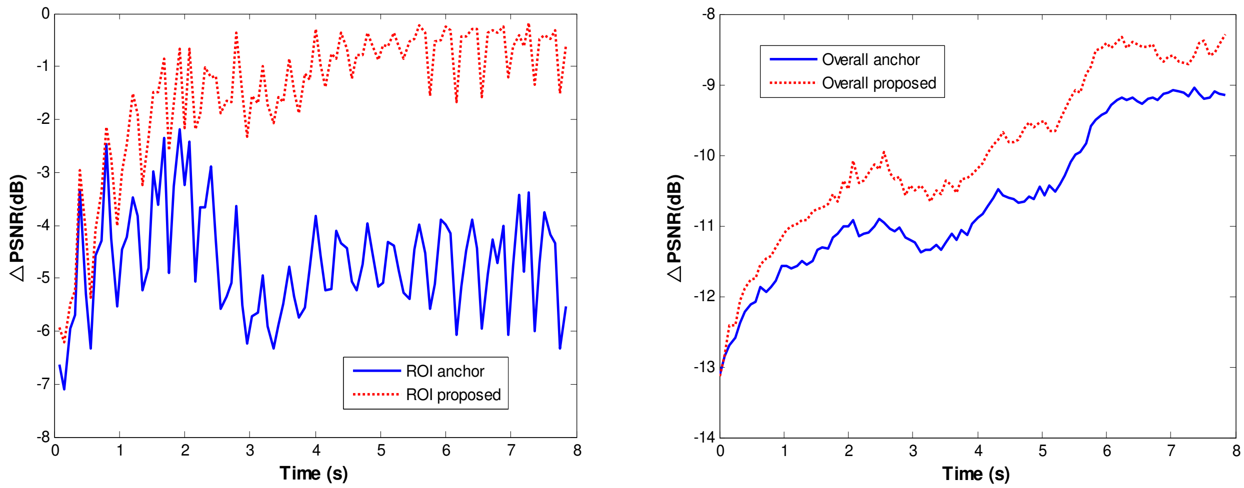
Fig. 8 shows the rate-distortion (RD) performance of the proposed scheme and the conventional coding in error-free case. On average, 1dB loss of PSNR is observed in slow motion sequence *foreman* in Fig.10(a) with LP-based ROI coding, though the PSNR gap narrows with bit rate increase. While in *stefan*, the typical sequence that features swift shift of ROI position and intense global motion, 2dB PSNR loss is observed in RD curve in Fig.10(b).

Fig. 9 shows RD performance of the two schemes in error case, or concretely, periodical error specified in Section III-A. Clearly, over a wide range of bit rate, the PSNR in both the overall picture and the ROI layer are of much higher magnitude employing our scheme. For slow motion sequence *foreman*, 6-13dB PSNR gain can be observed with the proposed scheme in Fig. 9(a), while 2-6dB PSNR gain can also be achieved in fast motion sequence *stefan*. Considering the significantly enhanced robustness in error case, the coding efficiency loss of the proposed scheme in error-free case might be tolerable.

