A RATE-DISTORTION OPTIMIZED ERROR CONTROL SCHEME FOR SCALABLE VIDEO STREAMING OVER THE INTERNET

Fan Zhai, Randall Berry, Thrasyvoulos N. Pappas, and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
{fzhai, rberry, pappas, aggk}@ece.northwestern.edu

ABSTRACT

Video streaming over the Internet is a challenging task due, in part to the wide range of bandwidth variations caused by network congestion. To deal with this challenge, we propose an optimal error control scheme for scalable video transmission over the Internet. The three major components of error control--error resilience, forward error correction (FEC), and error concealment--are considered in the proposed framework. Rate-distortion (R-D) optimization is carried out to determine the encoding mode for each packet and the channel coding rates, in order to minimize the overall expected end-to-end distortion. Our simulation study demonstrates that the proposed approach is robust to the wide range channel bandwidth variations and greatly outperforms the classical R-D optimization scheme.

1. INTRODUCTION

Streaming video applications, such as video on demand, videophone and videoconferencing, are becoming increasingly popular over the Internet. However, because of its best-effort design, the Internet suffers from packet loss and a wide range of capacity variations, caused by network congestion. Most frameworks for transporting streaming video over the Internet consist of two basic components, congestion control and error control. Specifically, congestion control includes rate control, rate-adaptive encoding, and rate shaping; error control includes error resilience, FEC, retransmission, and error concealment [1]. In this paper, we focus on error control techniques based on SNR scalable video.

To overcome a lossy packet channel, source encoding should be tailored to adapt to the channel errors. Error resilience schemes deal with packet loss at the source coding layer. Error resilience is usually composed of resynchronization marking, data partitioning and reversible variable-length coding (RVLC) for wireless video. For Internet video, error resilience usually turns out to be the optimal encoding mode selection for each packet, since different prediction modes result in different levels of coding efficiency and robustness. However, adaptation at the source cannot always overcome the large variations in packet loss and is also limited by the delay in the feedback and the low level of accuracy in estimating the bottleneck bandwidth. Therefore, FEC and/or retransmission are used to further protect packet losses. Conventional retransmission-based schemes such as automatic repeat request (ARQ) are not considered in this paper, since it may exceed timing requirements for streaming applications, especially for real-time applications. Error concealment refers to post-processing technique employed by the decoder [1].

SNR scalable video, where different parts of an encoded stream have unequal contributions to the overall quality, is a separate tool to overcome a wide range of capacity variations. When this property can be exploited in transmission by network mechanisms that use the available bandwidth to provide unequal error protection (UEP) for different parts with different importance, scalable video can maximize the perceived quality [2].

Recently, several studies have been carried out on error control for scalable video streaming over wireless channel or the Internet. In [3], the authors studied the impact of packet size within the framework of channel coding optimization for scalable video. The framework does not apply to the macroblock-level, and optimal mode selection is not considered therein. Kondi *et al*. [4] studied this problem in the context of wireless channels. Horn *et al*. [5] introduced the combination of scalable video coding and UEP to combat Internet packet losses. Neither of them, however, was macroblock-based. Rose *et al*. [6] considered optimal mode selection, but FEC was not part of the framework, while the focus was to study the "drift" management problem, which is not standard compliant.

In this paper, we propose an optimal error control scheme for scalable video streaming over the Internet, which covers optimal macroblock mode selection (prediction mode and quantizer selection), channel coding using Reed-Solomon (RS) codes, and error concealment for both base layer (BL) and enhancement layer (EL). Section 2 describes the problem formulation, and Section 3 describes the implementation details. The algorithm used to solve the optimization problem is laid out next, followed by the simulation results and discussion. Conclusions are drawn at the end.

2. PROBLEM FORMULATION

Classical R-D optimization considers only optimized mode selection at the source, as shown below:

$$\min_{\{\mu^k\}} D = \sum_{k=1}^{M} d^k(\mu^k) \qquad s.t. \ \sum_{k=1}^{M} R_s^k \le R_0 \,, \tag{1}$$

where d^k and R_s^k are the quantization distortion and the source bit rate, respectively, for the k-th packet with a particular mode

μ^k (prediction mode and quantizer). D and R_0 are the overall quantization distortion and bit budget for one frame respectively, and M denotes the number of packet in one frame. R_0 is usually obtained from a higher-level rate controller.

