FINE-GRAINED RATE SHAPING FOR VIDEO STREAMING OVER WIRELESS NETWORKS*

Trista Pei-chun Chen and Tsuhan Chen+
Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Email: {peichun, tsuhan}@andrew.cmu.edu

ABSTRACT

Streaming of precoded video, which is both source and channel coded, over wireless networks faces challenges of time-varying packet loss rate and fluctuating bandwidth. Rate shaping has been proposed in [1] to "shape" the precoded video to adapt to the real-time bandwidth requirement and the packet loss rate. In this paper, we propose a novel "fine-grained rate shaping (FGRS)" scheme to allow for bandwidth adaptation over a wide range of bandwidth and packet loss rate. The video is precoded with fine granularity scalability (FGS) followed by forward error correction (FEC) coding with erasure codes. Utilizing the fine granularity property of FGS and FEC, FGRS selectively drops part of the precoded video and still yields decodable bitstream at the decoder. A new two-stage rate-distortion (R-D) optimization, with model-based hyper-plane and hill-climbing based refinement, is proposed to select part of the precoded video to drop. Promising results of FGRS are shown.

1. INTRODUCTION

Due to the rapid growth of wireless communications, video over wireless networks has gained a lot of attention. Challenges as to cope with the time-varying error rate and fluctuating bandwidth bring out the need of error resilient video transport.

Joint source-channel coding techniques [2][3] are often applied to achieve error resilient video transport with online coding. Given the bandwidth requirement, the joint source-channel coder seeks the best allocation of bits for the source and channel coders. However, joint source-channel coding techniques are not suitable for streaming *precoded* video. The precoded video is both source and channel encoded prior to transmission. The network conditions are not known at the time of coding. We proposed to use *rate shaping* [1] to solve the problem. Given the source and channel encoded video, the rate shaper "shapes" the video, that is, reduces the bit rate by dropping part of the precoded video, according to the current network conditions.

To extend the capability of rate shaping proposed in [1], we adopt MPEG-4 fine granularity scalability (FGS) [4] for source coding, and the same erasure codes [5][6] as in [1] for forward error correction (FEC) coding. Unlike conventional scalability techniques such as SNR scalability used in H. 263 [7] and MPEG-2 [8], MPEG-4 FGS provides the video bitstream that is partially decodable over a wide range of bit rates. The more bits of the FGS bitstream is received, the better the video quality is. In addition, it has been known that partial FEC coded bitstream is still decodable within the error correction capability if erasure codes are used. Thus, both FGS and erasure codes provide fine-granularity properties in video quality and in packet loss protection.

With the FGS coded and FEC coded bitstream, "fine-grained rate shaping (FGRS)" is proposed to perform the bandwidth adaptation considering the current packet loss rate of the network. There are conceptually infinitely many possible combinations of dropping portion of the FGS bitstream and portion of the FEC codes. FGRS seeks the optimal solution in the rate-distortion (R-D) manner. A new two-stage rate-distortion (R-D) optimization is proposed for FGRS to select part of the precoded video to drop.

The proposed "two-stage R-D optimization" aims for both efficiency and optimality by using model-based hyper-plane and hill-climbing based refinement. In Stage 1, a model-based hyper-plane is first trained with a set of rate and distortion decrease pairs. We then find the solution in which the hyper-plane intersects the bandwidth constraint. In Stage 2, the near-optimal solution from Stage 1 (because of the model imperfection) is then refined with the hill-climbing based approach. We can see that Stage 1 aims to find the optimal solution globally with the model-based hyper-plane and Stage 2 refines the solution locally.

The paper is organized as follows. In Section 2, we introduce background materials. In Section 3, FGRS is proposed for streaming of the FEC coded FGS bitstream. R-D optimization problem is formulated. In Section 4, two-stage R-D optimization is illustrated. In Section 5, experiments are carried out to show the superior performance of the proposed FGRS. Concluding remarks are given in Section 6.

2. BACKGROUND

We first describe the source coding and channel coding in our system, as well as the wireless network characteristics. We adopt MPEG-4 fine granularity scalability (FGS) [4] for source coding and erasure codes as Reed-Solomon codes [5] for channel coding. The wireless network has time-varying packet loss rate and fluctuating bandwidth.

