CHANNEL-AWARE FRAME DROPPING FOR CELLULAR VIDEO STREAMING

Hao Liu, Wenjun Zhang, Songyu Yu, and Xiaokang Yang

Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, China

ABSTRACT

In the case of cellular video streaming over wireless channels, burst frame losses may be unavoidable. Considering the unequal importance of different frames in a group-of-pictures (GOP) and the burst-error characteristics of wireless channels, this paper proposes a channel-aware frame dropping scheme so as to shift burst losses into relatively unimportant frames in the same GOP. By using selective retransmission at the radio link layer, a base station can adaptively assign the unequal transmission attempts to different video frames. Simulation results show that the proposed scheme can be aware of the variation of wireless channel conditions, and thus significantly improve error resilience of cellular video streaming.

1. INTRODUCTION

Video streaming over 3G cellular networks has recently received increasing attention from the scientific and industrial communities. In cellular networks, wireless channels pose a major challenge for video streaming, as it has low bandwidth and high burst loss rate as compared to wired channels. Therefore, cellular video streaming over wireless channels is expected to experience burst frame losses and thus cause severe quality degradation. Compared with the passive frame losses in video transmission, active frame dropping at network nodes can discard some relatively unimportant frames according to time-varying channel conditions. In [1], a frame dropping strategy is proposed that discards those frames depending on the previously discarded frames. The authors in [2] proposed to discard video frames that are less important to human perception, and hence frames are discarded in order of the type importance (i.e. B-frame, P-frame and I-frame). The authors in [3] proposed to discard those frames that can minimize the likelihood of future frames being discarded. These schemes may be helpful to increase error resilience of video transmission. However, such mechanisms are implemented solely at the application layer, and may be insufficient for highly time-varying wireless channels where the frame-loss status is difficult to predict in advance. To enhance the robustness of cellular video streaming, joint consideration of two or more communication layers is desirable [4].

In this work, we extend the above idea about active frame dropping to the radio link-layer retransmission mechanism of cellular video streaming. Since cellular video streaming has loose delay constraint, we propose a channel-aware frame dropping scheme by adaptively assigning unequal link-layer transmission attempts to different frames in a GOP. In the proposed scheme, the unequal importance of different video frames at the application layer and the selective retransmission mechanism at the radio link layer are considered together. As a result, burst frame losses are shifted into relatively unimportant frames in the same GOP. The rest of the paper is organized as follows. In Section 2, we analyze the source-channel characteristics of cellular video streaming. Our proposed channel-aware frame dropping scheme is presented in details in Section 3. Section 4 gives simulation results that demonstrate the effectiveness of the proposed scheme. Finally, conclusions are drawn in Section 5.

2. CELLULAR VIDEO STREAMING

2.1. Burst-error characteristics of wireless channels

For typical cellular video, the compressed size of one video frame can become fairly small, and usually smaller than the maximum transmission unit (MTU) of the network. For example, if one considers transmission of QCIF video at 10 frames per second (fps) over a 64 kbit/s wireless channel, each coded frame occupies 800 bytes on average, not to mention the associated audio information. If the GOP structure is used, as often happens in many video streaming applications, the size of inter-frames can be even smaller [5]. On the other hand, the payload length larger than 750 bytes is normally required in order to ensure efficient packet header overhead [6]. In such scenarios, a single packet per video frame is frequently used, and thus packet losses correspond to whole-frame losses.

Fig.1 shows typical network architecture for cellular video streaming. At the sender, an encoder encapsulates the compressed bitstream into RTP/UDP/IP packets. A video streaming server responds to the user request, and delivery RTP/UDP/IP packets to the receivers through a base station. In the network architecture, the network resources bottleneck is at the radio interface. The base station plays a key role for robust video streaming, since it can obtain more timely feedback information than the server and adapt the video delivery to wireless channel conditions.

1-4244-0469-X/06/$20.00 ©2006 IEEE

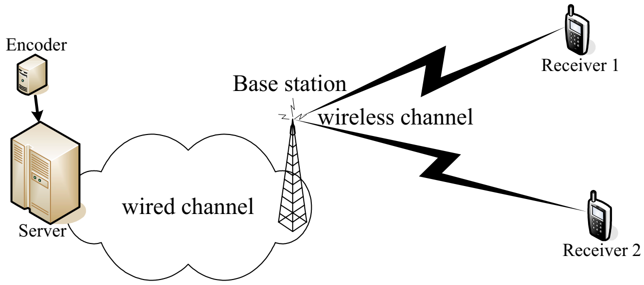


Fig.1. Network architecture for cellular video streaming.