C. Comparison with FEC

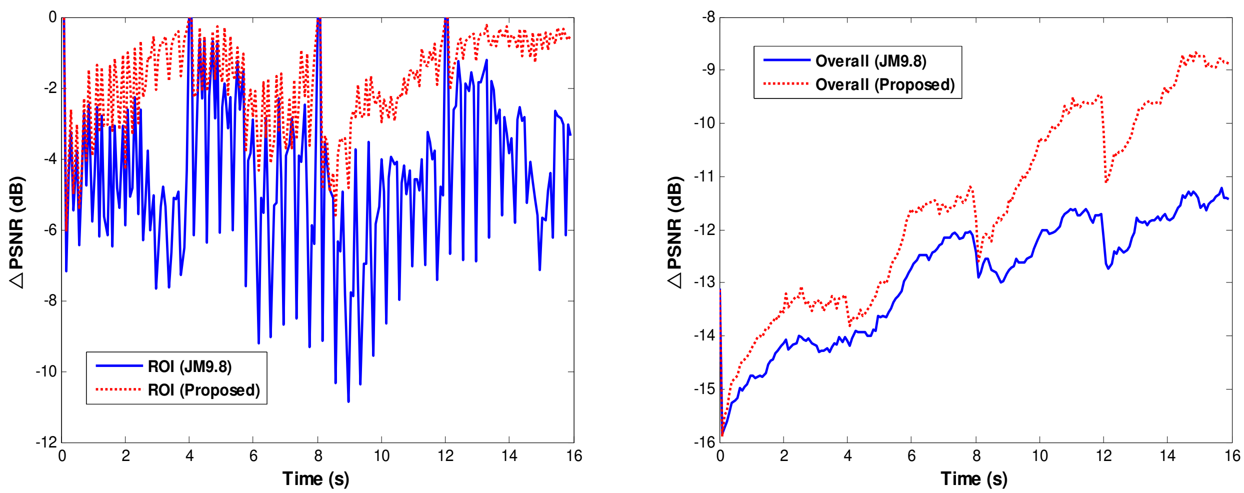
From the source-channel coding point of view, it is relevant to compare the separate source and channel coding (i.e. H.264 compression followed by error-correction codes) to the use of the joint source-channel coding approach (leaky prediction) proposed in this work. In this paper, we compare our scheme with FEC coding, typical interlaced RS(n,k) encoding. Basically, this scheme operates by aligning D successive data packets vertically, each of which contains k data and $n-k$ parity codes.

In our experiment, we set $n = 128$ for QCIF and $n = 255$ for CIF sequence, while k is adaptively adjusted to make sure



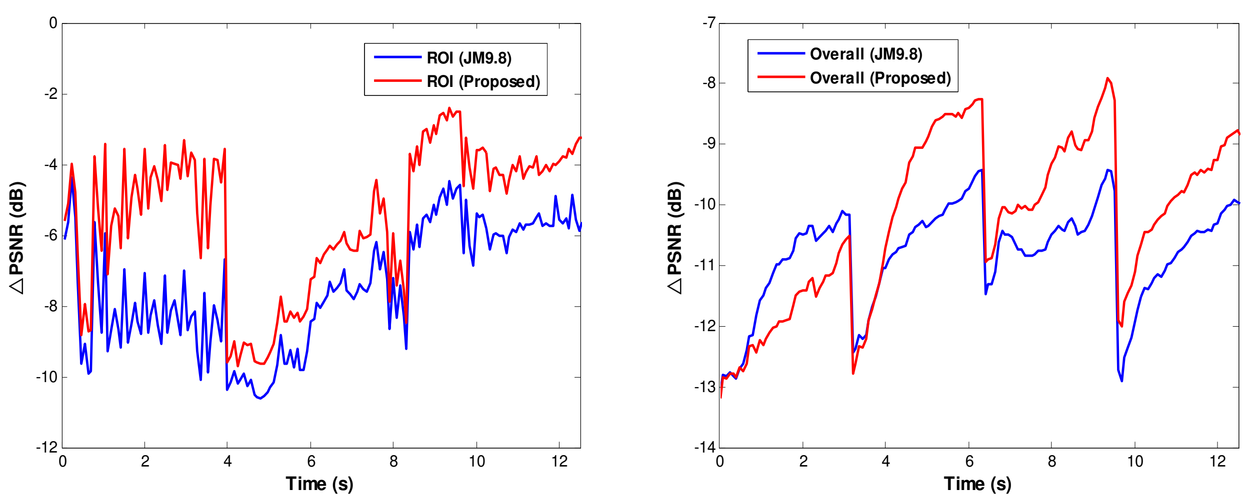
(a) (b)

Fig. 3. PSNR drop comparison in *foreman* QCIF, only the background slice of the first P frame is lost. For the proposed framework, $\alpha = 0.9$. (a) The ROI layer. (b) The overall frame.