2.1. MPEG-4 Fine Granularity Scalability (FGS)

FGS has been proposed to provide bitstreams that are still decodable when truncated. That is, FGS enhancement layer bitstream is decodable at any bit rate over a wide range of values. With such a property, FGS was adopted by MPEG-4 for

* Work supported in part by Industrial Technology Research Institute.
+ Corresponding author: Prof. Tsuhan Chen

streaming applications [4]. Through FGS encoding, two layers of bitstream are generated: one base layer and one enhancement layer (Figure 1). The base layer is predictive coded while the enhancement layer only uses the corresponding base layer as the reference.

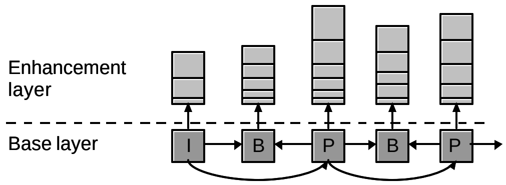

Figure 1. Dependency graph of the FGS base layer and enhancement layer

2.2. Reed-Solomon Codes

Erasure codes provide "fine-grained" packet loss protection with more and more symbols[1] received at the FEC decoder [5]. The "shaped" erasure code is still decodable if the number of erasures/losses from the transmission is no more than $d_{\min} - 1 - (\text{number of unsent symbols})$. An erasure code can successfully decode the message with the number of erasures up to $d_{\min} - 1$, considering both the unsent symbols and the losses taken place in the transmission. Therefore, the more symbols are sent, the better the sent bitstream can cope with the losses.

In this paper, we adopt Reed-Solomon codes as the erasure codes. In Reed-Solomon codes, $d_{\min} - 1$ equals $n - k$, where k is the message size in symbols and n is the code size in symbols. Thus, the partial code of size $r \leq n$ is still decodable if the number of losses from the transmission is no more than $r - k$.

2.3. Wireless Network Characteristics

Wireless networks are generally with time-varying packet loss rate and fluctuating bandwidth. The packet loss rate and bandwidth vary at each frame interval. We simulate random bandwidth fluctuation and use a two-state Markov-chain [9] to simulate the bursty bit errors.

The coded bitstream is transmitted in packets. Let us examine how the packet loss rate e_p relates to the transition probability p and the bit error rate e_b. First, we start without considering any header or extra information added to each packet. With the bit error rate e_b, the transition probability p, and the packet size s, we can derive and get the packet loss rate of the s-bit packet as,

$$e_p = 1 - (1 - e_b)(1 - p)^{s-1} \tag{1}$$

We observe two properties from (1) given the same bit error rate e_b: (i) the smaller the transition probability p, the smaller the packet loss rate e_p, and (ii) the smaller the packet size s, the smaller the packet loss rate e_p given the same e_b and p. However, to detect the loss of packets, some information as the packet number has to be added to each packet. The smaller the packet is, the heavier the overhead is. Therefore, it is a trade-off between the selection of the packet size and the resulting packet loss rate. We use $s = 280$ (bits) in this paper. Users can select the packet size s according to real system consideration using (1).

3. FINE-GRAINED RATE SHAPING (FGRS)

There are three stages for transmitting the video from the sender to the receiver: (1) precoding, (2) streaming with rate shaping, and (3) decoding, as shown in Figure 2 to Figure 4.

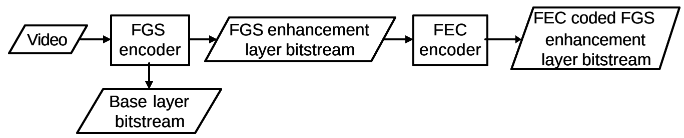

Figure 2 System diagram of the precoding process: FGS encoding followed by FEC encoding

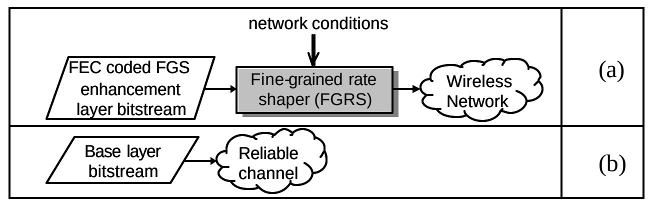

Figure 3. Transport of the precoded bitstreams: (a) transport of the FEC coded FGS enhancement layer bitstream with FGRS via the wireless network, and (b) transport of the base layer bitstream via the reliable channel

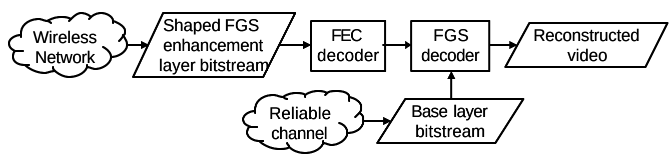

Figure 4. System diagram of the decoding process: FEC decoding followed by FGS decoding

Through FGS encoding, two layers of bitstream are generated: one base layer and one enhancement layer (Figure 1). We will consider hereafter the bandwidth adaptation and packet loss resilience for the FGS enhancement layer bitstream only assuming that the base layer bitstream is reliably transmitted as shown in Figure 3 (b).

