Transactions Letters

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

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

***Abstract*— A 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 the ROI stream when transporting ROI scalable video over an 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 letter, we propose an ROI scalable coding framework based on leaky prediction (LP) for robustly transporting video over an 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 (ME) within the ROI layer of the reference frame, and the other from the unrestricted ME in the overall reference frame. Simulation results show that the proposed scheme enhances the robustness of ROI scalability while maintaining coding efficiency.**

***Index Terms*—Error resilience, leaky prediction (LP), region-of-interest (ROI), scalable coding.**

I. INTRODUCTION

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* over an error-prone network, while the background layer cannot. However, since the erroneous background layer is probably employed as reference in decoding ROI layer, there is no guarantee that the ROI layer can be correctly *decoded*. As shown in Fig. 1, when the coding stream of background layer for $f(n-1)$ 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 that 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 letter. Even if all the following stream can be correctly received, the errors remain in frames $f(n+1)$, $f(n+2)\ldots$ until the next intra-coded frame is received.

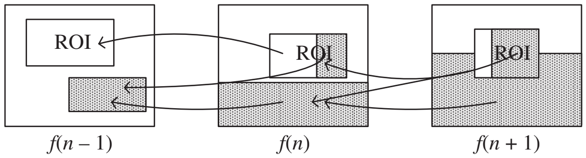


Fig. 1. Cross error propagation in ROI scalable coding (shaded area indicates error, and arrow indicates referring to previous frame).

To restrain background error propagation, error concealment can be employed to reduce the visible artefacts. However, the error after concealment propagate to successive frames and remain visible for a long period of time, making the resulting artefacts particularly annoying [1]. Another possible method to avoid background error propagation is to restrict the motion estimation (ME) range within the ROI layer of the reference frame, to completely prevent inter-frame dependency between ROI and background layers. 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, independent slice decoding 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 letter. It deploys a leaky factor to weigh the two predictions, one from the constrained ME within the ROI layer of the reference frame, and the other from the unrestricted ME in the overall reference frame.

For video coding, the leaky factor has been used to improve fine granularity scalability [3], [4] and layered video coding

---

Manuscript received September 11, 2007, revised March 11, 2008 and September 24, 2008. First version published May 12, 2009; current version published September 16, 2009. This work was in part supported by NSFC, Nos. 60502034, 60625103, 60828001, and 60702044, Hi-Tech Research and Development Program of China, No. 2006AA01Z124, NCET-06-0409, and the 111 Project. This paper was recommended by Associate Editor D. S. Turaga.

Q. Chen was with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China 200240. He is currently with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: qiantrue@ufl.edu).

X. Yang, L. Song, and W. Zhang are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China 200240 (e-mail: xkyang@sjtu.edu.cn; songli@sjtu.edu.cn; zhangwenjun@sjtu.edu.cn).

