INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 2005; **18**:449–464
Published online 15 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/dac.711

Comparing the streaming of FGS encoded video at different aggregation levels: frame, GoP, and scene

Philippe de Cuetos[1,‡,§], Patrick Seeling[2,¶], Martin Reisslein[2,*,†] and Keith W. Ross[3,‖]

[1] *Institut Eurecom, Sophia-Antipolis, France*
[2] *Department of Electrical Engineering, Arizona State University, Goldwater Center MC 5706, Tempe AZ 85287-5706, U.S.A.*
[3] *Polytechnic University, 6 MetroTech Center, Brooklyn, NY 11201, U.S.A.*

SUMMARY

Fine granularity scalability (FGS), a new coding technique that has recently been added to the MPEG-4 video coding standard, allows for the flexible scaling of each individual video frame at very fine granularity. This flexibility makes FGS video very well suited for rate-distortion optimized streaming mechanisms, which minimize the distortion (i.e. maximize the quality) of the streamed video by transmitting the optimal number of bits for each individual frame. The per-frame optimization of the transmission schedule, however, puts a significant computational burden on video servers and intermediate streaming gateways. In this paper we investigate the rate-distortion optimized streaming at different video frame aggregation levels. We find that compared to the optimization for each individual video frame, optimization at the level of video scenes reduces the computational effort dramatically, while reducing the video quality only very slightly. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: fine granularity scalability; multimedia communications; performance evaluation; rate-distortion optimized streaming; scalable video; video scene; video streaming

1. INTRODUCTION

The MPEG-4 video coding standard has recently been augmented by fine granularity scalability (FGS), which has been designed to increase the flexibility of video streaming. An FGS encoded

[*] Correspondence to: Martin Reisslein, Department of Electrical Engineering, Arizona State University, Goldwater Center MC 5706, Tempe AZ 85287-5706, U.S.A.
[†] E-mail: reisslein@asu.edu
[‡] E-mail: philippe.de-cuetos@enst.fr
[§] He is now with ENST, Paris, France.
[¶] E-mail: patrick.seeling@asu.edu
[‖] E-mail: ross@poly.edu

Contract/grant sponsor: National Science Foundation; contract/grant numbers: Career ANI-0133252; ANI-0136774
Contract/grant sponsor: State of Arizona; contract/grant number: IT301
Contract/grant sponsor: Sun Microsystems

Received March 2004
Revised September 2004
Accepted October 2004

video stream consists of one base layer and one enhancement layer. The base layer must be completely received to decode and display a basic quality video. The FGS enhancement layer can be cut anywhere at the granularity of bits and the received part (below the cut) can be decoded and improves upon the basic quality video. This fine granularity scalability, which is achieved by a bitplane coding technique [1, 2], allows the server or intermediate network nodes or gateways to adapt the transmission rate finely to changing network conditions. In a typical scenario for transmitting MPEG-4 FGS encoded videos over the Internet, the base layer is transmitted with high reliability (achieved through appropriate resource allocation and/or channel error correction) and the FGS enhancement layer is transmitted with low reliability (i.e. in a best effort manner and without error control).

The goal of video streaming is to maximize the overall quality of the delivered video, which is generally achieved by first maximizing the quality (minimizing the distortion) of the individual video frames and then minimizing the variations in quality between consecutive video frames [3]. These goals have been pursued at the relatively coarse basis of video layers by the streaming mechanisms for conventionally coded layered video that maximize the number of delivered layers and minimize the changes in the *number of completely delivered layers*, see for instance References [4–6]. Rate-distortion optimized streaming mechanisms [3, 7, 8] exploit the rate-distortion characteristics of the encoded video to *minimize the distortion*, i.e. maximize the overall video quality at the receiver, while meeting the constraints imposed by the underlying network. This optimization of the overall video quality is in general approached by algorithms that take the rate-distortion functions of all individual video frames into account. FGS-encoded video makes it possible to implement optimized transmission schedules that transmit the optimal number of enhancement layer bits for each individual video frame (image), subject to the bandwidth constraints. This per-frame optimization can be computationally very demanding, especially for the typically non-linear rate-distortion characteristics of the FGS enhancement layer. The large computational burden of the streaming optimization may thus limit the number of simultaneous streams that a video server may support. Also, the optimization may need to be done at intermediate streaming gateways (e.g. base stations in a wireless system) to optimize the video delivery subject to the available downstream bandwidth.