In the error prone channel, however, to achieve a good performance, a global optimization is required that takes into consideration both the path characteristics and the receiver behavior, in addition to the source behavior [1,5,9]. To account for this, we consider the minimization of the expected end-to-end distortion instead of just the distortion calculated by considering only quantization errors. Specifically we consider

$$\min_{\{\mu^k, R_c^k\}} E[D] = \sum_{k=1}^{M} E[d^k(\mu^k, R_c^k)] \qquad s.t. \ \sum_{k=1}^{M} (R_s^k / R_c^k) \le R_0 , \quad (2)$$

where $E[\cdot]$ is the expectation operator, which takes into consideration the packet loss and error concealment at the receiver, and R_c^k is channel rate, for the k-th packet. The source rates R_s^k are in bits per second (bps) while the channel rates R_c^k are the number of information bits per channel bit. With more bits allocated to the source, the coding efficiency becomes higher, but the bitstream is more likely to be corrupt. Therefore, the channel coding plays the role of trading-off the coding efficiency and the robustness of the bitstream. The goal of the optimization is to minimize the expected end-to-end distortion given the bit rate constraint. In terms of scalable video, with an additional bit rate constraint for the base layer (BL), the optimization problem of (2) can be solved on BL and enhancement layer (EL) sequentially, according to

$$\min_{\{\mu_b^k, R_{b,c}\}} E[D_b] = \sum_{k=1}^{M} E[d_b^k(\mu_b^k, R_{b,c})] \qquad s.t. \ \sum_{k=1}^{M} (R_{b,s}^k / R_{b,c}) \le R_{b,0} \quad (3)$$

$$\min_{\{\mu_e^k, R_{e,c}\}} E[D_e] = \sum_{k=1}^{M} E[d_e^k(\mu_e^k, R_{e,c})] \qquad \text{s.}t. \ \sum_{k=1}^{M} (R_{e,s}^k / R_{e,c}) \le R_{e,0} \quad (4)$$

where $R_{b,s}^k$ and $R_{e,s}^k$ are the source rates for the k-th packet, $R_{b,c}$ and $R_{e,c}$ are the channel rates, $R_{b,0}$ and $R_{e,0}$ ($R_0 = R_{b,0} + R_{e,0}$) are the bit rate constraints. Subscripts "*b*" and "*e*" denote BL and EL, respectively.

3. SYSTEM DETAILS

3.1. Packetization and Channel Model

For simplicity, the packet size is assumed to be one macroblock (MB) and every packet is independently decodable. However, the proposed framework can be easily extended to apply to other packetization strategies. The channel is modeled by Bernoulli process as a packet erasure channel, i.e., each packet for the BL (EL) is independently lost with probability ρ_b (ρ_e). When UEP is employed, it is reasonable to assume that ρ_b is always less than ρ_e. UEP can be realized by using priority channel coding [4,5], or by priority packet dropping schemes implemented in the routers, as in the differentiated services (DiffServ) [7]. If an idealized DiffServ is employed, assuming the channel capacity is C, the packet loss probabilities can be calculated as $\rho_b = max\{0, \ 1\text{-}C/R_b\}$, $\rho_e = max\{0, \ min\{1, 1\text{-}(C\text{-}R_b)/R_e\}\}$. That is, if C is greater than R_b, then the loss rate of BL, ρ_b is equal to zero, and packet loss only occurs in EL. We will use this channel model in the first experiment, which will be described in Section 4 together with experiment 2. In the second experiment, channel coding will be used to perform error protection. No Diffserv is employed in the second experiment.

3.2. FEC

For Internet applications, many researchers have considered using erasure codes to recover from packet losses. Specifically, a video stream is first chopped into segments, each of which is packetized into k packets; then for each segment, a block code is applied to the k packets to generate an n-packet block, where $n > k$. It allows the network and receivers to discard some of the packets that cannot be handled due to limited bandwidth or processing power. Here, we consider using Reed-Solomon (RS) code to perform channel coding. An RS code is represented as RS (n, k), where k is the length of source symbols and n-k is the length of parity symbols. An RS (n, k) code can correct up to n-k symbol erasures if symbol positions are known, regardless of which symbols are lost. The code rate of RS (n, k) is k/n.

Since RS codes are systematic codes, we say that a packet is lost after error recovery at the receiver only when the packet is lost and the block containing the lost packet cannot be recovered. Therefore, the probability of packet loss ρ after error recovery is defined as

$$\rho = \varepsilon[1 - \sum_{j=0}^{n-k-1} \binom{n-1}{j} \varepsilon^j (1-\varepsilon)^{n-1-j}] , \quad (5)$$

where ε is the probability of packet loss before error recovery. Generally speaking, the protection capability of the RS code depends on the block size and code rate. For Internet applications, the target number of video packets, n, can be determined according to the end-to-end system delay constraints. Since packet sizes (one MB per packet) in our framework are different, the maximum packet size of a block is first determined, and then all packets are padded with stuffing bits in the tail parts to make the size equal. The stuffing bits are removed after the parity codes are generated and are not transmitted [8].

