Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Unequal Forced-Intra-Refresh for robust video streaming*

LIU Hao†, ZHANG Wen-jun, YANG Xiao-kang

(*Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China*)

†E-mail: liu_hao@sjtu.edu.cn

Received Nov. 21, 2005; revision accepted Mar. 9, 2006

Abstract: For video streaming over lossy channels, intra refresh can mitigate the error-propagation effect caused by packet losses. Besides some intra-mode macroblocks (MBs) generated by the "Lagrangian rate-distortion" or "Sum of absolute difference" mode decision, the encoder or transcoder possibly needs to increase some "forced" intra-mode MBs for robust video streaming. Based on the error-propagation analysis in a group of pictures (GOP), we propose an unequal Forced-Intra-Refresh (FIR) scheme to improve packet loss resilience of video streaming. According to a GOP-level error-propagation model, the proposed unequal FIR scheme can optimally increase the unequal number of forced intra-mode MBs for different frames in a GOP. Simulation results showed that the proposed scheme can effectively enhance the robustness of video streaming under different channel conditions, and achieve about 0.1~0.9 dB gains over the average FIR scheme in H.264/AVC tools.

Key words: Video streaming, Error propagation, Intra refresh, Packet loss resilience, Group of pictures (GOP)
doi:10.1631/jzus.2006.A1695 **Document code:** A **CLC number:** TN919.81

INTRODUCTION

In the encoding or transcoding of video streaming, intra refresh is introduced as a non-normative error-resilience tool. Although the intra-refresh schemes based on the rate-distortion optimization (Côté *et al.*, 2000; Stockhammer *et al.*, 2002) are superior to the early heuristic intra-refresh strategies, these new schemes cannot yet accurately estimate the video distortion due to time-varying packet losses, and usually ignore the inter-frame error-propagation effect. Chiou *et al.*(2005) proposed a two-pass intra-refresh transcoding scheme for inserting error-resilience features into a pre-encoded video at the media gateway of a three-tier streaming system. Tu and Eckehard (2004) proposed an efficient intra-refresh scheme based on error tracking, which tracks the region corrupted by error propagation with feedback from the receiver. These schemes are very useful for some unicast systems, but may not be applicable to video multicast or broadcast systems, where we have to take into account the heterogeneity of different receivers so as to minimize the average video distortion experienced by each user.

Due to the randomness of time-varying channel errors, the video distortion caused by packet losses is a random variable. In a statistical sense, the random intra-refresh for a particular frame can achieve minimum average distortion (He *et al.*, 2002). In prevalent video coding standards such as the current H.264/AVC, some intra-mode microblocks (MBs) generated by the "Lagrangian rate-distortion" or "Sum of absolute difference" mode decision may be helpful for increasing packet loss resilience. However, such intra-mode MBs originated from the purpose of coding efficiency improvement may be insufficient for combating error propagation in most error-prone environments. To further enhance the robustness of video streaming under different channel conditions, the encoder or transcoder may utilize the Forced-Intra-Refresh (FIR) policy to increase some "forced" intra-mode MBs beyond the conventional mode decision. In current H.264/AVC tools, a straightforward

* Project (No. 60332030) supported by the National Natural Science Foundation of China

FIR scheme is to equally increase the forced intra-mode MBs for each inter-frame in a group of pictures (GOP), which may be non-optimal in terms of the trade-off between coding efficiency and loss resilience.

In this paper, we first analyze and model the effect of error propagation on the quality degradation caused by packet losses. We then propose an unequal FIR scheme to improve packet loss resilience of video streaming, which increases the unequal number of forced intra-mode MBs for different frames in a GOP, and then randomly refreshes the corresponding frame.

ERROR-PROPAGATION ANALYSIS

Video streaming normally needs to introduce the GOP structure so as to allow random access, interactive selection and multimedia synchronization. Since errors in B-frames do not interfere with other frames, we consider the FIR schemes in the typical GOP structure with one IDR-frame followed by some P-frames.

