ACCEPTED FROM OPEN CALL

Fine Granularity Scalable Video: Implications for Streaming and a Trace-Based Evaluation Methodology

Patrick Seeling, Arizona State University
Philippe de Cuetos, ENST, Paris
Martin Reisslein, Arizona State University

ABSTRACT

Fine granularity scalability (FGS) is a new development in the area of video coding, which is designed to facilitate video streaming over communication networks. With FGS coding, the video stream can be flexibly truncated at very fine granularity to adapt to the available network resources. In this article we introduce the communications generalist to the basic properties of FGS video coding to provide background for the design of video streaming mechanisms for FGS video. We then outline a methodology for evaluating streaming mechanisms for FGS encoded video. The methodology relies on traces of the rate-distortion characteristics of FGS encoded video and enables networking researchers and practitioners without access to video codecs and video sequences to develop and evaluate rate-distortion optimized streaming mechanisms for FGS encoded video.

INTRODUCTION

Fine granularity scalability (FGS) is a new development in the design of video coding mechanisms, which has as its goal increased flexibility in video streaming [1]. With FGS coding, the video is encoded into a base layer and one enhancement layer. Similar to conventional scalable video coding, the base layer must be received completely in order to decode and display basic quality video. In contrast to conventional scalable video coding, which requires the reception of complete enhancement layers to improve on the basic video quality, with FGS coding the enhancement layer stream can be cut anywhere at the granularity of bits before transmission. The received part of the FGS enhancement layer stream can be successfully decoded and improves on the basic video quality.

With the fine granularity property of the enhancement layer, FGS encoded videos can flexibly adapt to changes in the available bandwidth in wired and wireless networks. This flexibility can be exploited by video servers to adapt the streamed video to the available bandwidth in real time without requiring any computationally demanding re-encoding. In addition, the fine granularity property can be exploited by intermediate network nodes (including base stations in wireless networks) to adapt the video stream to the currently available downstream bandwidth.

In this article we first introduce the communications and networking generalist without specific prior knowledge in the area of video coding to the basic properties of FGS videos and contrast them with the basic properties of conventional scalable as well as nonscalable video coding. Our focus is primarily on the implications of the coding properties for the design and evaluation of video streaming mechanisms.

We then introduce a methodology for evaluating the performance of streaming mechanisms for FGS video. The methodology employs traces of the rate-distortion characteristics of the FGS enhancement layer and allows the designers of streaming mechanisms to evaluate the quality of the streamed video without requiring experiments with actual video.

IMPLICATIONS OF FGS VIDEO FOR STREAMING MECHANISMS

With conventionally encoded video, the goal of the streaming mechanism is to deliver the complete video stream (or complete layers) in a timely fashion to avoid starvation (loss) of video data at the decoder. Network streaming mechanisms for conventional video typically focus on minimizing the loss of video data subject to the available resources (available bandwidth, buffers, startup latency, etc.). This is very challenging due to the variabilities in the video traffic (bit rate) and the typically varying bandwidth available for video streaming (see,e.g., [2]). Because of the best effort nature of the Internet, streaming video should be able to adapt (i.e., scale) to changes in the available transmission rate [3]. Conventional nonscalable video coding is not

This work has been supported in part by the National Science Foundation under Grant no. Career ANI-0 133252 and Grant no. ANI-0136774 as well as the state of Arizona through the 1T 301 initiative, and two matching grants from Sun Microsystems.

0163-6804/05/$20.00 © 2005 IEEE

designed to adapt to changes in the available transmission rate, and any starvation of the decoder results in increased distortion (lower quality) of the displayed video. As detailed later, assessing the quality of the displayed video after lossy network transport of nonscalable video generally requires experiments with actual video. With conventional scalable (layered) encoded video, the encoded video can adapt to changes in the available transmission rate at the granularity of layers (i.e., by adding and dropping enhancement layers as the available transmission rate increases or decreases). As detailed later, the quality of the displayed video can be obtained at the granularity of layers from the rate-distortion characteristics of the encoded video.