3.3. Error Concealment

A simple but efficient error concealment scheme is used in this paper. In the BL, when a packet is lost, the corrupted packet is replaced with the MB from the BL of the previous frame that is pointed to by the motion vector of the previous packet. If the previous packet is also lost, the zero motion vector is used to perform concealment. When a packet in EL is lost, the decoder uses the "upward" concealment to replace the lost packet by the corresponding MB in the BL of the temporally simultaneous frame, which may either received or has been concealed.

3.4. Recursive Distortion Measurement

The distortion measurement is based on an algorithm called ROPE (Recursive Optimal Per-pixel Estimate), which ensures accurate estimation of the overall end-to-end distortion [9]. Assuming the mean squared error (MSE) criterion, the overall expected distortion levels of pixel i in frame n, at the BL and EL are respectively given by

$$E[d_n^i(b)] = E[(f_n^i - \tilde{f}_n^i(b))^2] = (f_n^i)^2 - 2 f_n^i E[\tilde{f}_n^i(b)] + E[\tilde{f}_n^i(b)^2] \quad (6)$$

$$E[d_n^i(e)] = E[(f_n^i - \tilde{f}_n^i(e))^2] = (f_n^i)^2 - 2 f_n^i E[\tilde{f}_n^i(e)] + E[\tilde{f}_n^i(e)^2] . \quad (7)$$

The parameters used here are defined as follows:

f_n^i : *i-th* pixel of the *n-th* original frame

$\tilde{f}_n^i(b)$, $\tilde{f}_n^i(e)$: *i-th* pixel of the *n-th* expected decoded frame

$\hat{f}_n^i(b)$, $\hat{f}_n^i(e)$: *i-th* pixel of the *n-th* reconstructed frame

$d_n^i(b)$, $d_n^i(e)$: distortion of the *i-th* pixel of the *n-th* frame

$\hat{e}_n^i(b)$, $\hat{e}_n^i(b)$: *i-th* quantized residue of the *n-th* frame

ρ_b, ρ_e: probability of packet loss with the use of channel coding, where, "*b*" and "*e*" represent BL and EL respectively.

The first and second order expected values of one pixel have different expressions with different modes and different layers, depending on the error concealment method. As an example, using the error concealment method described above, $E[\tilde{f}_n^i(b)]$ with INTRA and INTER mode would be, respectively, defined as

$$E[\tilde{f}_n^i(b)] = (1-\rho_b)\hat{f}_n^i(b) + \rho_b(1-\rho_b)E[\tilde{f}_{n-1}^k(b)] + \rho_b^2 E[\tilde{f}_{n-1}^i(b)] \quad (8)$$

$$E[\tilde{f}_n^i(b)] = (1-\rho_b)(\hat{e}_n^i(b) + E[\tilde{f}_n^i(b)]) + \rho_b(1-\rho_b)E[\tilde{f}_{n-1}^k(b)] + \rho_b^2 E[\tilde{f}_{n-1}^i(b)] \quad (9)$$

where the superscript k in $E[\tilde{f}_{n-1}^k(b)]$ is the position of the concealment pixel in the previous frame pointed by the concealment motion vector.

4. OPTIMAL CHANNEL RATE AND MODE SELECTION

Since (3) and (4) are of the same form, they are solved in the same way. Here we only discuss one, thus the subscript "*b*" and "*e*" are ignored. The constrained problems of (3) and (4) can be converted into unconstrained ones with the use of a Lagrange multiplier as shown below

$$\min_{\{\mu^k, R_c\}} \sum_{k=1}^{M} J_k(\mu^k, R_c) = \min_{\{\mu^k, R_c\}} \sum_{k=1}^{M} \{E[d^k(\mu^k, R_c)] + \lambda (R_s^k(\mu^k)/R_c)\} . \quad (10)$$

The above unconstrained minimization problem is equivalent to

$$\min_{\{R_c\}} \sum_{k=1}^{M} J_k(\mu^{k*}(R_c)) = \min_{\{R_c\}} \left\{ \min_{\{\mu^k\}} \sum_{k=1}^{M} J_k(\mu^k, R_c) \right\}, \quad (11)$$

which can be solved in two steps: optimal mode selection given the source bit rate constraint, and optimal bit allocation between the source coding and channel coding given the total transmission bit rate constraint.