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

Digital Object Identifier 10.1109/TCSVT.2009.2022703

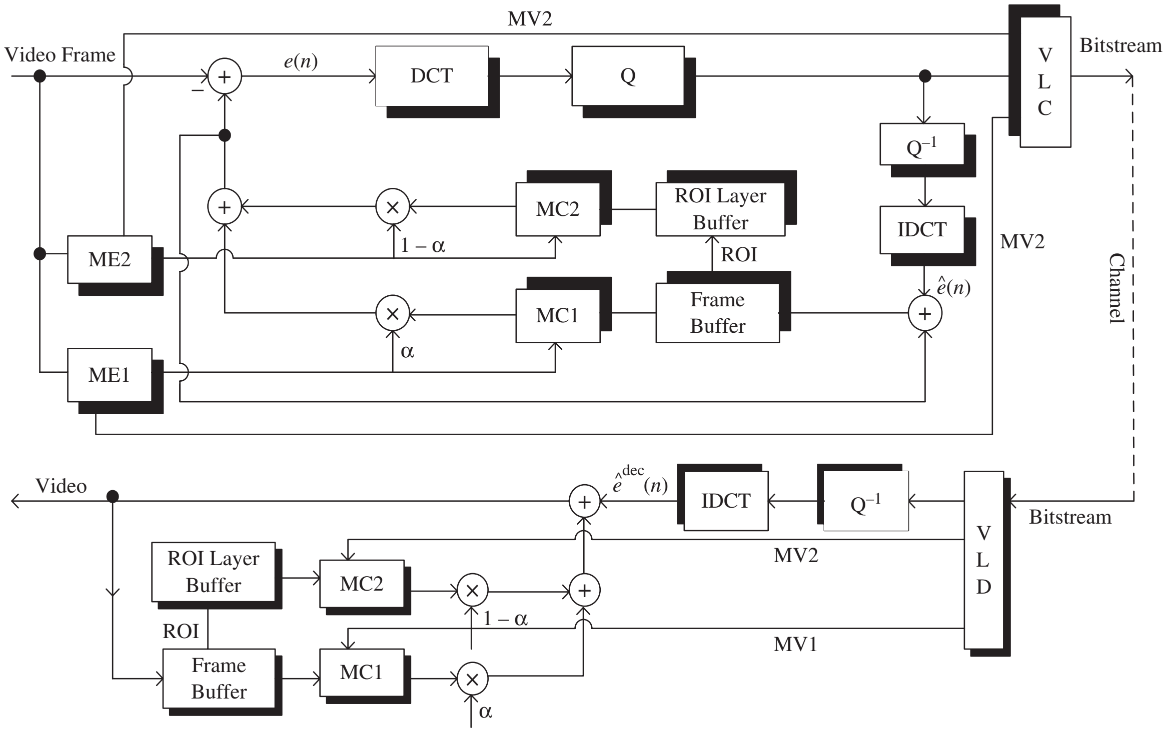


Fig. 2. Coding framework of the LP-based ROI coding scheme.

[5]. However, the previous leaky approaches are not applicable for ROI coding. Usually, in scalable and layered coding, the 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 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 letter are two fold: 1) we addressed the cross error propagation between ROI layer and background layers and 2) we developed robust video ROI coding system based on leaky prediction (LP).

II. LP-BASED ROI SCALABLE CODING

A. Basic Ideas to Build the LP-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 and 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

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

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 the MC loop. Hence $f_{MB}^{(p)}(n)$ is identical to that of the conventional coder without leaky prediction, which is optimal in the 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, by simply employing the prediction from ROI layer it reduces the 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. LP-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

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

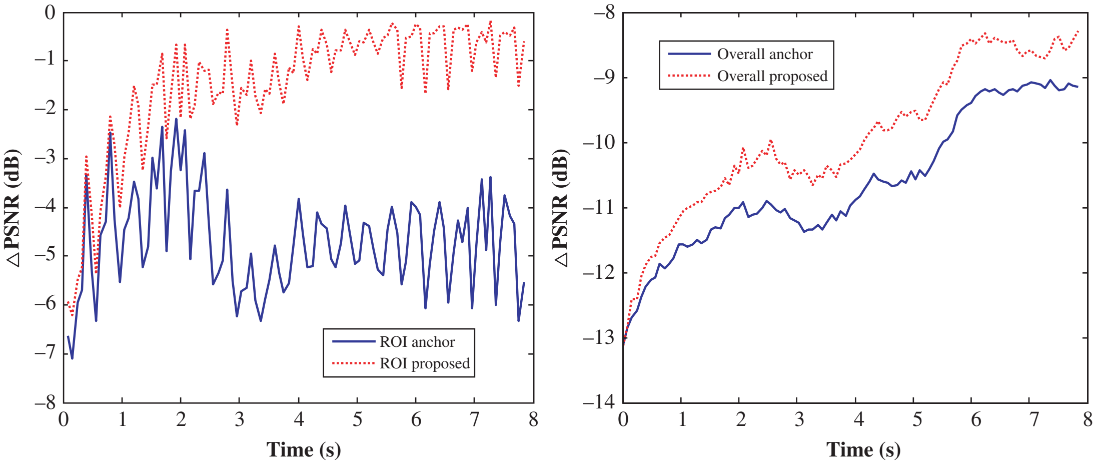
And the quantized residual is

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

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

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

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



(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) ROI layer and (b) overall frame.

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

To verify the enhanced error resilience of the proposed scheme, the error propagation performance in the conventional ROI coding and the proposed LP-based ROI coding are computed. We consider a video-over-network system as follows.

1) The leakage introduced by spatial filtering in a motion-compensated prediction [1] is not taken into account.
2) The background layer of $f(n-1)$ is damaged in transmission, and Δ is the error. Considering 1), the error would remain Δ in the following frames.
3) For a given ROI MB, let p denote the probability it can find prediction in the background layer of the reference, and $1-p$ the probability it refers to only the ROI layer. Note that the meaning of p differs in two schemes: a) in conventional coding, the proportion of p ROI layer MB get prediction from background layer reference; b) in the proposed scheme, the proportion of p ROI layer MB get mixed predictions from both background and ROI layers of the reference.

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 $e(n) = p\Delta$.