Error-propagation effect in a GOP

For motion-compensated video coding, the errors caused by packet losses not only impair the reconstruction quality of the current frame, but also lead to error propagation to subsequent frames. In this work, we consider the typical packetization method where each row of coded MBs is encapsulated into one RTP/UDP/IP packet, and every packet is independently decodable. Fig.1 illustrates a sketch map of the error-propagation effect caused by different packet losses, where two video packets marked as black rectangles are lost in different P-frames, and then the lighter color in subsequent P-frames represents the smaller distortion of reconstruction video. As can be seen, the reconstruction video can propagate errors both temporally and spatially to subsequent frames in the GOP.

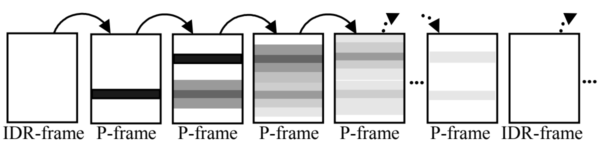

Fig.1 A sketch map of the error-propagation effect

Although there are many MB types in H.264/AVC, we divide these types into two basic modes: the intra-mode that can suppress error propagation, and the inter-mode that can result in error propagation. Let N_{GOP} and M respectively denote the total number of frames in a GOP and the total number of MBs in each frame, and let β_n denote the number of forced intra-mode MBs in Frame n ($1\leq n\leq N_{GOP}$). So β_n/M is the corresponding FIR ratio in Frame n. It is obvious that the number of forced intra-mode MBs in an IDR-frame is always equal to M. In the proposed unequal FIR scheme, the unequal numbers of forced intra-mode MBs are optimally increased to different P-frames in a GOP according to a GOP-level error-propagation model. In the following section, we will analyze the GOP-level error-propagation model.

GOP-level error-propagation model

For video streaming over lossy channels, the packet loss rate p can be estimated using the statistical reports such as RTCP. If each packet contains the same number of MBs (or pixels), the packet loss rate is equal to the loss ratio of a pixel. To take into account the numerical effect of error propagation, the pixel-level error propagation is estimated by modifying the Recursive Optimal Per-pixel Estimate (ROPE) (Zhang *et al.*, 2000; He *et al.*, 2002). Let $f(n,i)$ denote the original value of Pixel i in Frame n, $\hat{f}(n,i)$ and $\tilde{f}(n,i)$ denote its corresponding reconstruction value at the sender and that at the receiver, respectively. For simplicity but without loss of generality, we assume that error concealment in P-frames is employed by copying the pixel at the same location from the previous decoded frame. If the P-frame pixel $f(n,i)$ in the intra-mode is received correctly, its reconstruction value $\tilde{f}(n,i)$ at the receiver is equal to $\hat{f}(n,i)$; otherwise, its reconstruction value $\tilde{f}(n,i)$ is equal to $\tilde{f}(n-1,i)$, which is copied from the previous decoded frame. We assume that the source distortion and the channel distortion are uncorrelated with each other. Then the error-propagation estimation of the P-frame pixel $f(n,i)$ in the intra-mode can be expressed by the "Mean Square Error" criterion as follows:

$$\begin{aligned}
&PEP_{\text{intra}}(n,i,p)\\
&\quad = E\{[\hat{f}(n,i)-\tilde{f}(n,i)]^2\}\\
&\quad = (1-p)\cdot E\{[\hat{f}(n,i)-\hat{f}(n,i)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n,i)-\tilde{f}(n-1,i)]^2\}\\
&\quad \approx p\cdot E\{[\hat{f}(n,i)-\hat{f}(n-1,i)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n-1,i)-\tilde{f}(n-1,i)]^2\}.
\end{aligned} \tag{1}$$

