# CROSS-LAYER FRAME DISCARDING FOR CELLULAR VIDEO CODING

*Chongyang Zhang, Songyu Yu, Hua Yang, and Hongkai Xiong*

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

## ABSTRACT

In the case of delivering real-time video over the 3G cellular networks, burst frame losses may be inevitable and unpredictable, which may cause severe quality degradation. Based on cross-layer frame discarding (CLFD), this paper proposes an enhanced error-resilient video coding scheme for cellular video communication. By using unequal retransmission at the radio link (RL) layer, a base station can provide reliable transmission for the relatively important frames in one video sequence. Relying on the unequal protection at the RL layer, the encoder at the application (APP) layer can actively discard a certain number of frames according to the received acknowledgement messages. Thus, unpredictable burst frame losses during transmission can be transformed into selective frame discarding at the encoder. Experiments results show that the proposed scheme can enhance the error resilience of the cellular video communication significantly.

***Index Terms***— Cross-layer frame discarding, cellular network, error resilience, reference pictures selection

## 1. INTRODUCTION

With the great progress in digital video compression and network technology, video applications over 3G cellular networks have recently received increasing attention. Therefore, due to the low bandwidth and high burst loss rate of the wireless channels, cellular video communication over 3G networks is expected to experience burst frame losses and thus cause severe quality degradation. Fortunately, many effective error control techniques are proposed for the mobile video transmission [1]. Moreover, the prevalent video coding standards, such as H.264/AVC, have included many advanced features to enhance their error resilience for wireless system [2]. Besides these schemes implemented solely at the application layer, many more efficient cross-layer wireless multimedia compression and transmission strategies have emerged recently, which could be used to enhance the robustness of wireless video communication by joint consideration of two or more protocol stacks [3, 4].

However, most of the cross-layer video techniques are focused on transporting of the stored video streaming [3, 4]. In this work, we extend them to the live video coding domain. In other words, video coding at APP layer will be considered to cooperate with the lower RL layer transmission to achieve improved performance. By using unequal retransmission at the RL layer, a base station can provide reliable transmission for the relatively important frames in one video sequence [3]. Relying on the unequal protection at the RL layer, the encoder at the APP layer can actively discard a certain number of frames according to the received acknowledgement messages. Thus, unpredictable burst frame losses during transmission can be transformed into selective frame discarding at the encoder. In case of burst errors, the distance between the reference picture and the current frame (temporal-dependency-distance, TDD) may be large due to the long round-trip-time (RTT) delay of the feedback messages, which may reduce the compression efficiency of the reference-pictures-selection (RPS) based video coding system, such as RPS-based H.263+, NEWPRED-based MPEG-4 and LTMP-based H.264 [6]. However, the proposed video coding scheme can minimize the TDD by cross-layer frame discarding. As a result, increased rate-distortion performance and enhanced error resilience of the cellular video coding can be achieved.

The rest of this paper is organized as follows. In Section 2, we analyze the source-channel characteristics of the end-to-end cellular video communication and the RL layer unequal retransmission mechanism. Section 3 presents the cross-layer frame discard algorithm for the cellular video coding. Simulation results are shown in Section 4. Finally, conclusions are drawn in Section 5.

## 2. UNEQUAL RETRANSMISSION AT RADIO LINK LAYER

### 2.1. Video transmission over the cellular wireless channel

For typical cellular video, the compressed size of one video frame can become fairly small (800 bytes on average for 10 frame per second of QCIF video over 64 Kbit/s wireless channel), and a single packet per video frame is often used to ensure efficient packet header overhead [3]. Thus, packet losses correspond to whole frame losses.

This work was supported by the NSF of China under grant No. 60502033 and the PhD. Special Research under grant No. 20040248047.