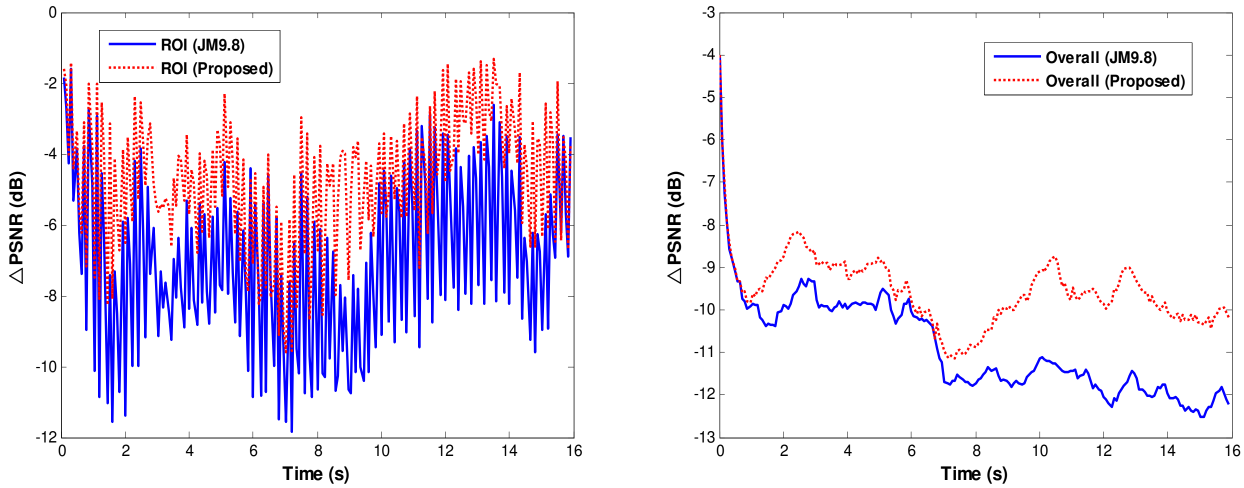
(a) (b)

Fig. 4. PSNR drop comparison in *foreman* QCIF in periodical frame error. For the proposed framework, $\alpha = 0.9$ (a) The ROI layer. (b) The overall frame.



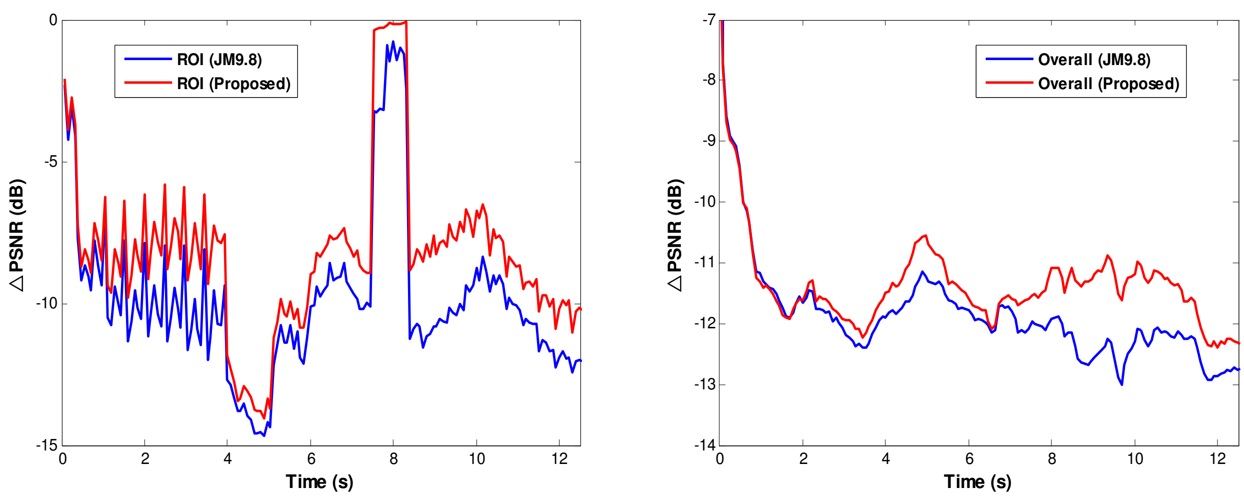
(a) (b)

Fig. 5. PSNR drop comparison in *stefan* CIF in periodical frame error. For the proposed framework, $\alpha = 0.9$. (a) The ROI layer. (b) The overall frame.



