

# Perceptual Video Streaming Protection Model by Priority Classification in Compress Domain

Weiwei Xu, Xiang Tian<sup>1</sup>, Yaowu Chen

**Abstract.** Quality degradation happens when Streaming video transmits over error-prone network. In this paper, a perceptual error protection model is proposed by utilizing the priority assignment schedule with FMO and ARQ. The prioritization is based on the property of human visual system and influence on the end-to-end quality. The MB coding mode, motion vector, partition size and bit-consumption are utilized for the derivation of priority from spatial and temporal domain. The local motion property and texture complexity are utilized for the derivation of perceptual saliency. The content-aware selective automatic repeat request mechanism is then realized by integrating the proposed classification method. The performance is examined under the condition of low latency and limited band width redundancy. Results of the experiment show that the proposed method outperforms former ARQ method in the objective and subjective quality of video transmission.

**Keywords:** Error protection. MB classification. Human visual system.

## 1.1 Introduction

Streaming video has largely increased, and the growing market of mobile entertainment and high-quality multimedia content over web site can be expected to continue. Hybrid codec like H.264 is widely used in modern multimedia applications. Motion compensation technique is applied in most video coding schemes to compress the video data into acceptable size for the limited transmission bandwidth. Video transmission over error-prone wired and wireless channel often suffers packet losses and bit errors which degrade the video quality <sup>[1]</sup>. The loss of the slice not only causes the quality to degrade itself, but also allows the error to propagate into the following frames. To overcome these shortcomings, various tech-

---

<sup>1</sup> Xiang Tian (✉)

*Institute of Advanced Digital Technology and Instrument, Zhejiang University, Hangzhou 310027, China*

e-mail: xuweiweimilk@zju.edu.cn

niques have been developed, which includes the error resilience methods on the source and the methods on the channel.

Past error resilience methods on the source utilize the adaptive intra refresh<sup>[2]</sup>, reference picture selection<sup>[3]</sup>, redundant pictures<sup>[4]</sup>, and flexible macro-block ordering (FMO)<sup>[5]</sup> on frame coding picture have been proposed. Past error resilience methods on the channel includes the Forward Error Correction (FEC) mechanism or the Automation Repeat request (ARQ) mechanism. The FEC mechanism protects the video by providing additional redundant packets<sup>[6]</sup>. Parity check codes<sup>[7]</sup> are utilized by the ARQ mechanism for the checking of packets, and the lost packets will be retransmitted to the receiver when the sender fails to receive an acknowledgement from the receiver or receives a retransmission request from the receiver. The ARQ mechanism can protect video transmission by the packet retransmission<sup>[8]</sup> according to channel conditions. The adoption of content-aware error handling and protection mechanisms can be realized by providing importance estimates at the packet level.

Due to the property of video compressing mechanism and video content, especially the spatial-temporal dependencies, some media units have more influence on the quality in case of loss than others. Several techniques have been proposed to determine the expected loss-distortion of individual units or groups of media units. The simplest video data classification is the distinction between I and non-I frame<sup>[9]</sup>. Intra-coded, inter-coded and bidirectional-coded frames can be assigned with different priorities in decreasing order as just listed to perform selective retransmissions<sup>[10]</sup>. But these methods can only be utilized for the simple prioritization assignment, and finer-grained prioritization mechanism can be better for system with adaptive reliability. The expected distortion can be estimated at the pixel level<sup>[11]</sup> under the condition of close knowledge of the loss rate and assumes a Bernoulli distribution. Estimating the overall channel-distortion by simulating the complete decoder behavior for the loss of each packet individually<sup>[12]</sup> can get quite close to the actual distortion derived at the receiver side, but derivation procedure is computationally expensive and is hard to carry out in real time. A lot of researches are devoted to the content-based or region-of-interest (ROI) video coding to improve the visual quality of the moving regions, and that can also be utilized on the prioritization assignment. A preprocessing step is necessary in order to detect the moving regions in the video scenes for these algorithms, which consumes a high computation resource<sup>[13-15]</sup>.