Given the error concealment strategy chosen, the mode selection for each MB only depends on the encoding of its previous MB. Therefore, dependency is constrained within one row. With K_r and R denoting the number of MBs in a row and the number of rows in a frame respectively, (11) can be decoupled into

$$\min_{\{R_c\}} \left\{ \min_{\{\mu^k\}} \sum_{k=1}^{M} J_k(\mu^k, R_c) \right\} = \min_{\{R_c\}} \left\{ \sum_{r=1}^{R} \{ \min_{\{\mu^k\}} \sum_{k=1}^{K_r} J_k(\mu^k, R_c) \} \right\}, \quad (12)$$

where minimization is independently performed within each row. This relaxed problem can be solved using techniques from Dynamic Programming (DP). The optimization problem of (3) and (4) can be solved through solving (12) by choosing an appropriate λ, which can be carried out by the bisection search or a fast convex search algorithm [10].

5. SIMULATION RESULTS AND DISCUSSION

The simulation is based on the H.263+ SNR scalable codec (Annex *O* supports SNR scalability and Annex *K* supports slice structure, which is used for packetization) [11]. The test sequence is Foreman with QCIF (176×144) format and frame rate 30 fps. The channel transmission rate is 360Kbps (which should not be confused with channel capacity—the theoretical maximum transmission rate at which information passes error free over the channel; channel transmission rate is obtained based on the estimated channel capacity), with 180Kbps for BL and 180Kbps for EL. Since we did not incorporate rate control in the proposed framework, the bit budget is 12000 bits per frame.

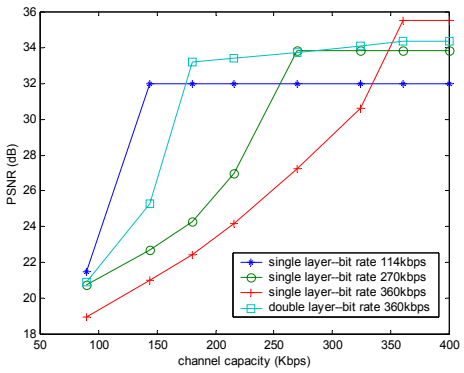

Fig. 1: Scalable vs. non-scalable video (double layer video is tuned to the estimated rate of 270Kbps)

Two experiments have been carried out. The first one, as shown in Fig. 1, is to compare the performance of double-layer to single-layer video delivery in the channel with wide range channel capacity variations. For simplicity, this experiment is based on the assumption of employing an idealized DiffServ to perform UEP for BL and EL, as discussed in Section 3. No channel coding is used in experiment one.

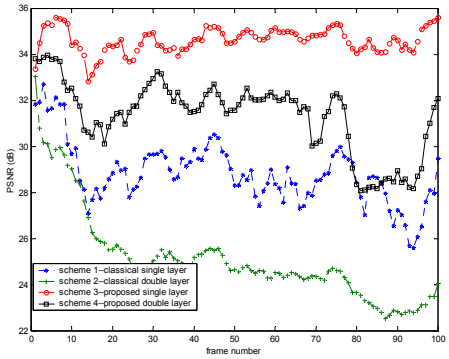

Fig.2: One realization of the four schemes (transmission rate: 360Kbps; channel capacity: 306Kbps)

In Fig. 1, the horizontal axis indicates the channel capacity, while the vertical axis indicates the video quality received by the receiver in terms of PSNR. For double layer video, we calculate the overall PSNR. The double layer video is optimized to the estimated channel capacity, which is 270Kbps in the experiment, using the proposed framework. Single layer cases are encoded at different rates, as shown in the figure. Each curve corresponds to one realization of one encoded bitstream at different channel capacities. The sharp dropping appears when the channel capacity is lower than the generated source bit rate, which corresponds to over estimation. It is clear that double layer video usually degrades more gracefully than single layer video with a

wide range of channel capacities, due to its flexibility to allow bit rate allocation to BL and EL and perform UEP.

The second experiment is to calculate the R-D bound of the proposed scheme, which is obtained based on the assumption that the encoder has accurate estimation of channel capacity. In order to study the efficiency of FEC, no DiffServ is used here, thus the BL and EL have the same probability of packet loss before error recovery. To illustrate the effectiveness of the proposed scheme, four schemes are compared: 1) classical optimized non-scalable scheme (without taking into account the channel error and without using channel coding); 2) classical optimized double-layer scheme; 3) proposed optimized scheme applied on single-layer video; and 4) proposed optimized scheme applied on double-layer video.