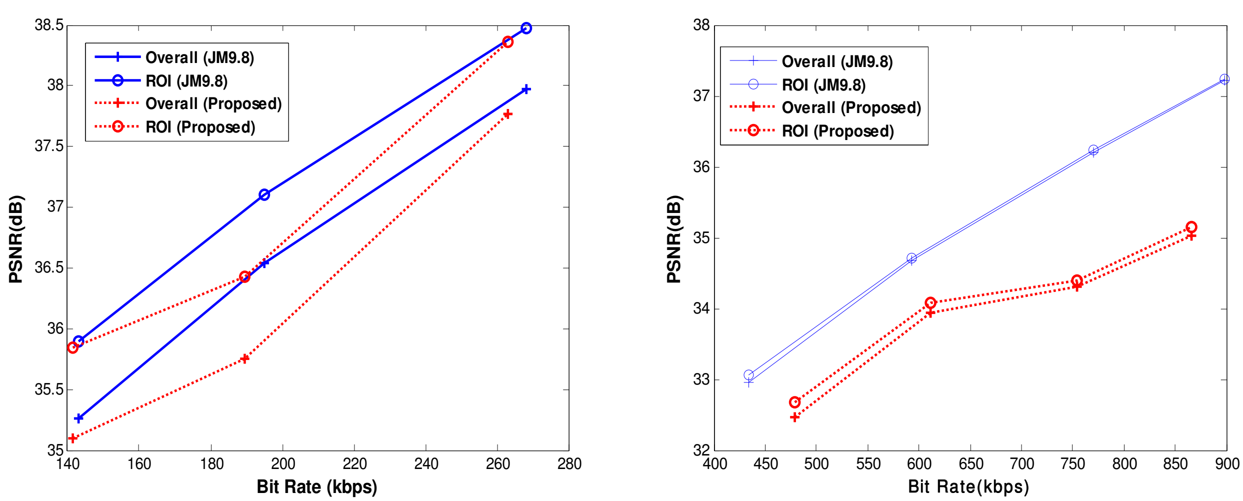
(a) (b)

Fig. 6. PSNR drop comparison in *foreman* QCIF in continuous frame error. For the proposed framework, $\alpha = 0.9$. (a) The ROI layer. (b) The overall frame.



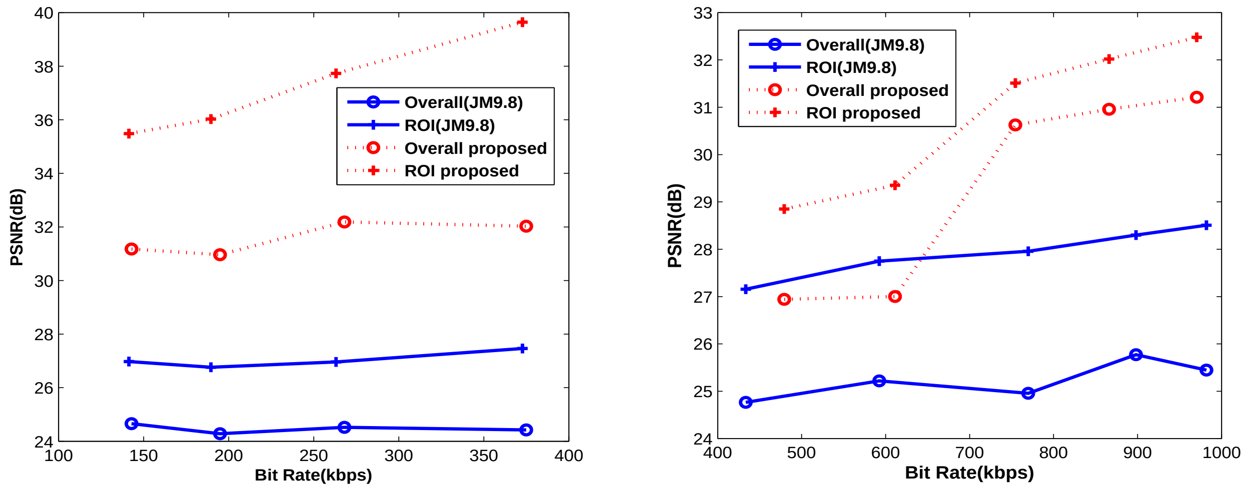
(a) (b)