Let $\hat{e}(n,i)$ be its corresponding motion-compensation difference for the P-frame pixel $f(n,i)$ in the inter-mode. Then in the case of no MB loss, its reconstruction value $\tilde{f}(n,i)$ at the receiver can be expressed as $\hat{e}(n,i)+\tilde{f}(n-1,j)$, where $\tilde{f}(n-1,j)$ is the corresponding motion-estimation value. If the corresponding MB is lost, its reconstruction value $\tilde{f}(n,i)$ is equal to $\tilde{f}(n-1,i)$. Thus the error-propagation estimation of the P-frame pixel $f(n,i)$ in the inter-mode can be expressed as follows:

$$\begin{aligned}
&PEP_{\text{inter}}(n,i,p)\\
&\quad = E\{[\hat{f}(n,i)-\tilde{f}(n,i)]^2\}\\
&\quad = (1-p)\cdot E\{[\hat{f}(n,i)-\hat{e}(n,i)-\tilde{f}(n-1,j)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n,i)-\tilde{f}(n-1,i)]^2\}\\
&\quad = (1-p)\cdot E\{[\hat{f}(n-1,j)-\tilde{f}(n-1,j)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n,i)-\hat{f}(n-1,i)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n-1,i)-\tilde{f}(n-1,i)]^2\}.
\end{aligned} \tag{2}$$

Therefore, using Eqs.(1) and (2), we can deduce the following pixel-level error-propagation estimation:

$$\begin{aligned}
&PEP(n,i,p,\beta_n)\\
&\quad = E\{[\hat{f}(n,i)-\tilde{f}(n,i)]^2\}\\
&\quad = \frac{\beta_n}{M}\cdot PEP_{\text{intra}}(n,i,p)+\left(1-\frac{\beta_n}{M}\right)\cdot PEP_{\text{inter}}(n,i,p)\\
&\quad = p\cdot E\{[\hat{f}(n,i)-\hat{f}(n-1,i)]^2\}\\
&\qquad +\left(1-\frac{\beta_n}{M}\right)(1-p)\cdot E\{[\hat{f}(n-1,j)-\tilde{f}(n-1,j)]^2\}\\
&\qquad +\left[\frac{\beta_n}{M}p+\left(1-\frac{\beta_n}{M}\right)p\right]\cdot E\{[\hat{f}(n-1,i)-\tilde{f}(n-1,i)]^2\}\\
&\quad = p\alpha\cdot E\{[f(n,i)-f(n-1,i)]^2\}\\
&\qquad + p\cdot E\{[\hat{f}(n-1,i)-\tilde{f}(n-1,i)]^2\}\\
&\qquad +\left(1-\frac{\beta_n}{M}\right)(1-p)\cdot E\{[\hat{f}(n-1,j)-\tilde{f}(n-1,j)]^2\}\\
&\quad = p\alpha\cdot E\{[f(n,i)-f(n-1,i)]^2\}\\
&\qquad +\left(1-\frac{\beta_n}{M}\right)(1-p)\cdot PEP(n-1,j,p,\beta_{n-1})\\
&\qquad + p\cdot PEP(n-1,i,p,\beta_{n-1}),\quad 2\le n\le N_{\text{GOP}},
\end{aligned} \tag{3}$$

where α is a constant that can be regarded as the energy leakage ratio caused by the coding distortion (Stuhlmuller *et al.*, 2000). After estimating the pixel-level error propagation, we sum up all $PEP(n,i,p,\beta_n)$ in Frame n to calculate the frame-level error-propagation estimation $FEP(n,p,\beta_n)$ recursively as follows:

$$\begin{aligned}
&FEP(n,p,\beta_n)\\
&\quad = \sum_{i\in frame} PEP(n,i,p,\beta_n)\\
&\quad = \sum_{i\in frame} p\alpha\cdot[f(n,i)-f(n-1,i)]^2\\
&\qquad + \sum_{i\in frame} p\cdot PEP(n-1,i,p,\beta_{n-1})\\
&\qquad + \sum_{j\in frame}\left(1-\frac{\beta_n}{M}\right)(1-p)\cdot PEP(n-1,j,p,\beta_{n-1})\\
&\quad = p\alpha\cdot\sum_{i\in frame}[f(n,i)-f(n-1,i)]^2\\
&\qquad + p\cdot FEP(n-1,p,\beta_{n-1})\\
&\qquad +\left(1-\frac{\beta_n}{M}\right)(1-p)\cdot FEP(n-1,p,\beta_{n-1})\\
&\quad =\left[1-(1-p)\frac{\beta_n}{M}\right]\cdot FEP(n-1,p,\beta_{n-1})\\
&\qquad + p\alpha\cdot\sum_{i\in frame}[f(n,i)-f(n-1,i)]^2,\ 2\le n\le N_{\text{GOP}},
\end{aligned} \tag{4}$$