Therefore, designing a low-complexity and effective content-adaptive video protection schedule is a challenging task. A joint analyzing-coding-transmitting framework is proposed in this paper with FMO and ARQ mechanism. The priority assignment and video coding is implemented together, and that greatly decrease the computational complexity of the error resilience method. The coding procedure is performed together with FMO by gathering the MBs with close priority in one slice group.

## 1.2 Prioritization

The lightweight approach in this paper is denoted as texture and motion-tracked prioritization (TMP), and targets on the computational complexity limited system. The proposed TMP mainly utilized the elements in compression domain.

The MBs with large spatial and temporal coordination have better error concealment quality, and then have small influence on the video quality in the decoder side. The priority of MB is related to influence on the end-to-end distortion in the decoder side when it gets lost. To achieve optimized video quality, the MBs with relatively large influence on the end-to-end quality should be labeled with higher priority. The property of the human visual system should also be taken into consideration. The prioritization of current slice can be calculated as the sum of all MBs' prioritization, and the TMP can be demonstrated as:

$$P_{TMP} = \sum_{i=0}^M \left( (1-w) \times P_T + w \times P_M \right) \times P_V \quad (1.1)$$

where  $w$  denotes the weight factor,  $P_T$  denotes the texture priority of MB,  $P_M$  denotes the motion priority of MB and  $P_V$  denotes the attention priority of MB.

### 1.2.1 Derivation of $P_M$

When a slice gets lost in inter-coded frame, the error concealment method in the temporal domain is utilized. The motion vector and reference frame of MBs in the neighbor slices are utilized to find suitable region in the temporally collocated frames for the restoration of the lost slice. Although different error concealment methods have different performance, they share some common property. The areas with global motion can be efficiently concealed from the lost. The areas with local motion have less coordination with MB in the same frame. That motion is more reliable than the other factors is a significant property of human visual system<sup>[13-15]</sup>. Several motion models and estimation methods have been proposed in the literature for global motion estimation<sup>[15-16]</sup>. In<sup>[15-16]</sup>, pixel level background segmentation methods were presented. However the goal of the proposed method is to be computationally efficient for target devices with low processing power. Simplified two-parameter global camera motion model under compression domain is utilized. The motion and texture information analysis of the encoded frame is performed after encoding process is finished. For the macro-blocks with smooth texture or regular edge, the motion estimation procedure is easily disturbed. The macro-blocks with complex texture can be "trusted" for the further analysis. the macro-block mask utilized is formulated as below;

$$T(i, j) = \begin{cases} 1 & \text{if } \left( M(i, j) \neq 16 \times 16 \right. \\ & \left. \text{or } \left( M(i, j) = 16 \times 16 \text{ and } B(i, j) > T_2 \right) \right) \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

where  $M(i, j)$  denotes the macro-block mode,  $B(i, j)$  denotes the bit consumption,  $i$  and  $j$  denote the coordination of macro-block. The area with global motion is abstracted from the trusted area  $T(i, j)$  by analyzing the direction of motion. The MB with global motion can be calculated:

$$G(i, j) = \begin{cases} 1 & \text{if } \left( D(i, j) = t \text{ and } H(t) = \max_{0 < i < N} (H(i)) \right) \\ 0 & \text{otherwise} \end{cases} \quad (1.3)$$

The coherence of local motion areas in successive frames is utilized for the decision of global motion;

$$G = \begin{cases} 2 & \text{if } (N_t > (N_T - N_L)) \\ 1 & \text{if } (N_t < N_L) \\ 0 & \text{otherwise} \end{cases} \quad (1.4)$$

where  $N_t$  denotes the number of  $G(i, j)$ ,  $N_t$  denotes estimated number of local motion macro-block,  $N_L$  denotes the number of  $T(i, j)$ .  $G$  equals 1 means that there is only local motion,  $G$  equals 2 means that there is global motion. Then the motion vector should be compensated for the scene with global motion;

$$\begin{bmatrix} H \\ V \end{bmatrix} = \begin{bmatrix} H_{org} \\ V_{org} \end{bmatrix} - \begin{bmatrix} H_G \\ V_G \end{bmatrix} \quad (1.5)$$