Fig. 7. PSNR drop comparison in *stefan* CIF in continuous frame error. For the proposed framework, $\alpha = 0.9$. (a) The ROI layer. (b) The overall frame.



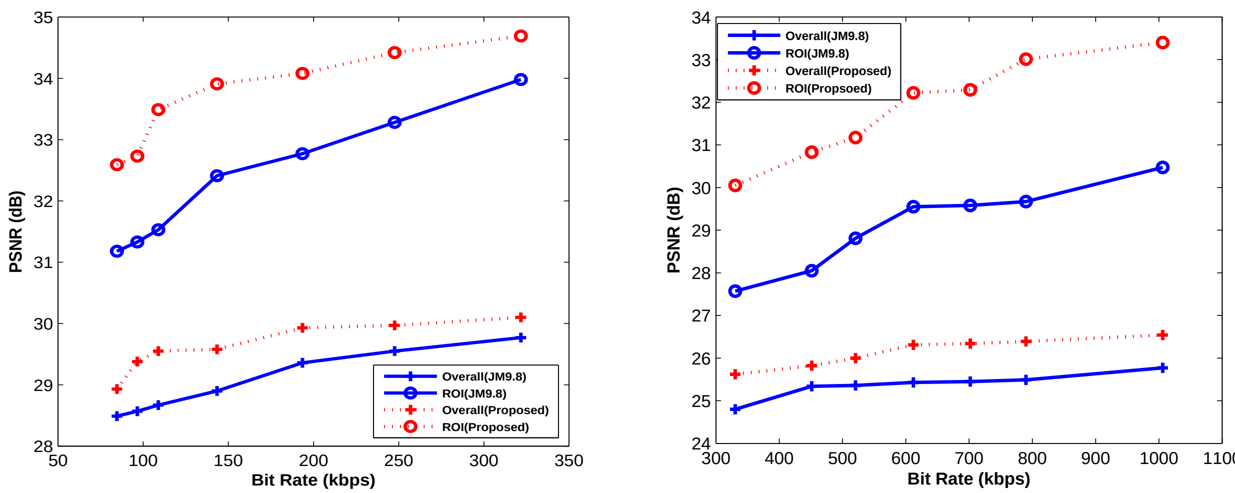
(a) (b)

Fig. 8. The comparison of rate-distortion performance in error-free case between the proposed framework ($\alpha = 0.9$) and the conventional coding. (a) *foreman* QCIF. (b) *stefan* CIF.



(a) (b)

Fig. 9. The comparison of rate-distortion in periodical error between the proposed framework ($\alpha = 0.9$) and the conventional coding. (a) RD performance in *foreman* QCIF. (b) RD performance in *stefan* CIF.



(a) (b)

Fig. 10. The comparison of rate-distortion performance between the proposed framework ($\alpha = 0.9$) and FEC. (a) RD performance in *foreman* QCIF. (b) RD performance in *stefan* CIF.

RS encoding has the equal coding bitrate with the proposed leaky prediction, and observe the PSNR performance of the two schemes given the same packet loss. Figure. 10 shows RD performance of the two schemes. In both *foreman* and *stefan* sequences, the overall PSNR gets an average $0.5dB$ gain, while ROI PSNR gets $1 - 2dB$ gain in the proposed leaky prediction.