UNEQUAL FORCED-INTRA-REFRESH

On the assumption that total number of forced intra-mode MBs in each GOP is L_{GOP}, β_n is determined using the following average FIR scheme in current H.264/AVC tools:

(1) n=1

$$\beta_n=M,$$

(2) $2\le n\le N_{\text{GOP}}$

$$\beta_n=\bar{\beta}=(L_{\text{GOP}}-M)/(N_{\text{GOP}}-1), \tag{5}$$

where $\overline{\beta}$ is a pre-determined coding parameter for the FIR. The number β_n of forced intra-mode MBs must be determined before Frame n is encoded, and the error-propagation effect from its previous frame may be regarded as an error fading process in the GOP due to the prediction loop and intra refresh. To allocate a reasonable FIR ratio for Frame n, the proposed scheme needs to estimate the following GOP-level error propagation from the previous frame of Frame n:

$$GEP(n,p,\overline{\beta}) = \begin{cases} FEP(1,p,M)\cdot \sum_{k=2}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-1}, \ n=2, \\ FEP(n-1,p,\overline{\beta})\cdot \sum_{k=n}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-n+1}, \\ \qquad 3\le n\le N_{\rm GOP}, \end{cases} \quad (6)$$

where $FEP(1,p,M)$ denotes the frame-level error propagation of an IDR-frame. The higher $GEP(n,p,\overline{\beta})$ value implies larger GOP-level error propagation. Therefore, performing more frequent FIR for Frame n with the higher $GEP(n,p,\overline{\beta})$ value usually can more effectively terminate error propagation. To realize real-time encoding or transcoding, we need to further analyze the above GOP-level error-propagation model. Based on Eqs.(4) and (6), we can deduce the following $GEP(n,p,\overline{\beta})$ values in a GOP:

$$GEP(2,p,\overline{\beta}) = FEP(1,p,M)\cdot \sum_{k=2}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-1},$$

$$GEP(3,p,\overline{\beta}) = FEP(2,p,\overline{\beta})\cdot \sum_{k=3}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-2}$$
$$=\{p\alpha\sum_{i\in frame}[f(2,i)-f(1,i)]^2+[1-(1-p)\overline{\beta}/M]$$
$$\cdot FEP(1,p,M)\}\sum_{k=3}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-2},$$

$$GEP(4,p,\overline{\beta}) = FEP(3,p,\overline{\beta})\cdot \sum_{k=4}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-3}$$
$$=\{p\alpha\sum_{i\in frame}[f(3,i)-f(2,i)]^2+[1-(1-p)\overline{\beta}/M]$$
$$\cdot FEP(2,p,\overline{\beta})\}\cdot \sum_{k=3}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-3},$$
$$\vdots$$
$$GEP(n,p,\overline{\beta})=FEP(n-1,p,\overline{\beta})\cdot \sum_{k=n}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-n+1}$$
$$=\{p\alpha\sum_{i\in frame}[f(n-1,i)-f(n-2,i)]^2+[1-(1-p)\overline{\beta}/M]$$
$$\cdot FEP(n-2,p,\overline{\beta})\}\cdot \sum_{k=n}^{N_{\rm GOP}}[1-(1-p)\overline{\beta}/M]^{k-n+1}, 5\le n\le N_{\rm GOP}. \quad (7)$$

In many multicast or broadcast systems, the most important IDR-frames are normally assigned a higher priority so that it has a great probability to be delivered almost error-free when network conditions worsen (Carpenter and Nichols, 2002; Kumwilaisak *et al.*, 2003). In this work, we assume that the packets in IDR-frames are transferred over an error-free channel to ensure random access and interactive selection for video streaming, while the packets in P-frames are transferred over a best-effort channel. Furthermore, $f(n-1,i)-f(n-2,i)$ at a short interval may be regarded as a stationary random variable. For a specific streaming application, the packet loss rate p usually is a certain value between l and u ($1\%\le l\le u<100\%$) where the step is 1%. So we can get the following average value of $GEP(n,p,\overline{\beta})$ within usual packet loss rates:

$$GEP(n,\overline{\beta}) = \frac{1\%}{u-l+1\%}\cdot \sum_{p=l}^{u} GEP(n,p,\overline{\beta}). \quad (8)$$