The magnitude of the motion vector difference can be utilized for the motion prioritization:

$$(1.6)$$

The error propagation is related to the reference frame list. For the different reference frames, the property of reference region should be taken into consideration. All the frames in the reference list are ranked and different weight are given for the area with global motion and local motion. The total motion priority can be derived utilizing motion track;

$$P_T = \sum_{j=0}^R \sum_{i=0}^S P_T(i, j) \quad (1.7)$$

Where  $R$  and  $S$  denotes the number of utilized reference frame and reference region in the reference frame.

### 1.2.2 Derivation of $P_T$

MBs is divided into 16x16, 16x8, 8x16, and 8x8 partitions, in which the latter is subdivided into four shaped sub-partitions. During the encoding procedure, the encoder decides which partition size to utilize based on the spatial and temporal complexity of the source frames. For Intra MBs, the spatial interpolation is utilized for the spatial prediction. For inter MB, the prediction within reference frame candidates will be decided, and the balance between quality and bit consumption should be made for the decision of partition size. The partition size is utilized for the classification of MB in the rate control<sup>[15]</sup>. When the MB with small partition size gets lost, the error concealment method cannot well restore it. The partition size and the bit consumption can also be utilized for the derivation of priority:

$$P_T(i) = \frac{\alpha \times B(i) \times M(i)}{\frac{1}{N} \sum_{j=1}^N B(j)} \quad (1.8)$$

Where B denotes the bit consumption of MB, M denotes the modes weight of the MB, and  $\alpha$  denotes the normalized factor to limit the max priority to 1.

### 1.2.3 Derivation of $P_V$

For the derivation of the attention priority, the motion activity should be first abstracted. According to the former researches<sup>[13-15]</sup>, the MB mode should also be taken into consideration. Due to the nature of animals, the motion takes the most important role in the detection of attention area. Utilizing the previous local motion derivation procedure, the attention priority can be formulated as;

$$P_V = \begin{cases} 1 & \text{if } \left( \sqrt{H^2 + V^2} > T \text{ and referred } P_V = 1 \right) \\ 0 & \text{otherwise} \end{cases} \quad (1.9)$$

## 1.3 Performance Evaluation

### 1.3.1 Simulation Setup

For encoding all test sequences, the bit-streams are encoded utilizing the JM references software. Due to the requirement of real time system, the bit-streams are en-

coded before the testing model, and the parameter like motion vector is also save to file. Only the compressed bit-streams and the relative information files are utilized in the test. All test sequences considered in experiments are commonly used in scientific literature and are publicly available. These sequences are encoded utilizing closed-GOP size of 16 frames. All the block sizes from  $4 \times 4$  to  $16 \times 16$  are allowed for the motion estimation/compensation. In the simulation, multi-slice mode is used with random IntraMB Refresh set to 0. The proposed method is designed for the real-time application, so the latency is limited to below 200ms under the frame rate of 30. The ARQ method in [7] is utilized as comparison for the proposed ARQ method.