D. Influence of Leaky Factor

Leaky factor serves as a parameter to trade off coding efficiency and robustness. In this section, we analyze the effect of leaky factor on the rate-distortion performance.

Table I gives the trends of PSNR and bit rate when α decreases in *foreman* sequence (all the other settings are the same as that mentioned in subsection III-A. Basically, the drop of α leads to bit rate increase. We also notice that PSNR does not always increase with the decrease of α. When α is about 0.4, we get the PSNR peak, but PSNR drops when α falls below 0.4. We found similar phenomenon in other sequences. Normally, a good trade-off between the ROI robustness and

TABLE I
PSNR AND BIT RATE VARYING WITH LEAKY FACTOR α IN *foreman* QCIF IN PERIODICAL ERROR CASE. THE PERCENT OF BIT RATE INCREASE BASES ON TRADITIONAL CODING.

α	PSNR(dB)	bit rate	bit rate increase(%)
1.0	31.1488	73.41	0
0.9	34.2501	83.77	14.11
0.8	34.8443	93.03	26.73
0.7	35.0765	101.62	38.43
0.6	35.2197	109.18	48.73
0.5	35.2352	116.48	48.73
0.4	35.3365	123.15	67.76
0.3	35.3187	129.95	77.02
0.2	35.3149	135.93	85.17
0.1	35.2940	141.66	92.97
0	35.1820	147.15	100.45

coding efficiency can be made in α range of $0.5 - 0.9$.

IV. Conclusion

We have presented a novel robust ROI scalable video coding scheme based on leaky prediction to circumvent the cross error propagation problem. The leaky factor accelerates the error decay process. Compared with the conventional ROI coding without leaky prediction, the proposed scheme makes a better tradeoff between error robustness and coding efficiency. Therefore, the proposed scheme can be used for robust ROI transmission in video applications, such as visual surveillance and handheld devices, so as to guarantee the quality of visually important regions.

Acknowledgment

The author would like to thank Prof. Xiaolin Wu and the anonymous reviewers for all the valuable and insightful comments that helped to improve the quality of the paper.

References

[1] B. Girod, M. Kalman, Y. Liang and R. Zhang, ”Advances in channel adaptive video streaming,” Proceedings IEEE International Conference on Image Processing, vol.1, pp.9 -12, Rochester, New York, USA, Sept. 2002.

[2] B. Girod and N. Färber, ”Feedback-based error control for mobile video transmission,”*Proceedings of IEEE*, vol.87, no.10, pp.1707-1723, Oct.1999.

[3] *ITU, Video coding for low bitrate comunication*, ITU-T Recommendation H.263,Version 2, Jan.1998.

[4] S. Han and B. Girod, ”Robust and efficient scalable video coding with leaky prediction,” IEEE International Conference on Image Processing, vol.2, pp.41-44, Rochester, New York, USA, Sept.2002.

[5] 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, Jun.2002.

[6] M. Ghanbari, V. Seferidis, ”Efficient H.261-based two-layer video codecs for ATM networks,” IEEE Transactions on Circuits Systems for Video Technology, vol. 5, no. 2, pp. 171-175, Apr.1995.

[7] Y. Liu, P. Salama, Z. Li and E. J. Delp, ”An enhancement of leaky prediction layered video coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol.15, no.11, pp.1317-1331, Nov.2005.

[8] JVT, H.264/AVC reference software JM9.8, online available at http://ftp3.itu.ch/av-arch/jvt-site/reference_software.

[9] G. Zhai, Q. Chen, X. Yang and W. Zhang, ”Scalable Visual Significance Profile Estimation,” IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, US, April, 2008, accepted.

[10] C. Chen, ”Error detection and concealment with an unsupervised MPEG2 video decoder,” Journal of Visual Communication and Image Representation, vol.6, no.3, pp.265-278, Sept. 1995.