In the proposed unequal FIR scheme, β_n is determined using the following recursive method:

(1) n=1

$$\beta_n=M,$$
$$L_{\rm GOP}=\overline{\beta}(N_{\rm GOP}-1)+M,$$

(2) $2\le n\le N_{\rm GOP}$

$$\beta_n=\min\left(\frac{GEP(n,\overline{\beta})\left(L_{\rm GOP}-\sum_{k=1}^{n-1}\beta_k\right)}{\sum_{k=n}^{N_{\rm GOP}}GEP(k,\overline{\beta})},\ M\right). \quad (9)$$

Thus we can obtain a numerical distribution of β_n in a GOP, which is not dependent on a timely feedback channel. Fig.2 gives an example about the distribution of β_n in the proposed scheme, where the packet loss rates are between 1% and 20%.

Although the corrupted P-frame closer to an IDR-frame usually impairs more frames in a GOP, its corresponding frame-level error propagation is relatively low when the packets in IDR-frames are protected better. As a result, the P-frames in the middle of a GOP usually have more serious GOP-level error propagation, and more forced intra-mode MBs should be assigned to these P-frames.

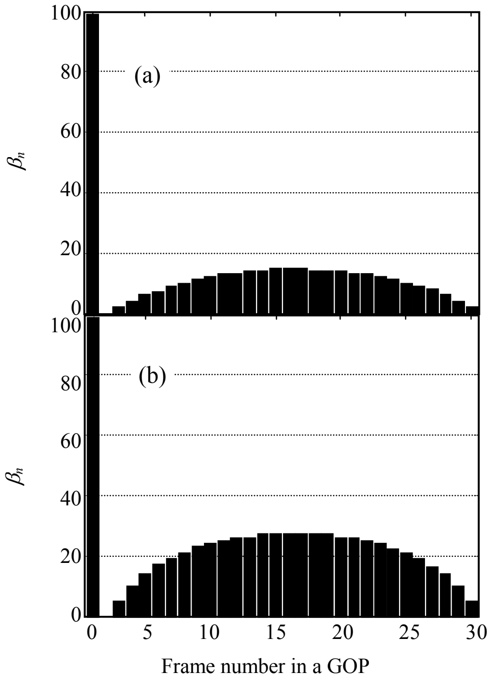

Fig.2 Distribution of β_n in the unequal FIR scheme. M=99, N_{GOP}=30. (a) $\bar{\beta}$=10; (b) $\bar{\beta}$=20

SIMULATION RESULTS

In this section, extensive simulations are carried out to evaluate the performance of the unequal FIR scheme. Video streaming is implemented based on the H.264/AVC reference software JM8.6. Two QCIF (M=99) test sequences News and Foreman are encoded at 30 frames per second. Each GOP is composed of one IDR-frame and 29 P-frames (N_{GOP}=30), and each row of coded MBs is encapsulated into one RTP/UDP/IP packet. To achieve relatively uniform perceptual quality, the codec outputs the variable-bit-rate video stream using the same quantization parameter. Inter pixels are not used for the intra prediction of intra-mode MBs. The proposed unequal FIR scheme is compared with the average FIR scheme in current H.264/AVC tools. The performance is measured by the peak signal-to-noise ratio of Y component ($PSNR$-Y). Table 1 gives the coding results of the two schemes. With the same $\bar{\beta}$ value, the two schemes can get very similar coding bit-rate and $PSNR$-Y at the sender.