We use a frame as the decision unit. Let us look at the FGS enhancement layer bitstream for a frame. FGS enhancement layer bitstream consists of bits of all the bit-planes for this frame. The most significant bit-plane (MSB plane) is coded before the less significant bit-planes until the least significant bit-plane (LSB plane). In addition, since the data in each bit-plane is variable length coded (VLC), if some part of the bit-plane is corrupted (due to packet losses), the remaining part of the bit-plane becomes un-decodable. The importance of the bits of the enhancement layer decreases from the beginning to the end.

Before appending the parity symbols to the FGS enhancement layer bitstream, we first divide all the symbols (recall that in this paper each symbol consists of 14 bits) for this frame into several *sublayers* (Figure 5 (a)). The way to divide the symbols into sublayers is arbitrary except that the later sublayers are longer than

[1] "Symbols" are used instead of "bits" since the FEC codes use a symbol as the encoding/decoding unit. In this paper, we use 14 bits to form one symbol. The selection of symbol size in bits depends on the user.

the previous ones, $k_1 \geq k_2 \geq \cdots \geq k_h$, since we want to achieve unequal packet loss protection (UPP). A nature way of division is to let Sublayer 1 consist of symbols of the MSB plane, Sublayer 2 consist of symbols of the MSB-1 plane, …, and Sublayer h consist of symbols of the LSB plane. Each sublayer is then FEC encoded with erasure codes to the same length n (Figure 5 (b)). The precoded video is stored and can be used at the time of delivery.

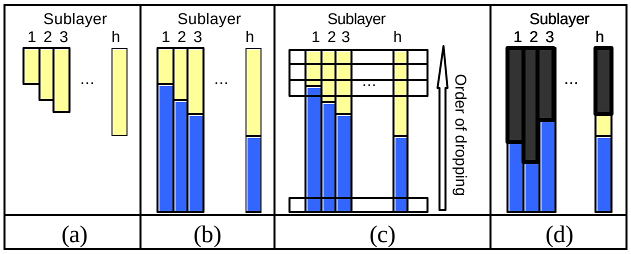

Figure 5. (a) FGS enhancement layer bitstream in sublayers, (b) FEC coded FGS enhancement layer bitstream in sublayers, (c) bandwidth adaptation with conventional approach, (d) bandwidth adaptation with FGRS

At the transport stage, FEC coded FGS bitstream is passed through FGRS for bandwidth adaptation under the current packet loss rate. Note that FGRS is different from joint source-channel coding like approaches, which perform FEC encoding for the FGS bitstream at the time of delivery with a bit allocation scheme that achieves certain objectives, as proposed by van der Schaar and Radha [10] and Yang et al [11].

With the precoded video, bandwidth adaptation can be implemented conventionally by dropping the symbols in the order shown in Figure 5 (c). Given a certain bandwidth requirement for this frame, Sublayer 1 has more parity symbols kept than Sublayer 2 and so on. Shaped bitstream with such a conventional bandwidth adaptation scheme has UPP to the sublayers. However, such scheme might not be optimal. We propose FGRS (Figure 5 (d)) for bandwidth adaptation given the current packet loss rate. The darken bars in Figure 5 (d) are selected to be sent by FGRS.

Let us start from the problem formulation. A FGS enhancement layer bitstream provides better and better video quality as more and more sublayers are correctly decoded. In other words, the total distortion is decreased as more sublayers are correctly decoded. With Sublayer 1 correctly decoded, we reduce the total distortion by DD_1 (accumulated *distortion decrease* is DD_1); with Sublayer 2 correctly decoded, we reduce the total distortion further by DD_2 (accumulated distortion decrease is $DD_1 + DD_2$); and so on. If Sublayer i is corrupted, the later sublayers $i+1$, $i+2$, etc., become un-decodable. The accumulated distortion decrease is illustrated as Figure 6. The dashed arrow indicates the sublayer is not correctly decoded. Note that DD_i of Sublayer i can be calculated given the FGS bitstream. DD_i of Sublayer i is different for every frame.