FGS video coding has the potential to fundamentally change video streaming in networks. In contrast to conventionally coded video, the FGS enhancement layer is designed to be cut (truncated) anywhere. Thus, it is not crucial to deliver the entire enhancement layer stream; instead, the enhancement layer can be flexibly cut to adapt to the available transmission rate. The received part (below the cut) can be decoded and contributes to the video quality according to the rate-distortion characteristics of the FGS enhancement layer. This has important implications for the design and evaluation of streaming mechanisms. The implication for the design of streaming mechanisms is that it is no longer necessary to deliver complete enhancement layers in a timely fashion. Instead, the streaming mechanisms can be designed to judiciously deliver a part of the FGS enhancement layer to maximize the delivered video quality (i.e., minimize distortion) subject to the available networking resources (including the available transmission rate). The development of such mechanisms has begun to attract significant interest [4, 5]. In the design of these mechanisms it is important to note that similar to conventional scalable encoding, the FGS enhancement layer is hierarchical in that "higher" bits require "lower" bits for successful decoding. This means that when cutting the enhancement layer bitstream before transmission, the lower part of the bitstream (below the cut) needs to be transmitted, and the higher part (above the cut) can be dropped.

The implication of FGS video for the evaluation of streaming mechanisms is that the decoded video quality corresponding to the received and decoded part of the enhancement layer can be determined directly from the enhancement layer rate-distortion characteristics at the granularity of bits. This enables the rate-distortion trace-based methodology we present later.

The development of FGS coding techniques is a topic of ongoing research in the video coding area. A form of FGS coding has recently been added to the MPEG-4 video coding standard [1]. In this form of FGS, however, the flexibility in cutting the enhancement layer comes at the expense of a quite significant degradation in compression performance. Ongoing research in video coding is seeking to maintain flexibility in enhancement layer cutting while reducing degradation in compression performance [6]. We note that the implications discussed above and the trace-based evaluation methodology are valid for any FGS coding technique that allows cutting of the FGS enhancement layer at the granularity of bits and apply analogously for FGS techniques that allow for the cutting of the enhancement layer at a coarser granularity.

Methodology for Evaluating FGS Streaming Mechanisms

In this section we outline a methodology for evaluating streaming mechanisms for FGS encoded video. This methodology is based on traces of the rate-distortion characteristics of individual video frames. Before we outline our methodology we give a brief overview of the issues and metrics involved in evaluating video streaming mechanisms.

Performance Metrics for Video Streaming Mechanisms

The key performance metric in networking studies on the streaming of conventionally encoded video has typically been the probability (or long run rate) of lost video data (i.e., data that misses its decoding and playout deadline at the client). This loss probability is a convenient metric for characterizing the performance of a video streaming mechanism as it can be obtained from video traffic models or frame size traces and does not require experiments with actual video codecs and video sequences. However, the loss probability is essentially a "network" metric and does not provide much quantitative insight into the video quality perceived by the user. Clearly, on a qualitative basis, a smaller starvation probability results generally in better video quality, but quantifying this relationship is very difficult without conducting experiments with actual video codecs and video sequences. This difficulty is due to the fact that for conventional video coding the encoder rate-distortion characteristics, which relate the bit rate at the encoder output to the video quality,[1] cannot directly be employed to assess the video quality after lossy network transport. (Note that for conventional scalable encoded video the decoded video quality can be obtained from the encoder rate-distortion characteristics at the granularity of complete layers.) Assessing video quality is further complicated by motion compensation and the resulting dependencies among the different frame types in H.26xand MPEG-encoded video. Also, a number of techniques have been developed to attempt to repair (conceal) losses or make encoded video more resilient to losses. All these issues need to be taken into consideration when assessing the decoded video quality after lossy network transport. We note that approximate heuristics that relate the loss of video data to the decoded video quality have been examined [7], but in general determining video quality after network transport requires experiments with actual video [8].

Experiments with actual video, however, pose a number of challenges. They require access to and expertise in operating video codecs. In addition, the long videos required

The evaluation methodology employs traces of the rate-distortion characteristics of the FGS enhancement layer and allows the designers of streaming mechanisms to evaluate the quality of the streamed video without requiring experiments with actual video.