Table 1 Coding results of the two schemes under different $\bar{\beta}$'s

Sequence	Scheme	Bit-rate (kb/s)		*PSNR-Y* (dB)	
		$\bar{\beta}$=10	$\bar{\beta}$=20	$\bar{\beta}$=10	$\bar{\beta}$=20
News	Proposed	199.45	281.22	36.16	36.05
	Average	200.33	281.66	36.07	35.97
Foreman	Proposed	269.86	339.60	35.38	35.27
	Average	270.77	341.53	35.36	35.24

For video streaming over simulated channels, we assume that the packets in IDR-frames are transmitted without packet losses, and that the packets in P-frames are transmitted at different packet loss rates. We adopt a Bernoulli channel model at the packet level to generate multiple packet loss patterns with four packet loss rates: 1%, 5%, 10% and 20%, respectively. For H.264/AVC decoder, error concealment is employed by the default method in JM8.6. Since the most important IDR-frames are error-free, the P-frames in the middle of a GOP usually have the larger GOP-level error propagation. Therefore, performing more frequent FIR for these P-frames can more effectively terminate error propagation. For other N_{GOP} values, we can draw the same conclusion through simulations. Fig.3 and Fig.4 illustrate the frame-by-frame $PSNR$-Y comparison of different schemes for different video streaming over simulated channels. From Fig.3 and Fig.4, we may observe that the proposed scheme can mitigate the overall error-propagation effect more effectively than the average FIR scheme in H.264/AVC tools.

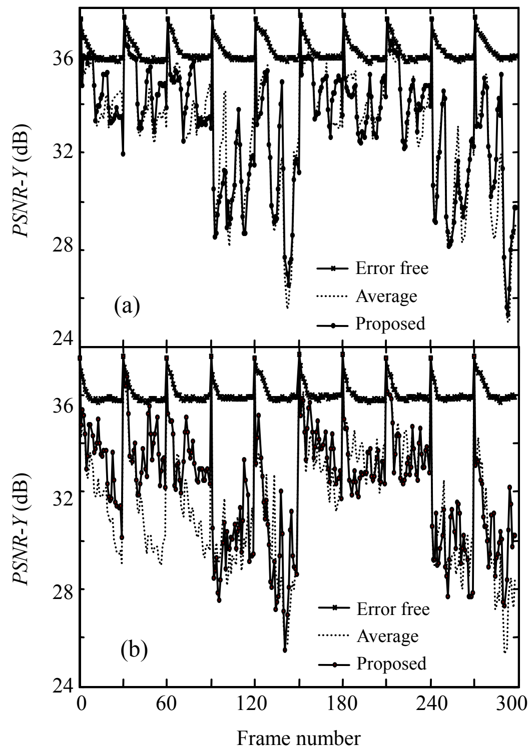

Fig.3 Frame-by-frame *PSNR-Y* comparison for News video streaming. (a) p=10%, $\bar{\beta}$ =10; (b) p=20%, $\bar{\beta}$ =20

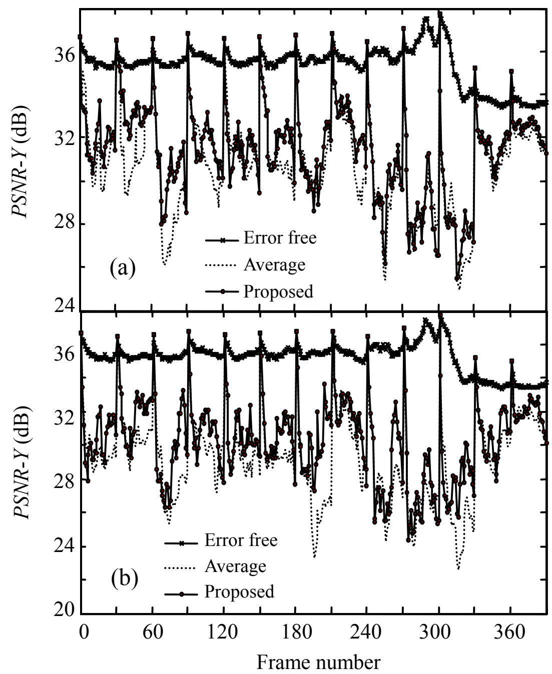

Fig.4 Frame-by-frame *PSNR-Y* comparison for Foreman video streaming. (a) p=10%, $\bar{\beta}$ =10; (b) p=20%, $\bar{\beta}$ =20

Fig.5 shows the average *PSNR-Y* comparison of different schemes at different packet loss rates and $\bar{\beta}$ values. We can see that the proposed scheme can achieve about 0.1~0.9 dB gains as compared to the average FIR scheme.