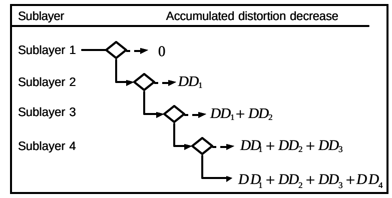

Figure 6. Accumulated distortion decrease

The expected accumulated distortion decrease is then:

$$DD = \sum_{i=1}^{h}\left(DD_i \prod_{j=1}^{i} v_j \right) \quad (2)$$

where v_j is the recovery rate of Sublayer j. Sublayer j is recoverable (or successfully decodable) if the number of erasures is no more than $r_j - k_j$. Thus, the recovery rate v_j is the summation of the probabilities that no loss occur, one erasure occurs, and so on until $r_j - k_j$ erasures occur.

$$v_j = \sum_{l=0}^{r_j - k_j}\left[\binom{r_j}{l} (e_{sym})^l (1-e_{sym})^{r_j - l} \right] \quad (3)$$

where h is the number of sublayers of this frame in total, k_j is the message (the bits from the FGS bitstream) size in Sublayer j, r_j is the number of symbols selected to send in Sublayer j, and e_{sym} is the symbol loss rate. If the packet loss rate is small, the symbol loss rate can be approximated by the packet loss rate divided by the number of symbols per packet $e_{sym} \approx e_p / (s/m)$, where s is the packet size and m is the symbol size in bits. By choosing different combinations of the number of symbols for each sublayer, the expected accumulated distortion decrease will be different. The rate shaping problem can be formulated as follows:

$$\text{maximize} \quad DD = \sum_{i=1}^{h}\left(DD_i \prod_{j=1}^{i} v_j \right)$$

$$\text{subject to} \quad \sum_{i=1}^{h} r_i \leq B \quad (4)$$

To solve the problem, we propose a new two-stage R-D optimization approach. We will elaborate in the next section.

4. TWO-STAGED RATE-DISTORTION OPTIMIZATION

The two-stage R-D optimization first finds the near-optimal solution globally. The near-optimal global solution is then refined by a hill climbing approach.

4.1. Stage 1

We can see from (2) and (3) that DD is related to $\mathbf{r} = [r_1 \quad r_2 \quad \cdots \quad r_h]$ implicitly through the recovery rates $\mathbf{v} = [v_1 \quad v_2 \quad \cdots \quad v_h]$. We can instead find a model-based hyper-plane that explicitly relates $\mathbf{r}$ and DD. The model parameters can be trained from a set of $(\mathbf{r}, DD)$ values. The optimal solution is then the intersection of this hyper-plane and the bandwidth constraint. The complexity of the model determines the preciseness of the model in finding the optimal solution.

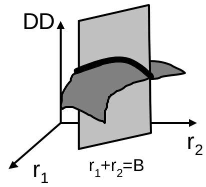

Figure 7. Intersection of the model-based hyper-plane and the bandwidth constraint, illustrated with $h = 2$

4.2. Stage 2

Stage 1 of the two-stage R-D optimization gives a near-optimal solution. The solution can be refined by a hill-climbing based approach (Figure 8). We perturb the solution from Stage 1 to yield a larger accumulated distortion decrease under the bandwidth constraint. The process can be iterated until the solution reaches a stopping criterion such as the convergence.

```
While (stop == false)
    z_i = r_i for all i=1~h
    For (j=1; j<=h; j++)
        For (k=1; k<=h; k++)
            z_k = z_k + delta for k==j //Increase sublayer j
            z_k = z_k - delta/(h-1) for k!=j //Decrease others
        End-for
        Evaluate DD_j by equations (2) and (3)
    End-for
    Find the j* with the largest DD_j*.
    For (i=1; i<=h; i++)
        r_i = r_i + delta for i==j*
        r_i = r_i - delta/(h-1) for i!=j*
    End-for
    Calculate the stop criterion.
End-while
```

Figure 8. Pseudocodes of hill-climbing algorithm

5. EXPERIEMENT

The proposed FGRS is compared with the conventional approach described in Figure 5 (c). The test video sequences are “akiyo”, “foreman”, and “stefan” in common intermediate format (CIF). The frame rate of MPEG-4 FGS coding is three frames/sec.