[1] *Strictly speaking, the rate distortion characteristics give the distortion (not the quality) as a function of the bit rate, but it is common to refer to the quality as a function of rate as rate-distortion characteristics.*

The enhancement layer rate-distortion characteristics can be used to obtain the video quality for an arbitrary amount of enhancement layer bits delivered in time to the client.

for statistically valid networking studies imply very large file sizes, which are difficult to manage. Also, the copyright on the video content limits the exchange of videos among networking researchers for collaboration and the verification of video streaming protocols. Despite these challenges, the distortion (or quality) of the delivered and displayed video more closely reflects the usefulness or value of the video to the user than the network metric loss probability, and is thus a preferable performance metric for video streaming. Indeed, rate-distortion optimized streaming mechanisms that minimize the distortion (i.e., maximize the quality) of the received video (and not the loss probability) subject to the networking resources have recently started to attract significant attention [9, 10].

While the distortion of conventionally encoded video that has suffered losses during network transport can generally only be determined from experiments with actual video, the explicit design of FGS video coding for cuts in the enhancement layer makes it possible to determine the distortion of the received video frames from the rate-distortion characteristics. As detailed in the next section, the received enhancement layer part of a given video frame contributes to the decoded quality of that frame according to its rate-distortion characteristics.

Rate-Distortion Trace Methodology for Evaluating FGS Video Streaming

In this section we outline our methodology for evaluating the quality of the video delivered by an FGS video streaming mechanism. Recall that FGS encoding is designed to allow for cuts at an arbitrary bit rate of the enhancement layer. Thus, the enhancement layer rate-distortion characteristics can be used to obtain the video quality for an arbitrary amount of enhancement layer bits delivered in time to the client.

To make these ideas more precise we formulate the following evaluation methodology. We assume that the frame period (display time of one video frame) is constant and denote it by T s. Let N denote the number of frames in a given video and let n, $n = 1, ..., N$, index the individual video frames. Frame n is supposed to be decoded and displayed at the discrete instant $t = n \cdot T$. Suppose the base layer was encoded with some coding mechanism, resulting in either a constant bit rate or variable bit rate base layer (i.e., either constant or variable base layer frame sizes). For the sake of clarity we assume that in either case the transmission of each individual frame is spread out equally over the frame period preceding the actual display frame period; that is, the frame is transmitted at a constant bit rate over one frame period from $t = (n - 1) \cdot T$ to $t = n \cdot T$, $n = 1, \ldots, N$, such that it arrives just in time for its display. (The outlined methodology can be adapted to streaming mechanisms that transmit frames ahead of time in a straightforward fashion.) More formally, let X_n^b denote the size of the base layer of frame n (in bits or bytes). Let $X_{n,comp}^e$ denote the size of the complete FGS enhancement layer of frame n (i.e., the enhancement layer without any cuts). The base layer is thus transmitted with constant bit rate X_n^b/T. Similarly, the complete enhancement layer would be streamed at the constant bit rate $C_{n,comp} = X_{n,comp}^e/T$. We refer to the part of the enhancement layer of a frame that is actually received and decoded as the *enhancement layer subframe*. More formally, we introduce the following terminology. We say that the enhancement layer subframe is encoded at rate C_n, $0 \le C \le C_{n,comp}$, when the first $C_n \cdot T$ enhancement layer bits of frame n are received and decoded together with the base layer. In other words, the enhancement layer subframe is said to be encoded with rate C_n when the last $(C_{n,comp} - C_n) \cdot T$ bits have been cut from the FGS enhancement layer and are not decoded. Note that with the enhancement layer subframe coded at rate C_n, the total size of frame n is $X_n^b + C_n \cdot T$.

We now turn to the quality increase that can be achieved with the available enhancement layer bit rate C_n. Let $Q_n(C_n)$, $n = 1, \ldots, N$, denote the quality of the nth decoded video frame when the enhancement layer subframe is encoded with rate C_n. Let $Q_n^b = Q_n(0)$, denote the quality of the same video frame when only the base layer is decoded. We define $Q_n^e(C_n) = Q_n(C_n) - Q_n^b$ as the improvement (increase) in quality achieved when decoding the enhancement layer subframe encoded with rate C_n together with the base layer of frame n.