We assume no more background error occurs in the following frames [4]. The error that propagates from background in $f(n)$ to ROI in $f(n+1)$ would be

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

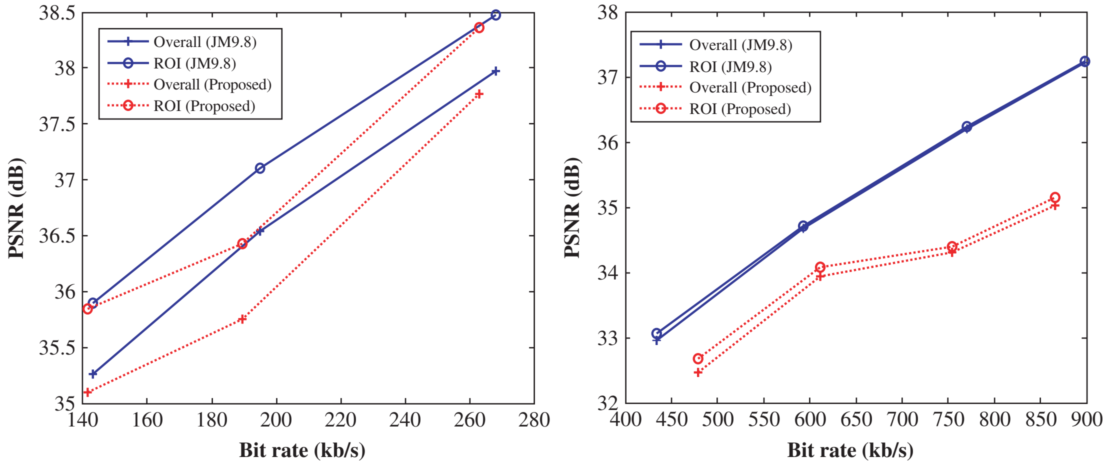
Therefore

$$e(n+t) = \Delta[1-(1-p)^{(t+1)}]. \tag{9}$$

Suppose p is constant; 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 Δ.

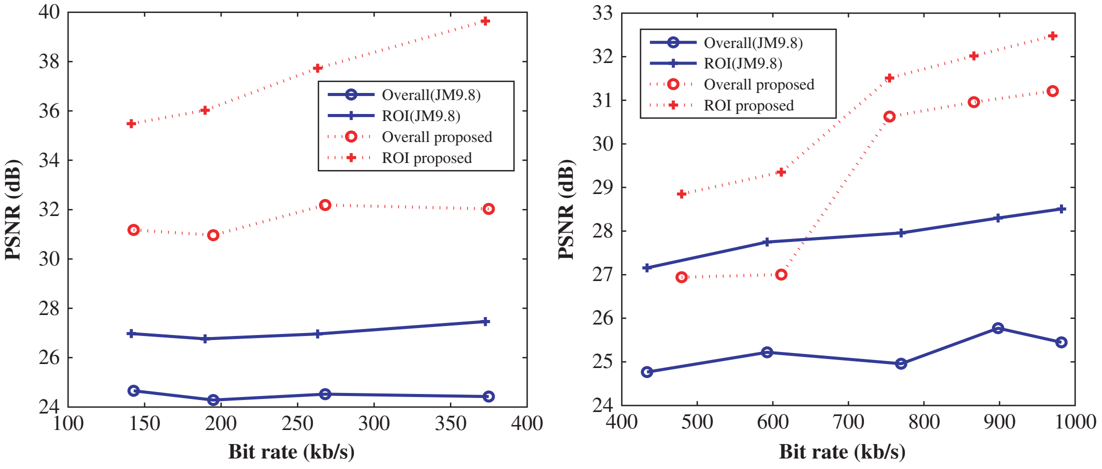
Likewise, we compute the error propagation in the LP-based ROI coding scheme. The prediction of ROI layer in $f(n)$ is

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



(a) (b)

Fig. 4. Comparison of RD performance in error-free case between the proposed framework ($\alpha = 0.9$) and the conventional coding. (a) *Foreman* QCIF and (b) *Stefan* CIF.



(a) (b)

Fig. 5. Comparison of RD in periodical error between the proposed framework ($\alpha = 0.9$) and the conventional coding. (a) RD performance in *Foreman* QCIF and (b) RD performance in *Stefan* CIF.

We rewrite (10) as

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

$$O(n) = MC_1(f^{(r,BG)}(n-1)) \tag{12}$$

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

Consequently, for $f(n+t)$

$$e(n+t) = \Delta[1-(1-p\alpha)^{(t+1)}]. \tag{13}$$

Comparing (13) 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 the α the slower the background error would propagate to ROI.