To address this large computational burden on video servers and intermediate streaming gateways we examine in this paper the rate-distortion optimization for different aggregation levels of video frames. We explore the optimization approaches where the server/intermediate streaming gateway groups several consecutive frames of the video into *streaming sequences* and performs rate-distortion optimization over the streaming sequences. With this aggregation approach, the optimization assigns a specific bit budget to each streaming sequence. Within a given streaming sequence we then split the bit budget evenly across the frames of the streaming sequence, i.e. we allocate each frame within a given sequence the same number of bits.

More specifically, we consider a commonly studied two-tiered streaming system which employs (i) a coarse-grained streaming strategy which negotiates bandwidth (i.e. a bit budget) with the network for so-called *allocation segments* that are on the order of tens of seconds or a minute long, and (ii) a fine-grained streaming strategy which conducts rate-distortion optimization to optimize the video quality of the allocation segment given the allocated bit budget. This second tier—the fine-grained streaming strategy—is where generally the computationally highly demanding frame-by-frame rate-distortion optimization is conducted.

We seek to overcome this high computational burden by examining the aggregation of the frames over (i) a group of pictures (GoP), (ii) a scene of the video, (iii) a constant length segment with the duration of the average scene length in the allocation segment, or (iv) over the entire rate allocation segment. We compare the video quality achieved with these different aggregation levels with the video quality achieved with the frame-by-frame optimization. We demonstrate that by exploiting the strong correlations in quality between the consecutive video frames within a video scene, the scene aggregation approach dramatically decreases the computational requirement of the optimization procedure while reducing the video quality only very slightly. In typical scenarios, the scene based optimization approach reduces the computational load by two or more orders of magnitude while reducing the average PSNR video frame quality by less than 0.1 dB.

This paper is organized as follows. In the following subsection we discuss the related work. In Section 2, we describe the streaming scenario and the different aggregation levels studied in this paper. In Section 3, we present the evaluation framework and define the notation used to formulate the streaming optimization problem. In Section 4, we formulate the optimization problem and describe its solution. We present the results of the optimization for the different aggregation levels in Section 5 and briefly summarize our conclusions in Section 6.

1.1. Related work

Over the past few years, streaming video over the Internet has been the focus of many research efforts (see References [9, 10] for comprehensive surveys). Because of the best-effort nature of the Internet, streaming video should adapt to the changing network conditions. One of the most popular techniques for network-adaptive streaming of stored video is using scalable video (see for instance References [11, 12]). Video streaming applications should also adapt to the properties of the particular encoded video [8]. Recently, rate-distortion optimized streaming algorithms have been proposed (e.g. [7, 13]) to minimize the end-to-end distortion of media, for transmission over the Internet. Our work is complementary to these studies in that we evaluate video streaming mechanisms for a specific type of video, namely FGS encoded video.

Significant efforts have gone into the development of the FGS amendment to the MPEG-4 standard, see for instance References [1, 2] for an overview of these efforts. Models for the rate-distortion characteristics of FGS video are developed in References [14, 15]. Recently, the streaming of FGS video has been examined in a number of studies, all of which are complementary to our work. General frameworks for FGS video streaming are discussed in References [16–18]. The error resilience of FGS video streaming is studied in References [3, 19, 20]. In Reference [21] the FGS enhancement layer bits are assigned to different priority levels, which represent the importance of the carried content. In Reference [22] a real-time algorithm for the network adaptive streaming of FGS-encoded video is proposed. The proposed algorithm does not take the rate-distortion characteristics of the encoded video into consideration. The concept of scene-based streaming is briefly introduced in Reference [23], but not evaluated with rate-distortion data. Streaming mechanisms which allocate the FGS enhancement layer bits over fixed length segments are studied in References [24, 25] and evaluated using the well-known short MPEG test sequences. In contrast, in this paper we study the allocation of the FGS enhancement layer bits on individual frame, fixed-length segment, and video scene basis using traces of long videos. A bit allocation scheme with sliding window is developed in Reference [26].