In general, the Q_n notation can be used to denote any arbitrary quality metric of the decoded video frame. The most common quality metric is the peak signal-to-noise ratio (PSNR). The PSNR gives the inverse of the mean square error, which is a distortion measure between the original video frame and the encoded and subsequently decoded video frame on a logarithmic scale (in dB). Higher PSNR values indicate higher video quality. One reason for the popularity of the PSNR is a recent Video Quality Expert Group (VQEG) report [11] which describes extensive experiments that compared several different objective quality measures with subjective quality evaluations (viewing and scoring by humans). It was found that none of the objective measures (some of them quite sophisticated and computationally demanding) performed better than the computationally very simple PSNR in predicting (matching) the scores assigned by humans.

The key characterization of each FGS encoded frame is the rate-distortion characterization of the FGS enhancement layer. This rate-distortion characterization of a given frame n gives the improvement in video frame quality Q_n^e as a function of the enhancement layer subframe bit rate C_n. This rate-distortion characterization can be recorded in a *rate-distortion trace*. The rate-distortion trace gives for each frame n of a given video the improvement in video quality Q_n^e for several sample values of the enhancement layer rate C_n. Such a trace is illustrated in Table 1 for an MPEG-4 FGS encoding of a few frames n from the *Table Tennis* video for a few sample rates C_n. We refer the interested reader to [12] for a discussion of more detailed traces, which also provide the base layer frame size X_n^b and base layer quality Q_n^b.

The rate-distortion trace can be exploited for

Frame n	C_n (in kb/s)								
	0	1000	2000	3000	4000	5000	6000	7000	8000
13	0	1.40	2.67	4.30	6.73	7.72	8.95	10.65	12.89
14	0	2.78	4.38	5.84	8.08	9.66	10.77	12.16	14.42
15	0	1.82	3.56	4.98	7.18	8.91	10.02	11.39	13.55
16	0	1.75	3.36	4.86	7.19	8.63	9.76	11.18	13.52

Table 1. *An excerpt of a rate-distortion trace of* Table Tennis*: quality improvement* Q_n^e *(in dB) for some FGS enhancement layer bit rates* C_n *for some frames* n.

The rate-distortion trace can be exploited for evaluating a given streaming mechanism for FGS enhanced video.

evaluating a given streaming mechanism for FGS encoded video as follows. The streaming mechanism is simulated with standard techniques from discrete event simulation, and the number of bits the streaming mechanism is able to deliver in time for the enhancement layer subframe n is determined. Suppose the streaming mechanism was able to deliver $C_n \cdot T$ bits (i.e., the enhancement layer subframe at rate C_n). Then we can read off the corresponding improvement in quality as $Q_n^e(C_n)$ from the rate-distortion trace for video frame n (whereby the quality improvement can be obtained by interpolating the nearest sample values for C_n). Together with the base layer quality Q_n^b we obtain the decoded video frame quality as $Q_n(C_n) = Q_n^b + Q_n^e(C_n)$. The video frame qualities $Q_n(C_n)$ thus obtained can be used to assess the statistics of the qualities of the decoded video frames. The mean video frame quality, for instance, is estimated as

$$\bar{Q} = \frac{1}{N}\sum_{n=1}^{N} Q_n(C_n).$$

Other elementary statistics of the video frame quality (e.g., the standard deviation) can be estimated in analogous fashion from the sequence $Q_1(C_1), Q_2(C_2), \ldots, Q_n(C_N)$. Maximization of the overall quality is generally achieved by maximizing the quality of the individual video frames and minimizing the variations in quality between consecutive video frames [2].