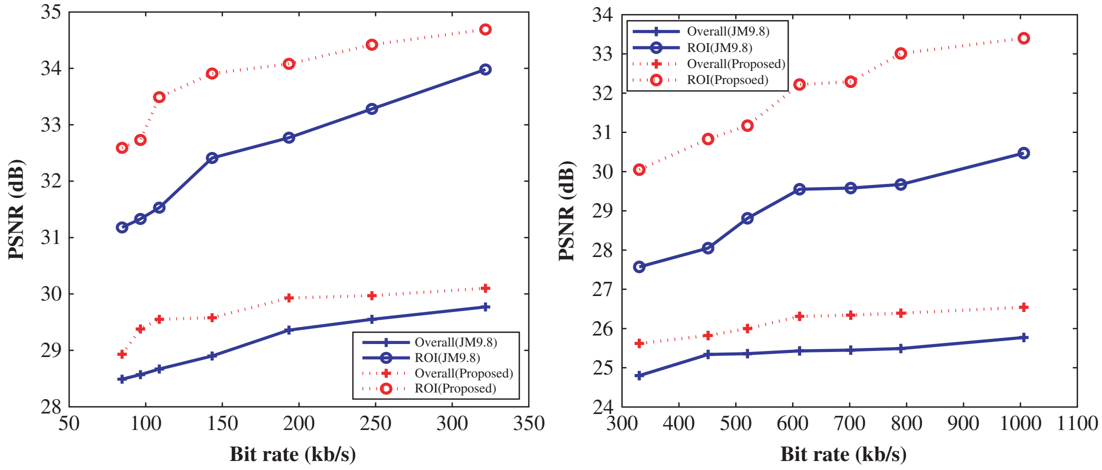
III. SIMULATION RESULTS AND ANALYSIS

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 [6]. In simulation, ROI is defined by the method in [7], and γ is the size ratio of ROI area to its overall frame.

A. Error Resilience Performance

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, 12.5 frames/s, 100 frames, QP = 28, IPPP mode, γ = 36%, bitrate = 79.26 kb/s). 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 [8].

As can be observed in Fig. 3(a) and (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



(a) (b)

Fig. 6. Comparison of RD performance between the proposed framework ($\alpha = 0.9$) and FEC. (a) RD performance in *Foreman* QCIF and (b) RD performance in *Stefan* CIF.

scheme provides better ROI quality while preserving the background quality similar to that in the conventional coder. Hence, it upgrades the quality of the overall picture.

B. Coding Efficiency

Leaky prediction would inevitably introduce bit redundancy in *error-free* case. We now investigate the coding efficiency of the proposed scheme.

Fig. 4 shows the rate-distortion (RD) performance of the proposed scheme and the conventional coding in error-free case. On average, 1 dB loss of PSNR is observed in slow motion sequence *Foreman* in Fig. 4(a) with LP-based ROI coding, though the PSNR gap narrows with bit rate increase. While in *Stefan*, the sequence features swift shift of ROI position and intense global motion, 2 dB PSNR loss is observed in RD curve in Fig. 4(b).

Fig. 5 shows RD performance of the two schemes in the error case. Clearly, the PSNR in both the overall picture and the ROI layer are of much higher magnitude by employing our scheme. For slow motion sequence *Foreman*, 6–13 dB PSNR gain can be observed, while 2–6 dB PSNR gain can be achieved in fast motion sequence *Stefan*.

C. Comparison With Forward Error Correction (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 (LP) proposed in this letter. In this letter, we compare our scheme with FEC coding, which is a 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 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. Fig. 6 shows RD performance of the two schemes. In both *Foreman* and *Stefan* sequences, the overall PSNR gets an average 0.5 dB gain, while ROI PSNR gets 1–2 dB gain in the proposed leaky prediction.

TABLE I

PSNR AND BIT RATE VARYING WITH LEAKY FACTOR α IN FOREMAN QCIF IN PERIODICAL ERROR CASE. THE PERCENT OF BIT RATE INCREASE BASED 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

D. Influence of Leaky Factor

Leaky factor serves as a parameter to trade off coding efficiency and robustness.

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 Section 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 a similar phenomenon in other sequences. Normally, a good tradeoff between the ROI robustness and 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.

REFERENCES

[1] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 97, no. 10, pp. 1707–1723, Oct. 1999.

[2] *ITU-T, Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263, Version 2, Jan. 1998.

[3] S. Han and B. Girod, "Robust and efficient scalable video coding with leaky prediction," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2. Rochester, NY, Sep. 2002, pp. 41–44.

[4] H. Huang, C. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 372–385, Jun. 2002.

[5] Y. Liu, P. Salama, Z. Li, and E. J. Delp, "An enhancement of leaky prediction layered video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 11, pp. 1317–1331, Nov. 2005.

[6] *JVT, H.264/AVC Reference Software JM9.8.* Available at http://wftp3.itu.int/av-arch/jvt-site/reference/software

[7] G. Zhai, Q. Chen, X. Yang, and W. Zhang, "Scalable visual significance profile estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 268–271.

[8] C. T. Chen, "Error detection and concealment with an unsupervised MPEG2 video decoder," *J. Visual Commun. Image Representation*, vol. 6, no. 3, pp. 265–279, Sep. 1995.