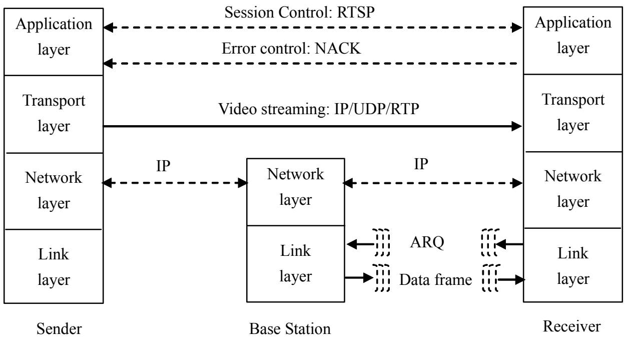


**Fig.1. The open end-to-end architecture for cellular video communication**

Fig.1 illustrates the typical network architecture for end-to-end cellular video communication, where coded video streaming is delivered from the sender to the receiver through a base station. At the sender, each captured live video frame is compressed firstly, and then the compressed bits are mapped to a RTP/UDP/IP payload. After robust header compression, the RTP/UDP/IP packet is segmented into several radio link layer frames with fixed size. For cellular video communication, the application-level framing policy, where each video frame is encapsulated into one link-layer SDU (Service Data Unit) [3], is adopted in the proposed scheme. Due to most burst errors occur at the wireless channel, the radio interface between the base station and the receiver is the network resources bottleneck. As a solving measure, timely feedback information (auto repeat request, ARQ)-based retransmission is utilized by most base stations. At APP layer, negative acknowledgement (NACK)-based error control technique is utilized to adapt the encoder to the channel.

### 2.2. Unequal retransmission at the RL layer

If a base station receives the ARQ message about the current link-layer SDU, the base station will selectively retransmit this SDU according to a certain retransmission limit. The existing link-layer mechanism utilizes the equal retransmission limit: each link-layer SDU has the same retransmission limit "$\eta$" ($\eta \geq 0$), which result in the non-optimal performance. Considering the unequal importance of different video frame in a group-of-pictures (GOP), more efficient unequal retransmission policy is proposed in [3]: the base station keeps retransmission video frame $i$ when a burst error occurs under its maximum retransmission times ($\Phi_i$, which can be calculated by equation (2) in [3]); then the corresponding number of the latter frames in the same GOP are discarded.

Therefore, the scheme in [3] is focused on the stored video streaming, where the frame dropping is implemented at RL layer. To enhance the system performance further, we shift the frame dropping from RL layer to the APP layer (video coding domain). At RL layer, the base station keeps retransmission video frame $i$ when a burst error occurs until its maximum retransmission-time-delay (RTD) limitation (denoted by $M$, an assumed maximum successive retransmission times) is reached. Thus, different from that in [3], the definition of "$\Phi_i$" can be formulated as:

$$\Phi_i = \begin{cases} (\eta+1)\cdot\left(1+\left\lfloor\dfrac{M-1}{l}\right\rfloor\right) & (\textit{for I-frame}) \\ (\eta+1)\cdot M & (\textit{for P-frame}) \end{cases} \tag{1}$$

Here $l$ is an assumed value according to the statistics of the average I-frame size vs. that of P-frame.

At the receiver, once a frame is not received correctly after "$\eta+1$" times ARQ, the RL layer will inform the APP layer to send a NACK message to the sender. At the sender, when the NACK is received after one RTT delay, the encoder will update the statistics of the acknowledgement messages and make a decision to either encode the current normally or discard it entirely. The frame discarding is used to tradeoff the overall retransmission attempts at the RL layer. Section 3 gives its algorithm.

## 3. CROSS-LAYER FRAME DISCARDING AT THE ENCODER

### 3.1. RPS-based error resilient video coding

In the current low bit-rate video coding schemes, the typical video GOP structure is with one I-frame followed by $N_{GOP}$-1 P and B-frames. As B-frame losses do not interfere with other frames, we only consider the IPPP…IPPP…structure. In this structure, frames losses will result in interframe error propagation. Many video coding and transporting strategies have been proposed to avoid or reduce the error propagation. In [5], dynamic reference picture replacing scheme is utilized to stop the interframe error propagation. In [6], the prevalent feedback-based RPS modes, such as RPS of H.263+, NEWPRED of MPEG-4, are reviewed and a more efficient mode, Long-term memory prediction (LTMP) used for H.264, is proposed.