2.2. Unequal importance in a GOP

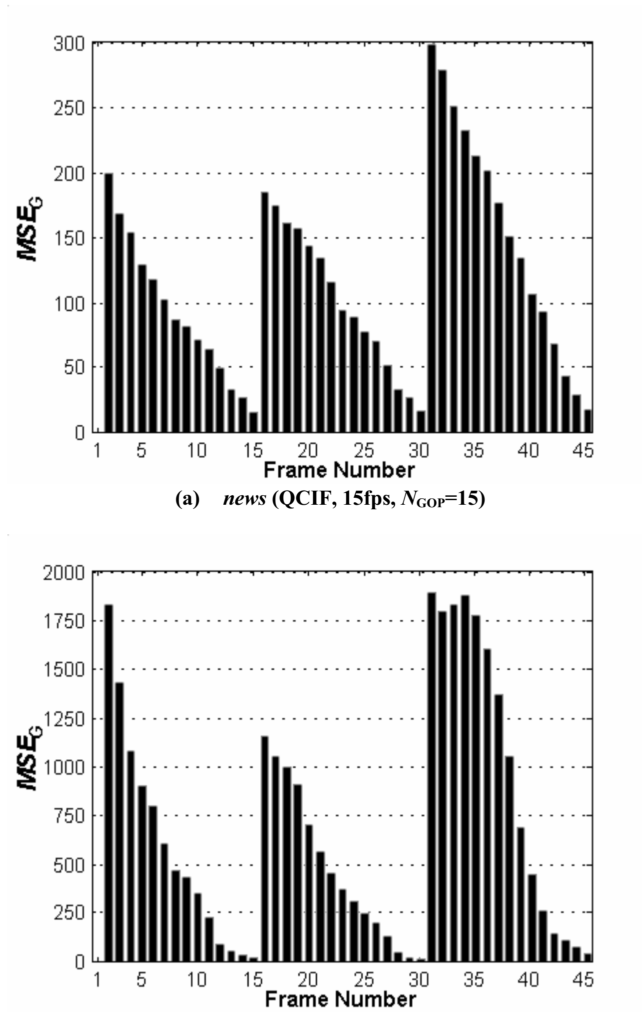
In prevalent video coding standards, the inter-frame prediction can efficiently remove the temporal redundancy between successive frames. For analysis convenience, let N_{GOP} denote the total number of frames in each GOP. Since B-frame losses do not interfere with other frames, we consider the frame dropping schemes in a typical GOP structure with one I-frame followed by "N_{GOP}-1" P-frames. In the structure, losing the preceding (the first or the first few) frames of a GOP often leads to the worse result. There have been some video transmission strategies that consider the unequal importance in a GOP. In [7], unequal forward-error-correction (FEC) codes are optimally assigned to different P-frames with the degressive importance in a GOP. In [8], different FEC codes are selected for the video transmission according to the channel conditions, frame type, frame delivery deadline, etc. However, the sole use of FEC is not effective to combat time-varying burst losses. Compared with FEC, automatic-repeat-request (ARQ) is more aware of the variation of wireless channel conditions. Channel-aware capability at a base station is achieved by an additional ARQ mechanism that privileges the important frames in a GOP.

Let $f(n,j)$ and $\tilde{f}(n,j)$ denote the original value of pixel j in Frame n ($1 \le n \le N_{\text{GOP}}$) of a GOP and its corresponding reconstruction value, respectively. When one frame loss occurs in the same GOP, the GOP-level distortion "MSE_G" defined in (1) is used to compute the mean square errors between the original GOP and reconstruction GOP.

$$MSE_G = \frac{1}{N_{\text{GOP}}} \sum_{n=1}^{N_{\text{GOP}}} \sum_{j=1}^{width \times height} \frac{\left[f(n,j) - \tilde{f}(n,j)\right]^2}{width \times height} \qquad (1)$$

where *width* and *height* are the dimensions of the video sequence. Although the proposed channel-aware frame dropping scheme does not depend on any error concealment algorithm, in this work, we assume a type of error concealment that is referred to as *previous frame repetition*. In the simplest yet commonly used way for error concealment of cellular video, an incorrectly decoded frame and all subsequent frames in the same GOP are replaced by the most recent correctly decoded frame. For instance, losing the last frame of a GOP leads to little distortion as just the second last frame of the GOP is displayed twice. *Previous frame repetition* can avoid the annoying artifact effect of the conventional concealment algorithms, especially in the presence of scene changes, fast motion, rotation and deformation of objects.