A packetization and packet drop policy for FGS video streaming is proposed in Reference [27]. The transmission of FGS video over multiple network paths is studied in Reference [28]. An efficient approach for the decoding of streamed FGS video is proposed in Reference [29]. Refined video coding mechanisms and bit allocation mechanisms to equalize the quality over the area of a given video frame are developed in References [30, 31]. Finally, streaming of FGS video over multicast [32] and to wireless clients [33–39] has also been considered, while issues of FGS complexity scaling and universal media access are addressed in [40, 41].

For completeness we also note that streaming algorithms for various refinements of the basic fine granularity scalability considered in this paper have begun to attract interest, see for instance References [42–46].

2. COMPARISON SET-UP: DEFINITION OF DIFFERENT STREAMING SEQUENCES

In this section we describe our set-up for the comparison of the rate-distortion optimized streaming of FGS-encoded video at different levels of video frame aggregation. We suppose that the transmission of the base layer is made reliable, and we focus on the streaming of the enhancement layer. When streaming video over the best-effort Internet, the available bandwidth typically fluctuates over many time-scales. However, for the streaming of stored video, the user can usually tolerate an initial build-up delay, during which some initial part of the video is prefetched into the client before the start of the playback. Maintaining a sufficient playback delay throughout the rendering allows the application to accommodate future bandwidth variations (see for instance References [11, 12]).

To account for bandwidth variability, we model bandwidth constraints and client buffering resources as follows. As shown in Figure 1, a video consisting of N video frames is partitioned into L *allocation segments*, with each allocation segment l containing the same number of frames

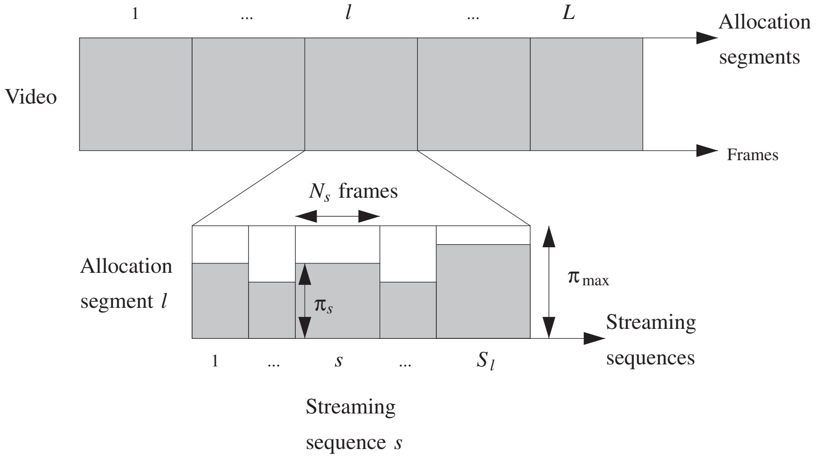


Figure 1. The video is partitioned into L allocation segments, each consisting of S_l streaming sequences.

N/L. While the server is streaming the video, for each allocation segment l, the server assigns a maximum bandwidth budget of $B_{\max} = C_{\max}NT/L$ bits to be allocated across all the frames in the segment, where T denotes the frame period (display time of one video frame). The maximum average bit rate $C_{\max}$ typically varies from one allocation segment to the next and is typically the outcome of a negotiation between the video server/intermediate streaming gateway with the downstream network for the available bandwidth. The values for $C_{\max}$ are determined by a coarse-grain streaming strategy, such as those given in References [11, 22, 47]. In this study, we focus on the fine-grain streaming strategy, namely, the allocation of the bandwidth budget to the individual frames within an allocation segment. In our experiments, we use allocation segments consisting of $N/L = 1000$ frames, which correspond to about 30 s of a 30 frames/s video.

Due to the client buffering, the server has great flexibility in allocating the given bandwidth budget to the frames within an allocation segment. Given the rate-distortion functions of all video frames, the server can optimize the streaming within the allocation segment by allocating bits from the bandwidth budget to the individual frames so as to maximize the video quality. Alternatively, the server can group several consecutive video frames of an allocation segment into sub-segments, which we refer to as *streaming sequences*, and perform rate-distortion optimization on the granularity of streaming sequences. In this case, each frame in a given streaming sequence (that is sub-segment) is allocated the same number of bits. We denote by S_l the number of streaming sequences in a given allocation segment l, $l = 1, \ldots, L$.