To illustrate the described evaluation procedure we consider the streaming of frames n = 13, 14, 15, and 16 for which the rate-distortion curves are provided in Table 1. Suppose for this illustrative example that the varying available network bandwidth allowed the server to transmit enhancement layer frames 13 and 14 with a bit rate of 4000 kb/s and enhancement layer frames 15 and 16 with a bit rate of 3000 kb/s, and that aside from this bandwidth fluctuation no additional transmission errors occurred on the transmission path from the server to the client. Furthermore, suppose that the base layer for all four frames was fully transmitted and provides a basic video quality of 27 dB for each video frame (i.e., $Q_{13}^b = Q_{14}^b = Q_{15}^b = Q_{16}^b = 27$ dB). Suppose we are interested in finding the average quality $\bar{Q}_n$ of the received video. To obtain $\bar{Q}$ we first find the improvement in quality achieved by the actually transmitted and received parts of the enhancement layer frames. For frame n = 13 whose enhancement layer was transmitted at the bit rate C_{13} = 4000 kb/s we obtain the quality improvement $Q_{13}^e(4000 \text{ kb/s}) = 6.73$ dB from the rate-distortion curve in Table 1. Similarly, we find $Q_{14}^e(4000 \text{ kb/s}) = 8.08$ dB, $Q_{15}^e(3000 \text{ kb/s}) = 4.98$ dB, and $Q_{16}^e(3000 \text{ kb/s}) = 4.86$ dB from the table. The decoded video quality is then obtained by adding the quality improvement from the enhancement layer to the base layer quality; that is, we obtain the decoded quality of frame n = 13 as $Q_{13}(4000 \text{ kb/s}) = Q_{13}^b + Q_{13}^e(4000 \text{ kb/s}) = 27 \text{ dB} + 6.73 \text{ dB} = 33.73$ dB. Analogously we find $Q_{14}(4000 \text{ kb/s}) = 35.08$ dB, $Q_{15}(3000 \text{ kb/s}) = 31.98$ dB, and $Q_{16}(3000 \text{ kb/s}) = 31.86$ dB, which results in average decoded video quality $\bar{Q} = 33.16$ dB.

Note that with the outlined evaluation methodology the encoding and decoding of the actual videos is decoupled from the simulation of the streaming mechanism. The encoding of the videos and decoding at a range of sample cutting rates are done offline, and the results are stored in rate-distortion traces. The simulation of the streaming mechanism is then simply conducted in terms of the numbers of bits to be transported for the individual frames instead of simulating the transport of the actual bits. With the number of received bits obtained from the simulation of the streaming mechanism, the traces are consulted to obtain the corresponding quality level.

Conclusions

We have given an overview of the implications of fine granularity scalable video coding for the design and evaluation of streaming mechanisms. We have outlined a methodology for evaluating the streaming of FGS video over networks. The methodology employs rate-distortion traces of FGS coded video and allows accurate evaluation of the quality of streamed FGS encoded video without requiring experimentation with actual video. The outlined methodology relies on the availability of rate-distortion traces of FGS encoded video. A basic library of such traces has been made publicly available [12], and is being continuously expanded and updated as new techniques for FGS coding are developed.

The outlined methodology relies on the availability of ratedistortion traces of FGS encoded video. A basic library of such traces has been made publicly available and is being continuously expanded and updated.

ACKNOWLEDGMENTS

We are grateful to Osama Lotfallah and Sethuraman Panchanathan of Arizona State University for explaining the intricacies of the MPEG-4 FGS reference software to us, and to Frank Fitzek of Acticom GmbH, Berlin, Germany, and Aalborg University, Denmark, who helped in setting up the trace library Web site. We are grateful to Keith W. Ross of Polytechnic University for support of this work. We thank the anonymous reviewers for their insightful and detailed comments which helped to greatly improve the quality of this article.

REFERENCES

[1] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, Mar. 2001, pp. 53–68.

[2] D. Wu *et al.*, "Streaming Video over the Internet: Approaches and Directions," *IEEE Trans. Circuits and Sys. for Video Tech.*, vol. 11, no. 3, Mar. 2001, pp. 1–20.

[3] R. Rejaie and A. Reibman, "Design Issues for Layered Quality-Adaptive Internet Video Playback," *Proc. Wksp. Digital Commun.*, Taormina, Italy, Sept. 2001, pp. 433–51.