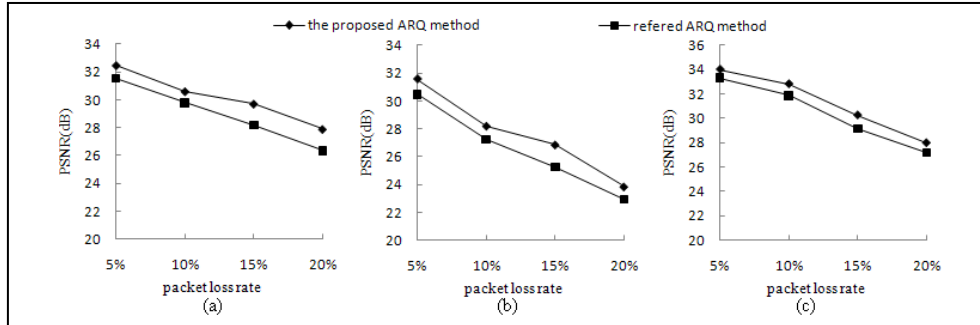
For decoding and concealment, the JM references software is also utilized. Again due to the requirement of real time application, the bit-streams are not reconstructed after received, and only the parsing process are performed. The final bit-streams are saved to file and send to the JM decoder after communication is over. The decrease in video quality caused by transmission errors is expressed in signal-to-noise ratio (PSNR) and the WPSNR [19]. Due to the length of video sequences, the average PSNR value is regarded as the final result.

The testbed of video transmission system includes a sender node, a receiving node, and a network node. The packet loss and delay of the network can be simulated by replacing the network node with software network simulator to conduct quick of-line experiments. The available bandwidth is calculated as the bit-rate of bit-stream plus an additional 5%-10% of redundancy. For the condition with more redundancy and delay, many other error resilience methods can achieve better video transmission quality. From experiment results, it turned out that the results obtained by using software simulator are similar to those coming from the wired testbed. For simplification, the duration between a packet being sent and sender receiving response (RTT) is set as  $4 \times 33$ ms. RTO is estimated by SRTT and RTTV AR, SRTT is the exponentially-weighted moving average (EWMA) of the past RTT samples, RTTV AR is smoothed RTT standard deviation estimator [18]. The RTO is set as  $4 \times 33 + 4 \times 5$ ms. Although the ARQ is utilized for the loss slice, there are still packets getting lost, due to the timing constraints and bandwidth limit. Under most situations, the lost packet can be restored by retransmission. Under certain situation, there are too many packets that need to be retransmitted that some of them get too late for second transmission.

### ***1.3.1 Simulation results***

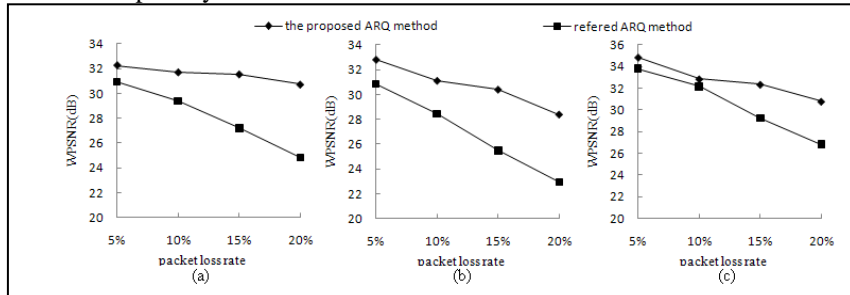
The experiments are performed under the condition of independent loss utilizing the coastguard, Edberg, and foreman sequences. The Bernoulli model at loss rates of 5%, 10%, 15% and 20% are utilized. The bitrates of these sequences are set as 900Kbps, 1200Kbps and 500Kbps. The experimental results are illustrated in Fig.1.1. The proposed ARQ method outperforms the referred ARQ method 1.07dB in average PSNR. For these sequences, the proposed method improves the precision of ranking slices, and then improves the end-to-end quality of video

transmission.



**Fig. 1.1** The experiment PSNR results of the coastguard (a), Edberg (b), and foreman (c).

The experiments are also performed to evaluate the subjective visual quality utilizing WPSNR<sup>[19]</sup>. The experimental results are illustrated in Fig.1.2. Significant improvement can be achieved by using the proposed method, due to the utilizing of attention priority.