We consider five aggregation cases for streaming sequences:

- Video frames—each video frame of the considered allocation segment forms a distinct streaming sequence ($S_l = 1000$ with the allocation segment length of 1000 frames considered in this study).
- GoPs—we group all video frames from the same GoP into one streaming sequence. In this case, the number of streaming sequences in allocation segment l is equal to the number of distinct GoPs in the allocation segment ($S_l = 1000/12 \approx 83$ with the 12 video frame GoP used in this study).
- Scenes—we group all video frames from the same video scene into one streaming sequence. In this case $S_l = S_l^{\text{scene}}$, where S_l^{scene} denotes the number of distinct scenes in allocation segment l.
- Constant—allocation segment l is divided into $S_l^{\text{const}} = S_l^{\text{scene}}$ streaming sequences, each containing the same number of frames. Consequently, each streaming sequence contains a number of frames equal to the average scene length of the allocation segment.
- Total—all the video frames from allocation segment l form one streaming sequence ($S_l = 1$).

In order to simplify the notation, we focus in the following on the streaming of a particular allocation segment l. We remove the index l from all notations whenever there is no ambiguity. Let S denote the number of streaming sequences in the considered allocation segment. Let N_s denote the number of frames in streaming sequence s, $s = 1, \ldots, S$ (see Figure 1). Furthermore, let n, $n = 1, \ldots, N_s$, denote the individual frames in streaming sequence s. The main notation used in this paper is summarized in Table I.

Table I. Summary of Notation.

N	Total number of frames in video
$\bar{N}$	Average number of frames in a scene in given video
L	Number of allocation segments in video ($= N/1000$ in this study)
T	Frame period (1/30 s in this study)
$B_{\max}$	Maximum bit budget for an allocation segment ($= C_{\max}NT/L$)
S	Number of streaming sequences in a given allocation segment
N_s	Number of video frames in streaming sequence s, $s = 1, \ldots, S$
π_s	Number of bits allocated to a frame in streaming sequence s (decision variable in opt. problem)
$d_{s,n}(\pi_s)$	Distortion of frame n in streaming sequence s when π_s bit are allocated to it

3. EVALUATION SET-UP: RATE-DISTORTION TRACES AND PERFORMANCE METRICS

In this section we briefly describe the FGS rate-distortion traces employed in our study and define the considered performance metrics. We also define the notation used in the formulation of the rate-distortion streaming optimization.

We use the rate-distortion traces of the videos listed in Table II, which we obtained from Reference [48]. The traces give the quality of each decoded video frame q_n as a function of the number of enhancement layer bits π_n that have been delivered to the client for frame n, i.e. the traces give $q_n(\pi_n)$, see References [49, 50] for details. The video frame quality is given in terms of the peak signal-to-noise ratio (PSNR) in dB units, which is a widely used metric for the quality of video frames [51]. For ease of formulating the streaming optimization, we will consider the video frame distortion in terms of the mean square error (MSE) d_n in our formulation. The quality and distortion of a video frame n are related by

$$q_n = 10 \log \frac{255^2}{d_n} \tag{1}$$

To assess the quality of a streaming sequence s containing video frames n, $n = 1, \ldots, N_s$, we average the MSE distortions of the individual frames, which is common in assessing the quality of sequences of video frames. Thus, the distortion of streaming sequence s is obtained as

$$D_s = \frac{1}{N_s} \sum_{n=1}^{N_s} d_{s,n} \tag{2}$$

where $d_{s,n}$ denotes the distortion of frame n of sequence s. The quality Q_s of the sequence s is then obtained by converting the MSE distortion D_s into PSNR quality using (1). Analogously we define D for the distortion and Q for the quality of a given allocation segment containing the streaming sequences s, $s = 1, \ldots, S$, i.e.

$$D = \frac{1}{S} \sum_{s=1}^{S} D_s \tag{3}$$

The traces from Reference [48] provide the boundaries of the *scene shots*, which we employ in our evaluation of the scene-based streaming optimization. The total number of scene shots, the average length $\bar{N}$ of the scene shots in number of video frames, as well as the coefficient of

Table II. Scene shot length statistics of considered videos.