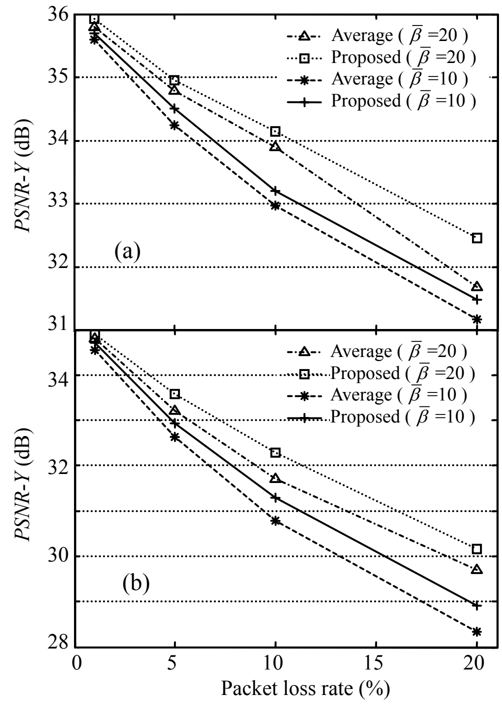

Fig.5 Average *PSNR-Y* comparison for News (a) and Foreman (b) video streaming

CONCLUSION

In this paper, we have proposed an unequal Forced-Intra-Refresh (FIR) scheme for robust video streaming. The proposed scheme can effectively mitigate the error-propagation effect caused by different packet losses, thus improve the overall video quality. Because the proposed scheme does not need to introduce the high-complexity mode decision, it can be a useful error-resilience tool in real-time encoding or transcoding of video streaming.

References

Carpenter, B.E., Nichols, K., 2002. Differentiated services in the Internet. *Proceedings of the IEEE*, **90**(9):1479-1494. [doi:10.1109/JPROC.2002.802000]

Chiou, H.J., Lee, Y.R., Lin, C.W., 2005. Content-aware error resilient transcoding using prioritized intra-refresh for video streaming. *Journal of Visual Communication and Image Representation*, **16**(4-5):563-588. [doi:10.1016/j.jvcir.2004.11.009]

Côté, G., Shirani, S., Kossentini, F., 2000. Optimal mode selection and synchronization for robust video communications over error-prone networks. *IEEE J. Select. Areas Commun.*, **18**(6):952-965. [doi:10.1109/49.848249]

He, Z.H., Cai, J.F., Chen, C.W., 2002. Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. *IEEE Trans. Circuits Syst. Video Technol.*, **12**(6):511-523. [doi:10.1109/TCSVT.2002.800313]

Kumwilaisak, W., Hou, Y.T., Zhang, Q., Zhu, W., Kuo, C.C., Zhang, Y.Q., 2003. A cross-layer quality of service mapping architecture for video delivery in wireless networks. *IEEE J. Select. Areas Commun.*, **21**(10):1685-1698. [doi:10.1109/JSAC.2003.816445]

Stockhammer, T., Kontopodis, D., Wiegand, T., 2002. Rate-distortion Optimization for JVT/H.26L Coding in Packet Loss Environment. Packet Video Workshop'2002. Pittsburgh, PY.

Stuhlmuller, K., Farber, N., Link, M., Girod, B., 2000. Analysis of video transmission over lossy channels. *IEEE J. Select. Areas Commun.*, **18**(6):1012-1032. [doi:10.1109/49.848253]

Tu, W., Eckehard, S., 2004. Proxy-based Error Tracking for H.264 Based Real-time Video Transmission in Mobile Environments. IEEE International Conference on Multimedia & Expo (ICME'04). Taipei, p.1367-1370.

Zhang, R., Regunathan, S.L., Rose, K., 2000. Video coding with optimal inter/intra-mode switching for packet loss resilience. *IEEE J. Select. Areas Commun.*, **18**(6):966-976. [doi:10.1109/49.848250]