There are two features in the above schemes: (1) they are focused on the single frame (or group-of-blocks, GOB) loss; (2) the encoder has to use older reference pictures for the motion-compensated prediction with increasing round-trip time, which results in decreased coding efficiency. Unlike the existing error resilient video coding schemes, the proposed scheme can transform the unpredictable burst frame losses into active frame discarding by using cross-layer cooperation. Thus the encoder can selectively discard those frames that can minimize the TDD to achieve higher rate-distortion performance. As to the consecutive frame losses occurred at cellular network frequently, the proposed scheme could spread out the long burst errors effectively. In the following, for the sake of comparison, only the nearest error-free picture is adopted as the reference.

### 3.2. Cross-layer frame discard algorithm at the encoder

Here we define a policy vector $\Pi = \{\pi_i, \pi_{i+1}, \ldots, \pi_{N\text{-}1}\}$ for the possible discarded frame queue $Q$. When $\pi_j$ is set to "0", the frame $p_j$ ($i \le j \le$(N-1)) is discarded (the discarding can be achieved by skipping without coding). Otherwise, $p_j$ is encoded as I-frame or P-frame. If the frame $p_j$ is discarded, then the distortion introduced by its loss is $\Delta d_j$. The overall distortion $D(\Pi)$ and the total coding bit rate $R(\Pi)$ for the frame queue $Q$ can be expressed as:

$$D(\Pi) = \sum_{j=i}^{N-1} \Delta d_j (1 - \pi_j) \tag{2}$$

$$R(\Pi) = \sum_{j=i}^{N-1} r_j (1 - \pi_j) \tag{3}$$

Here $r_j$ is the coding bit rate of $p_j$. With equations (2) and (3) for the possible discarded frame queue, our goal is to seek the optimal policy vector $\Pi_{opt}$ that minimizes the overall distortion $D(\Pi)$ under the bandwidth budget $R_{max}$:

$$Minimize\ \ D(\Pi)\ \ s.t.\ \ R(\Pi) \le R_{\max} \tag{4}$$

Note that optimization approaches, such as Lagrangian Relaxation and Dynamic Programming, can be used to solve the constrained nonlinear optimization problem. However, in the video coding system, the estimation of the frame distortion $\Delta d_j$ is quite complex due to the time-varying temporal correlation, which must be taken into account. To simplify the rate-distortion optimization problem, we propose an efficient frame discard algorithm, which flowchart is generalized as Fig.2. The main intuition for this scheme is that: the smaller the distance that between the reference picture and the current frame (TDD) is, the more interframe correlation can be exploited and the higher compression efficiency will be achieved. Since minimum Consecutive Loss Factor (CLF, denoted by $k_0$) will lead to minimum TDD [7], we can calculate the value of $k_0$ by formula (5) according to [7], and then $\pi_j$ can be determined by formula (6).

$$k_0 = \begin{cases} 0 & (n_{NACK} = 0) \\ 1 & (0 < n_{NACK} \le \dfrac{m}{2}) \\ \left\lfloor \dfrac{n_{NACK}}{m - n_{NACK} + 1} \right\rfloor + 1 & (\dfrac{m}{2} < n_{NACK} < m) \\ m & (n_{NACK} \ge m) \end{cases} \tag{5}$$

$$\pi_j = \begin{cases} 1 & (n_{skip} \ge k_0 \ or\ p_j\ is\ I\text{-}frame) \\ 0 & (n_{skip} < k_0\ and\ p_j\ is\ P\text{-}frame) \end{cases} \tag{6}$$

Here, the "buffer" size $m$ (the reserved frame number that used to spread out the errors) is initialized to the RTD limitation $M$. Then, its value will decrease till be re-initialized conditionally to adapt the discard algorithm to the channel.