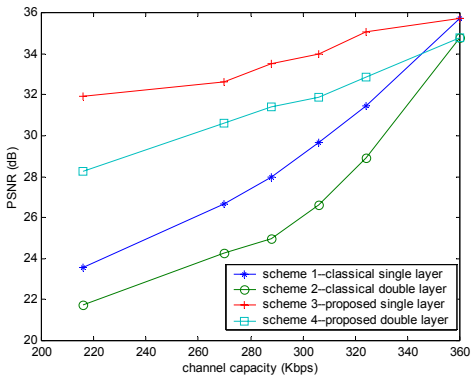

Fig.3: R-D bounds of the four schemes (transmission rate is 360Kbps, with 180Kbps for BL and EL respectively for double layer video)

Figure 2 shows a realization of the above four schemes in terms of qualify versus frames, where the transmission rate is 360Kbps, and the channel capacity 306Kbps. Figure 3 depicts the R-D bounds of the four schemes. It can be seen from Fig. 2 that scheme 3 and 4 outperform schemes 1 and 2 by 0-8.5 dB and 0-6.5 dB, respectively. Scheme 3 has higher R-D bounds than that of scheme 4. This makes sense because when the encoder can be tailored accurately to the channel, non-scalable methods can achieve better performance than scalable ones due to the redundancy for layered approaches at the source coding and the overhead of packet headers. However, this does not mean that non-scalable methods are superior to scalable ones, because when the encoder cannot be tailored to the channel accurately, the scalable method is more robust to a wide range channel variations, as shown in Fig. 1. Another observation from this experiment is that, although we did not explicitly use UEP for BL and EL and we optimized BL and EL sequentially, the optimization automatically results in such UEP. As shown in Table 1, the protection ratio for BL is always higher than that of EL. It is achieved by using error concealment at the decoder (the error concealment is known at the encoder), which makes EL more robust to the packet loss than BL. This is because if a packet in EL is lost, it can be concealed from the BL of the same frame, while if a packet in BL is lost, it can only get concealment from the previous frame. In addition, as the channel gets worse, the encoder turns out to allocate more resources to protect the bitstream, which makes sense since in this case, transporting the bitstream to the decoder is more important than the coding efficiency.

6. CONCLUSION

This paper proposes an R-D optimized error control scheme for scalable video streaming over the Internet based on H.263+ SNR scalable codec. The proposed scheme is robust to the wide range channel capacity variations in the Internet. It is achieved by jointly considering the three major components of error control: error resilience, channel coding and error concealment. By jointly optimizing the channel rate and packet mode, the optimization automatically results in UEP for BL and EL, giving more protection to the most important parts of the bitstream and therefore achieves the maximum video quality received.

Channel capacity	216 Kbps	270 Kbps	288 Kbps	306 Kbps	324 Kbps	360 Kbps
protection ratio, BL	0.50	0.33	0.28	0.21	0.15	0
protection ratio, EL	0.04	0.07	0.05	0.04	0.02	0

Table 1. Protection ratio for scheme 4 (transmission rate: 360Kbps, with 180Kbps for BL and EL respectively)

7. REFERENCES

[1] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: challenges and approaches," *Proc. IEEE*, vol. 88, pp. 1855-1877, Dec. 2000.

[2] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301-317, March 2001.

[3] B. Hong and A. Nosratinia, "Rate-constrained scalable video transmission over the Internet," in *Proc. Packet Video Workshop*, Pittsburg, PA, 2002.

[4] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint scource-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Processing*, vol. 11, pp. 1043-1052, Sept. 2002.

[5] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *IEEE Trans. Image Processing*, vol. 15, pp. 77-94, Sept. 1999.

[6] H. Yang, R. Zhang, and K. Rose, "Drift management and adaptive bit rate allocation in scalable video coding," in *Proc. IEEE ICIP*, Rochester, New York, Sept. 2002.

[7] A. Markopoulou, and S. Han, "Transmitting scalable video over a DiffServ network," Final Project, Stanford Univ., 2001.

[8] X. Yang, C. Zhu, Z. Li, G. Feng, S. Wu, and N. Ling, "Unequal error protection for motion compensated video streaming over the Internet," in *Proc. IEEE ICIP*, Rochester, New York, Sept. 2002.

[9] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 966-976, June 2000.

[10] G. M. Schuster and A. K. Katsaggelos, "Rate-distortion based video compression: optimal video frame compression and object boundary encoding," *Kluwer Academic Publishers*, 1997.

[11] ITU Telecom. Standardization Sector of ITU, Video coding for low bitrate communication, *Draft ITU-T Recommendation H.263 Version 2*, Sept. 1997.