**Fig. 1.2** The experiment WPSNR results of the coastguard (a), Edberg (b), and foreman (c).

## 1.4 Conclusion

In this paper, we proposed a real-time-capable technique to prioritize video packets according to their impact on the end-to-end quality. The MB coding mode, motion vector, partition size and bit-consumption are utilized for the derivation of priority from spatial and temporal domain. The property of the human visual system is also taken into consideration. Utilizing this priority assignment schedule, the adaptive send schedule in the ARQ improves the total video transmission performance under error-prone channel. Significant performance improvements could be observed in terms of PSNR. Under the condition of long delay, the FEC can be utilized together as the HARQ in the future research. The priority assignment method can be utilized for the assign of redundancy for the FEC.

## 1.5 Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 40927001), the project of Key Scientific and Technological Innovation Team of Zhejiang Province, (Grant No. 2011R09021-06), the Fundamental Research Funds for the Central Universities.

## 1.6 References

1. Tsai M, Chilamkurti N, Park J et al (2010) Multi-path transmission control scheme combining bandwidth aggregation and packet scheduling for real-time streaming in multi-path environment. *IET Commun.* 4(8): 937–945
2. Shu H Y, Chau L P (2008) Macroblock Mode Decision for Error Resilience Transcoding. *IEEE Transactions on Multimedia.* 10(1):97–104
3. Psannis K E, Ishibashi, Y (2009) Efficient error resilient algorithm for H.264/AVC: mobility management in wireless video streaming. *Telecommun. Syst.*, 41(2):65–76.
4. Zhu C B, Wang Y K, Miska M et al (2009) Error Resilient Video Coding Using Redundant Pictures. *IEEE Transactions on Circuits and System for Video Technology.* 19(1):3–14
5. Shih J Y, Tsai W J (2010) A New Unequal Error Protection Scheme Based on FMO. *Multimedia Tools and Applications.* 47(3):461–476
6. Argyriou A (2008) Cross-layer error control for multimedia streaming in wireless/wireline packet networks. *IEEE Trans. Multimedia.* 10(6):1121–1127
7. Han R, Messerschmitt D (1999) A progressively reliable transport protocol for interactive wireless multimedia. *Multimed. Syst.* 7(2):141–156
8. Ding J, Tseng S, Huang Y (2003) Packet permutation: a robust transmission technique for continuous media streaming over the Internet. *Multimed. Tools Appl.* 21(3):281–305
9. Chang H S, Chang R I, Ho M J et al (2007) A priority selected cache algorithm for video relay in streaming applications. *IEEE Trans. Broadcast.* 53: 79–91
10. Zhang W, Zheng Q, Lian Y (2009) Tree-aware selective frame discard for P2P IPTV system on set-top boxes. *IEEE Trans. Consum. Electron.*, 55: 1982–1987
11. Shen Y, Cosman P C, Milstein B (2006) Video coding with fixed-length packetization for a tandem channel. *IEEE Trans. Image Process.*, 15: 273–288
12. Baldi M, De Martin J C, Masala E et al (2008) Quality-oriented video transmission with pipeline forwarding. *IEEE Trans. Broadcast.*, 54:542–556
13. Song H, Kuo C C J (2004) A region-based H.263+ codec and its rate control for low VBR video. *IEEE Trans. Multimedia.* 6(3):489–500
14. Lin H C, Wu J L (1997) Content-based rate control scheme for very low bit-rate video coding. *IEEE Trans. Consumer Electronics.* 43(2):123–133
15. Chen J, Zheng J, He Y (2007) Macroblock-level adaptive frequency weighting for perceptual video coding. *IEEE Trans. Consumer Elec.*, 53(2):775–781
16. Liu S, Kim J, Kuo C J (2003) Hybrid global/local motion-compensated frame interpolation for low bitrate video. *J. Vis. Commun. Image Represent.*, 14: 61–79
17. Su Y, Sun M, Hsu V (2005) Global motion estimation from coarsely sampled motion vector field and the applications. *IEEE Trans. Circuits Syst. Video Technol.*, 15(2):232–242
18. Paxson V, Allman M (2000) RFC 2988:Computing TCP's Retransmission
19. Song H, Kuo C C J (2004) A region-based H.263+ codec and its rate control for low VBR video. *IEEE Trans. Multimedia.* 6(3):489–500