From the frame-by-frame PSNR performance in Figure 9, we see that the proposed FGRS gives superior results to the conventional approach. Comparing performance with different sequences (Table 1 and Figure 10), the PSNR gain of FGRS over conventional method is the most significant in sequence “akiyo”, followed by sequence “foreman” and “stefan”. Sequence “stefan” is the most challenging one with the most complex scene and the highest motion. Considering the gain in the Y component, FGRS yields 0.76 dB to 1.38 dB gains compared with the conventional approach.

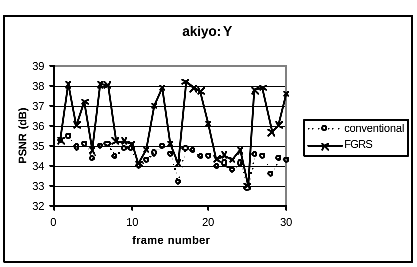

Figure 9. Frame by frame PSNR of conventional approach and FGRS with sequence “akiyo” in Y component

Table 1. PSNR gains in Y, U, and V components with three sequences “akiyo”, “foreman”, and “stefan”

PSNR gain (dB)	Y	U	V
akiyo	1.38	1.28	0.87
foreman	0.86	0.44	0.52
stefan	0.76	0.34	0.38

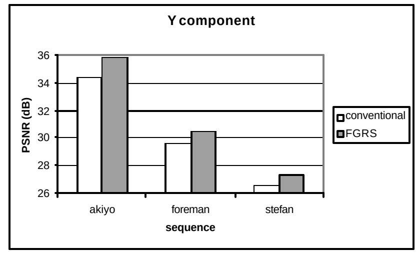

Figure 10. Overall PSNR of conventional approach and FGRS with three sequences in Y components

6. CONCLUSION

We proposed in this paper a novel fine-grained rate shaping (FGRS) approach to perform bandwidth adaptation for a precoded video, which is both FGS coded and FEC coded. A two-stage rate-distortion (R-D) optimization approach is used. The two-stage R-D optimization first obtains the near-optimal solution by finding the intersection of the model-based hyper-plane and the bandwidth constraint. The near-optimal solution is then refined by a hill-climbing based approach. The two-stage R-D optimization aims for both the efficiency and optimality. The proposed FGRS outperforms the conventional approach.

REFERENCES

[1] T. P.-C. Chen and T. Chen, “Adaptive Joint Source-Channel Coding using Rate Shaping”, *ICASSP 2002*, Orlando, FL, U.S.A., May 2002.

[2] G. Cheung and A. Zakhor, “Bit Allocation for Joint Source/Channel Coding of Scalable Video”, *IEEE Transactions on Image Processing*, 9(3), March 2000.

[3] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, **“**Joint Source-Channel Coding for Motion-Compensated DCT-based SNR Scalable Video”, *IEEE Transactions on Image Processing*, 11(9), September 2002.

[4] W. Li, “Overview of Fine Granularity Scalability in MPEG-4 Video Coding” Standard”, *IEEE Transactions on CSVT*, 11(3), pp. 301-317, March 2001.

[5] S. Wicker, Error Control Systems for Digital Communication and Storage, Prentice-Hall, 1995.

[6] L. Rizzo, “Effective Erasure Codes for Reliable Computer Communication Protocols”, *Computer Communication Review*, April 1997.

[7] D. S. Turaga and T. Chen, “Fundamentals of Video Compression: H.263 as an Example”, in Compressed Video over Networks, edited by M.-T. Sun and A. R. Reibman, Marcel Dekker, Inc., 2001.

[8] B. G. Haskell, A. Puri, and A. N. Netravali, Digital Video: An Introduction to MPEG-2, Chapman & Hall 1997.

[9] F. Alajaji and T. Fuja, “A Communication Channel Modeled on Contagion”, *IEEE Transactions on Information Theory*, 40(6), pp. 2035-2041, 1994.

[10] M. van der Schaar and H. Radha, “Unequal Packet Loss Resilience for Fine-Granular-Scalability Video”, *IEEE Transactions on Multimedia*, 3(4), pp 381-394, December 2001.

[11] X. K. Yang, C. Zhu, Z. G. Li, G. N. Feng, S. Wu, and N. Ling, “A Degressive Error Protection Algorithm for MPEG-4 FGS Video Streaming”, *ICIP 2002*, pp. 737-740, September 2002.