Using *previous frame repetition*, Fig.2 plots the corresponding GOP-level distortion MSE_G when losing different frames for two typical QCIF sequences, namely *news* and *foreman*. Each GOP is composed of 15 frames, and the vertical bars indicate the MSE_G value when the frame is lost individually. As can be seen from Fig.2, in most cases, the corresponding GOP-level distortion is larger when losing the preceding frame of the same GOP, and the MSE_G value decreases as the "Frame Number" increases in the same GOP. If there is little motion in the video sequence such as *news*, the loss of a frame has little effect on the overall quality of the reconstruction video; otherwise, if there is significant motion in the video sequence such as *foreman*, the effect of a frame loss can be dramatic.



(a) ***news*** **(QCIF, 15fps, N_{GOP}=15)**

(b) ***foreman*** **(QCIF, 15fps, N_{GOP}=15)**

Fig.2. GOP-level distortion versus Frame Number for two video sequences.

3. CHANNEL-AWARE FRAME DROPPING

For cellular video over wireless channels, we adopt the concept of application-level framing [9] to analyze the proposed scheme, where each video frame is encapsulated into one radio link-layer SDU (Service Data Unit) and any link-layer SDU with a corrupted bit-number greater than the capacity of error correction has to be completely discarded. In general, each P-frame in the same GOP has the similar size. We assume that the average size of an I-frame is $\boldsymbol{m}$ times the average size of a P-frame, where $\boldsymbol{m}$ can be obtained statistically. Thus one I-frame SDU is equivalent to $\boldsymbol{m}$ P-frame SDUs. If a base station receives the ARQ message about the current link-layer SDU, the base station will selectively retransmit the lost link-layer SDU according to a certain retransmission limit. The existing link-layer mechanism does not consider the unequal importance of different frames in a GOP, and each link-layer SDU has the same retransmission limit "$\boldsymbol{\eta}$" ($\boldsymbol{\eta} \geqslant 0$), which results in the non-optimal performance.

Obviously, a tradeoff exists between protection of the preceding frames and that of the subsequent frames in a GOP. If we assign more transmission attempts to the link-layer SDU of the preceding frames of the GOP, we have to assign less transmission attempts to the link-layer SDU of the subsequent frame. As the preceding frames of a GOP are more important, the proposed scheme can increase the probability of their error-free reception by discarding the latter frames. The frame set $\boldsymbol{S}_{\text{GOP}}=\{\boldsymbol{i},\boldsymbol{i}+1,\cdots,\boldsymbol{n}_{\text{GOP}}-1,\boldsymbol{n}_{\text{GOP}}\}$ ($1 \le \boldsymbol{i} \le \boldsymbol{n}_{\text{GOP}} \le \boldsymbol{N}_{\text{GOP}}$) is composed of the video frames to be transmitted in a GOP. If burst frame losses occur in Frame $\boldsymbol{i}$ of the GOP, the proposed scheme firstly computes by (2) its maximum transmission times "$\boldsymbol{\Phi}_i$" for Frame $\boldsymbol{i}$ so as to obtain the similar transmission cost compared with the traditional scheme.

$$\boldsymbol{\Phi}_i = \begin{cases} (\boldsymbol{\eta}+1)\cdot\left(1+\left\lfloor \dfrac{N_{\text{GOP}}-1}{\boldsymbol{m}} \right\rfloor\right) & (i=1) \\ (\boldsymbol{\eta}+1)\cdot(n_{\text{GOP}}-\boldsymbol{i}+1) & (i \ge 2) \end{cases} \qquad (2)$$

During the burst-error period, the proposed scheme keeps retransmitting Frame $\boldsymbol{i}$ under its maximum transmission times "$\boldsymbol{\Phi}_i$", and cumulates its ARQ times "$\boldsymbol{C}_i$". When either burst losses end or the cumulative value $\boldsymbol{C}_i$ reaches $\boldsymbol{\Phi}_i$, the proposed scheme stops retransmitting Frame $\boldsymbol{i}$, and then the number "$\boldsymbol{A}_i$" of the latter frames to be discarded in the frame set $\boldsymbol{S}_{\text{GOP}}$ can be formulated as:

$$\boldsymbol{A}_i = \begin{cases} \min\{m\cdot\left\lfloor \dfrac{\boldsymbol{C}_i}{\eta+1} \right\rfloor,\ N_{\text{GOP}}\} & (i=1) \\ \left\lfloor \dfrac{\boldsymbol{C}_i}{\eta+1} \right\rfloor & (i \ge 2) \end{cases} \qquad (3)$$

Fig.3 generalizes the flowchart of the proposed channel-aware frame dropping scheme. When the frame set $\boldsymbol{S}_{\text{GOP}}$ become empty, the base station will transmit the next GOP.