[4] K. W. Stuhlmuller *et al.*, "Scalable Internet Video Streaming with Unequal Error Protection," *Proc. Packet Video Wksp.*, New York, NY, Apr. 1999.

[5] J. Vieron *et al.*, "TCP-Compatible Rate Control for FGS Layered Multicast Video Transmission Based on a Clustering Algorithm," *Proc. IEEE Int'l. Symp. Circuits and Sys.*, Scottsdale, AZ, May 2002, pp. 453–56.

[6] S.-R. Chen, C.-P. Chang, and C.-W. Lin, "MPEG-4 FGS Coding Performance Improvement using Adaptive Inter-layer Prediction," *Proc. IEEE Int'l. Conf. on Acoustics, Speech, and Sig. Proc.*, Montreal, Canada, May 2004, pp. 265–68.

[7] N. Duffield, K. Ramakrishnan, and A. Reibman, "Issues of Quality and Multiplexing When Smoothing Rate Adaptive Video," *IEEE Trans. Multimedia*, vol. 1, no. 4, Dec. 1999, pp. 53–68.

[8] X. Lu, R. 0. Morando, and M. ElZarki, "Understanding Video Quality and Its Use In Feedback Control," *Proc. Packet Video Wksp.*, Pittsburgh, PA, Apr. 2002.

[9] P. A. Chou and A. Sehgal, "Rate-Distortion Optimized Receiver-Driven Streaming over Best-Effort Networks," *Proc. Packet Video Wksp.*, Pittsburg, PA, Apr. 2002.

[10] Z. Miao and A. Ortega, "Expected Run-time Distortion Based Scheduling for Delivery of Scalable Media," *Proc. Packet Video Wksp.*, Pittsburg, PA, Apr. 2002.

[11] A. M. Rohaly *et al.*, "Video Quality Experts Group: Current Results and Future Directions," *Proc. SPIE Visual Commun. and Image Proc.*, vol. 4067, Perth, Australia, June 2000, pp. 742–53.

[12] P. de Cuetos *et al.*, "Evaluating the Streaming of FGS-Encoded Video with Rate-Distortion Traces," Arizona State Univ., tech. rep., 2004; traces available from http://trace.eas.asu.edu/indexfgs.html

BIOGRAPHIES

PATRICK SEELING [StM] (patrick.seeling@asu.edu) received a Dipl.-Ing. degree in industrial engineering and management (specializing in electrical engineering) from the Technical University of Berlin (TUB), Germany, in 2002. Since 2003 he has been a Ph.D. student in the Department of Electrical Engineering at Arizona State University. His research interests are in the area of video communications in wired and wireless networks. He is a student member of the ACM.

PHILIPPE DE CUETOS (philippe. de-cuetos@erist. fr) is a researcher with the Ecole Nationale Superieure des Telecommunications (ENST), Paris, France. He received a Ph.D. degree from the Institut Eurecom/University of Nice in 2003. His research interests are in the area of video communication, in particular exploiting fine granular scalability for efficient network transport.

MARTIN REISSLEIN (reisslein@asu.edu) is an assistant professor in the Department of Electrical Engineering at Arizona State University, Tempe. He received a Dipl.-Ing. (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and an M.S.E. degree from the University of Pennsylvania, Philadelphia, in 1996, both in electrical engineering. He received his Ph.D. in systems engineering from the University of Pennsylvania in 1998. During academic year 1994–1995 he visited the University of Pennsylvania as a Fulbright scholar. From July 1998 through October 2000 he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin, and a lecturer at TUB. He is Editor-in-Chief of *IEEE Communications Surveys and Tutorials* and has served on the Technical Program Committees of IEEE INFOCOM, IEEE GLOBECOM, and the IEEE International Symposium on Computers and Communications. He has organized sessions at the IEEE Computer Communications Workshop (CCW). He maintains an extensive library of video traces for network performance evaluation, including frame size traces of MPEG-4 and H.263 encoded video, at http://trace.eas.asu.edu. He is co-recipient of the Best Paper Award of the SPIE Photonics East 2000 — Terabit Optical Networking conference. His research interests are in the areas of Internet QoS, video traffic characterization, wireless networking, and optical networking.