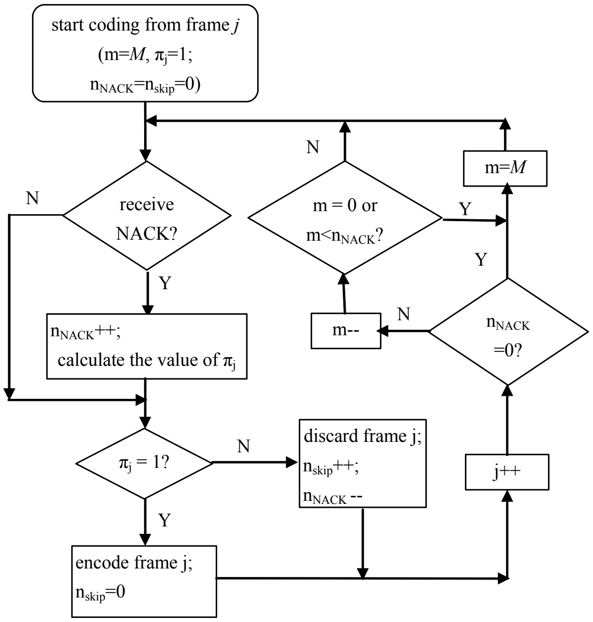
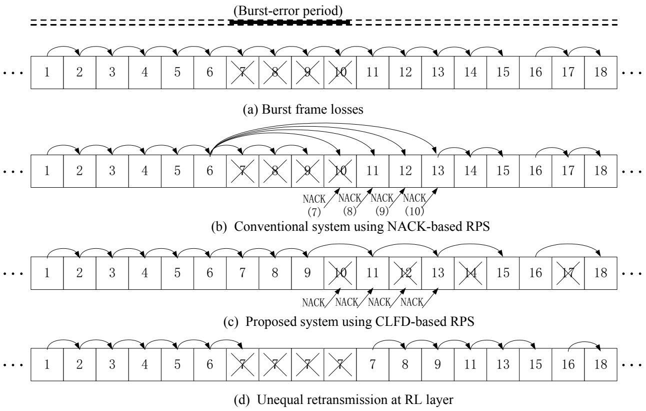


**Fig.2. The flowchart of frame discard algorithm at the encoder**



(a) Burst frame losses

(b) Conventional system using NACK-based RPS

(c) Proposed system using CLFD-based RPS

(d) Unequal retransmission at RL layer

**Fig.3. Illustration of coding algorithms for the conventional system and the proposed system**

Fig.3 illustrates the coding algorithms of the conventional system and the proposed system. Here we assume that NACK's are received error free after a relatively large round trip delay, e.g., 3 times of the coding interval. This delay covers several retransmission attempts and may be considered as a bad case estimate for the actual delay. The interval between two I-frames ($N_{GOP}$) is set to 15 and "$M$=8". When burst losses occur in frame 7 of a sequence (Fig.3 (a)), conventional system, such as H.263+ and [5], will code the frames 11—13 with reference to frame 6 (indicated by the arrows in (Fig.3 (b)). For the proposed system, selective frame discarding is adopted at the encoder (Fig.3 (c)) and unequal retransmission is utilized at RL layer (Fig.3 (d)), thus burst errors are shifted to frame 10, 12, 14 and 17. From Fig. 3 we can find that: the

TDD is four frames or more when using the traditional RPS schemes; while it is no more than two frames in the proposed scheme. Obviously, the proposed scheme will outperform the traditional schemes used in [1, 5] due to the smaller TDD achieved by using CLFD.

### 3.3. Decoding and displaying at the receiver

At the receiver, considering of the time delay for unequal retransmission at RL layer, a relatively longer decoding and displaying time delay may be needed. The length of the additional delay depends on the RTD limitation *M*, which can be controlled or traded-off expediently.