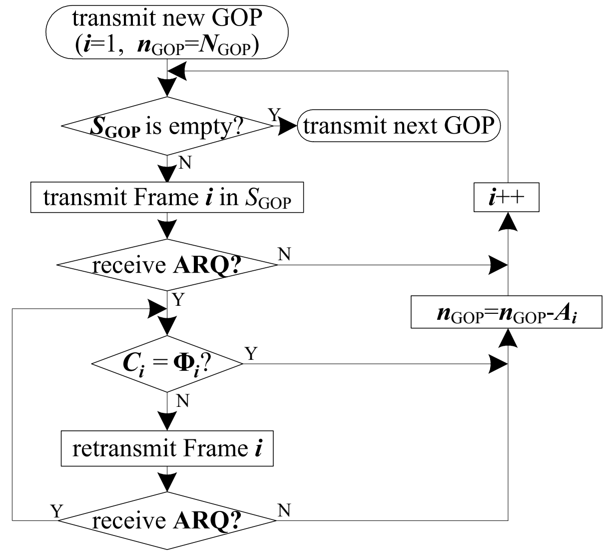


Fig.3. The flowchart of channel-aware frame dropping.

Fig.4 illustrates a simple example of the proposed scheme used for cellular video streaming. When burst losses occur in Frame *7* of a GOP, the proposed scheme keeps retransmitting Frame *7* under its transmission limit $\boldsymbol{\Phi}_7$. After Frame *7* is successfully received, the latter "$\boldsymbol{A}_7$=4" frames including Frame ***12***, Frame ***13***, Frame ***14*** and Frame ***15*** have to be discarded. In this way, burst losses are shifted into relatively unimportant frames. Therefore, abrupt quality degradation of received video can be mitigated.

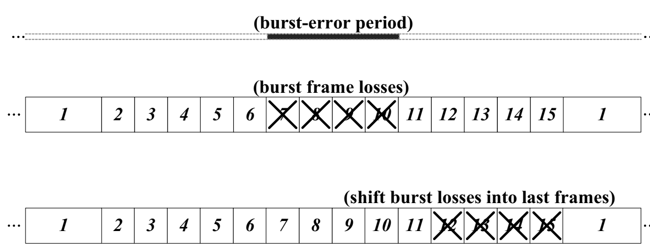


Fig.4. An example of channel-aware frame dropping.

4. SIMULATION RESULTS

Simulations have been carried out to evaluate the performance of the proposed channel-aware frame dropping scheme. Two typical QCIF sequences, namely *foreman* and *news*, are encoded using H.264/AVC reference software JM9.6. Each GOP is composed of "$\boldsymbol{N}_{\text{GOP}}$=15" frames. The original video (30 fps) is encoded once every two frames with the coding format of IPPP···IPPP···. Using the constant quantization parameter, the coding bitrate is set to about 64 kbit/s. Without loss of generality, we assume "$\boldsymbol{m}$=5". In simulations, we assume that the losses in wired channels are negligible. The video streaming is directly transported over a simulated wireless channel, which is realized as a twostate Markov model. The trace statistics clearly present a burst-error behavior when wireless channel condition is poor. In the decoder, error concealment is employed by using *previous frame repetition*. We will compare the performance of two schemes, i.e., (a) **CA-FD**: the proposed channel-aware frame dropping; (b) **EL-FD**: the traditional frame dropping with the equal retransmission limit. The performance is measured by the peak signal-to-noise ratio (PSNR) of the reconstruction frame.

The average PSNR comparison of different schemes is shown in Table I. The simulation results show that the proposed **CA-FD** scheme can achieve significant average PSNR gain over the traditional **EL-FD** scheme. Furthermore, if the video sequence such as *foreman* has larger motion, the proposed scheme can achieve the larger average PSNR gain.

Table I Average PSNR comparison of different schemes for two video sequences.