## 4. SIMULATION RESULTS

Sample results from two typical QCIF sequences, namely *News* and *Foreman*, are shown to evaluate the performance of the proposed cross-layer frame discarding scheme. By using H.264/AVC reference software JM10.2, the original video (30 fps) is encoded once every two frames with the coding format of IPPP…IPPP… ($N_{GOP}$=15). Without loss of generality, we set "$l$=5" for formula (1) and "$M$=8" for the cross-layer frame discard algorithm. In simulation, we assume that the feedback channels are error-free and the RTT delay is 3 times coding interval. The video streaming is directly transported over a simulated wireless channel, which is realized as a two-sate Markov model. The trace statistics clearly present a burst-error behavior when wireless channel condition is poor [3]. In the encoder, the rate control option is utilized to limit the encoding bit rate to about 64kbit/s. The previous frame repetition is used as the decoder's error concealment. For the sake of performance comparison, the two sequences are coded with three modes: (1) Error free; (2) W/O CLFD: Conventional RPS-based video coding without CLFD (used in [1, 5]); (3) With CLFD: Proposed video coding with CLFD.

The average PSNR-Y (peak signal-to-noise ratio of the luminance) comparison of the above three modes is shown in Table 1. From the comparison we can find that the proposed system with CLFD can improve the average PSNR gain significantly over the traditional scheme without CLFD. Furthermore, the higher motion the scenario is (*Foreman* sequence), the more PSNR gains can be achieved.

**Table 1 Average PSNR-Y comparison of different modes**

| | Error free | W/O CLFD | With CLFD |
|---|---|---|---|
| ***News*** | 36.73 dB | 33.39 dB | 34.88 dB |
| ***Foreman*** | 33.63 dB | 29.01 dB | 31.04 dB |

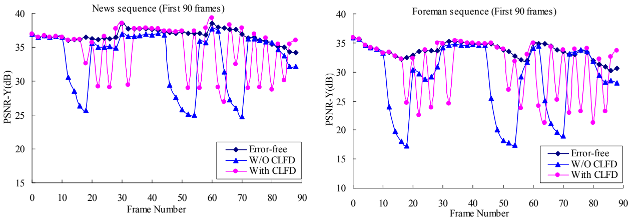


**Fig.4. Frame-by-frame PSNR-Y comparison of different schemes for *News* (left) and *Foreman* (right) sequences**

The frame-by-frame PSNR-Y comparison is shown in Fig.4, where the PSNR-Y is presented once every second frame. We can observe from the figure that: (1) the proposed scheme can spread out the consecutive frame losses; (2) in the proposed scheme, the later pictures in the sequence can get higher quality. The reason is that the selective frame discarding at APP layer results in minimum TDD and reduced encoding bits, and the saved bits are assigned to the following frames by the rate-control-based coding technique.

## 5. CONCLUSIONS

We have proposed a resilient video coding scheme for cellular video coding using cross-layer frame discarding. Simulation results show that the proposed scheme can distribute successive frame losses over some time period to increase rate-distortion performance, and thus significantly improve error resilience of cellular video communication.

## 6. REFERENCES

[1] B. Girod, N. Färber, "Feedback-based error control for mobile video transmission," in *Proc. IEEE,* special issue on video for mobile multimedia, vol. 87, no. 10, pp. 1707-1723, Oct. 1999.

[2] T. Stockhammer and M. M. Hannuksela, "H.264/AVC video for wireless transmission," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 6-13, Aug. 2005.

[3] H. Liu, W.J. Zhang, S.Y. Yu and X.K. Yang, "Channel-aware frame dropping for cellular video streaming," In *proc. Int. Conf. Acoust., Speech, Signal Processing,* Thulouse, France, May. 2006.

[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. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. IEEE Global Telecommunication Conf.*, London, U.K., Nov. 1996.

[6] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Select. Areas Commun.,* Vol. 18, No. 6, June. 2000.

[7] S. Varadarajan, H.Q. Ngo, and J. Srivastava, "Error spreading: a perception-driven approach to handling error in continuous media streaming," *IEEE/ACM trans. on networking,* vol. 10, No.1, Feb. 2002.