	Error free	EL-FD	CA-FD
news	36.23 **dB**	28.60 **dB**	32.59 **dB**
foreman	35.16 **dB**	23.44 **dB**	30.74 **dB**

The frame-by-frame PSNR comparison is shown in Fig.5, where the PSNR is presented once every second frame. The proposed **CA-FD** scheme can discard relatively unimportant frames by more frequently retransmitting important frames in the same GOP. Thus, it is observed that the proposed **CA-FD** scheme can more effectively mitigate burst-error effects than the traditional **EL-FD** scheme in the case of frame losses.

5. CONCLUSIONS

We have proposed a channel-aware frame dropping scheme for cellular video streaming over wireless channels. Simulation results show that the proposed scheme can shift burst losses into relatively unimportant frames in a GOP, thus significantly improve error resilience of cellular video streaming. Since the proposed scheme is not based on scalable coding, it can be applied to any video coding standard.

REFERENCES

[1] G. Ravindra, N. Balakrishnan, and K.R. Ramakrishnan, "Active router approach for selective packet discard of streamed MPEG video under low bandwidth conditions", in *Proc. IEEE Int. Conf. Multimedia and Expo*, New York, July 2000.

[2] E. Gurses, G.B. Akar, and N. Akar, "Selective frame discarding for video streaming in TCP/IP networks", in *Proc. Packet Video Workshop*, Nantes, France, Apr. 2003

[3] Z.L. Zhang, S. Nelakuditi, R. Aggarwal, and R.P. Tsang, "Efficient selective frame discard algorithms for stored video delivery across resource constrained networks", *Real-Time Imaging*, vol. 7, no. 3, pp. 255-273, June 2001.

[4] Lai-U Choi, W. Kellerer, and E. Steinbach, "Cross-layer optimization for wireless multi-user video streaming", in *Proc. Int. Conf. Image Processing*, Singapore, Oct. 2004

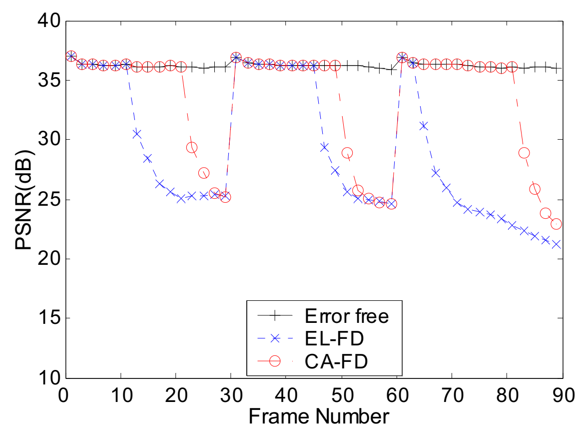
[5] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 316-329, Apr. 2005.

[6] 3GPP TR 26.937 V6.0.0, "Transparent end-to-end packet switched streaming service (PSS); RTP usage model (Release 6)", Mar. 2004.

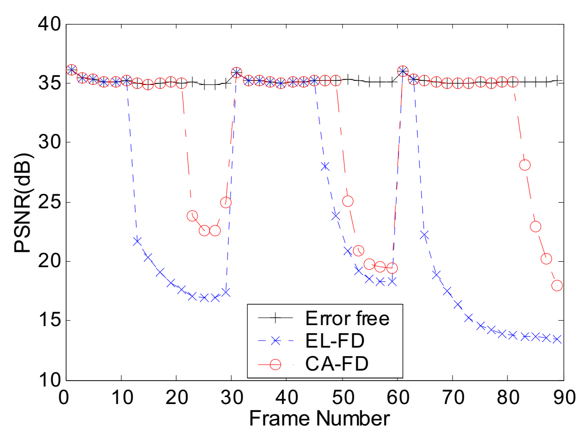
[7] X.K. Yang, C. Zhu, Z.G. Li, X. Lin, G.N. Feng, S. Wu and N. Ling, "Unequal loss protection for motion compensated video streaming over the Internet", *Signal Processing: Image Communication*, vol. 18, no. 3, pp. 157-167, Mar. 2003.

[8] D. Qiao and K.G. Shin, "A two-step adaptive error recovery scheme for video transmission over wireless networks", in *Proc. IEEE INFOCOM'2000* , Tel-Aviv, Israel, Mar. 2000

[9] S. Wenger, M.M. Hannuksela, T. Stockhammar, M. Westerlund, and D. Singer, "RTP payload format for H.264 video", *IETF, Request for Comments: 3984*, Feb. 2005.



(a) "*news*" video streaming



(b) "*foreman*" video streaming

Fig.5. Frame-by-frame PSNR comparison of different schemes for two video sequences.