	Run time (min)	# of frames N in video	# of scenes in video	Avg. scene length $\bar{N}$ (video frames)	Coefficient of variation of scene length	Peak-to-mean ratio of scene length
The Firm	60	108 000	890	121	0.94	9.36
Oprah with Commercials	60	108 000	621	173	2.46	39.70
Oprah	38	68 000	320	215	1.83	23.86
News	60	108 000	399	270	1.67	9.72
Star Wars	60	108 000	984	109	1.53	19.28
Silence of the Lambs	30	54 000	184	292	0.96	6.89
Toy Story	60	108 000	1225	88	0.95	10.74
Football	60	108 000	876	123	2.34	31.47
Lecture	49	88 000	16	5457	1.62	6.18

variation (standard deviation divided by mean), and ratio of largest to average scene shot length are provided in Table II. We note that the segmentation of the video into scene shots is relatively coarse in that it only considers director cuts and ignores other changes in the motion or visual content between two successive director cuts. A finer scene segmentation that takes into consideration those changes between director cuts is largely still a subject of ongoing research. Nevertheless, distinct scene shots are likely to have distinct visual characteristics, so we believe that considering scene shot segmentation instead of a finer segmentation does not have a strong effect on the conclusions of our study. In fact, a finer segmentation would only increase the total number of distinct video scenes, and increase the correlation between the qualities of the frames in a scene. Henceforth we refer to a scene shot in short as *scene*.

4. FORMULATION OF RATE-DISTORTION STREAMING OPTIMIZATION

In this section we formally describe the rate-distortion optimization of the streaming of a given allocation segment. Recall that we consider a typical scenario, where a coarse-grain streaming strategy gives the allowed average bit rate $C_{\max}$, i.e. the bit budget $B_{\max} = C_{\max}NT/L$, allocated to the transmission of the enhancement layer frames in the allocation segment. Our goal is to compare different frame aggregation levels in the streaming optimization, whereby the same number of bits is allocated to each frame inside a given frame aggregate (streaming sequence), i.e. the enhancement layer streaming bit rate is adjusted more or less frequently. Formally, we denote π_s for the number of bits allocated to each video frame among the N_s video frames in streaming sequence s. We define $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S)$ as the streaming policy for a given allocation segment. We denote by $\pi_{\max}$ the maximum number of enhancement layer bits that can be allocated to any video frame of the video.

We study the quality $Q(\boldsymbol{\pi})$ of the current allocation segment as a function of the streaming policy $\boldsymbol{\pi}$. We denote $D(\boldsymbol{\pi})$ for the distortion of the allocation segment as a function of the streaming policy $\boldsymbol{\pi}$.

With these definitions we can formulate the streaming optimization problem as follows:

For the given allocation segment, a given bandwidth constraint $C_{\max}$, and a given aggregation case, the optimization procedure at the server consists of finding the policy $\boldsymbol{\pi}^* = (\pi_1^*, \ldots, \pi_S^*)$

that optimizes:

$$\begin{aligned} \text{minimize} \quad & D(\boldsymbol{\pi}) = \textstyle\sum_{s=1}^{S} \left(\dfrac{1}{N_s} \textstyle\sum_{n=1}^{N_s} d_{s,n}(\pi_s) \right) \\ \text{subject to} \quad & \textstyle\sum_{s=1}^{S} N_s \pi_s \leqslant B_{\max} \\ & \pi_s \leqslant \pi_{\max}, \ s = 1, \ldots, S \end{aligned}$$

We denote $D^* = D(\boldsymbol{\pi}^*)$ (respectively Q^*) for the minimum distortion (maximum quality) achieved for the considered allocation segment. Our problem is a resource allocation problem, which can be solved by dynamic programming [52]. Dynamic programming is a set of techniques that are used to solve various decision problems. In a typical decision problem, the system transitions from state to state, according to the decision taken for each state. Each transition is associated with a profit. The problem is to find the optimal decisions from the starting state to the ending state of the system, i.e. the decisions that maximize the total profit, or in our context minimize the average distortion.

The most popular technique to solve such an optimization problem is recursive fixing. Recursive fixing recursively evaluates the optimal decisions from the ending state to the starting state of the system. This is similar to the well-known Dijkstra algorithm which is used to solve shortest-path problems. We note that due to the non-linear shape of the rate-distortion curves, which as observed in Reference [48] are neither convex nor concave, we cannot employ the computationally less demanding marginal analysis technique of dynamic programming. It was found in Reference [48] that sampling the values of enhancement layer bits per video frame in steps of 833 bytes with a maximum of $\pi_{\max} = 8333$ bytes per video frame accurately captures the characteristics of the rate-distortion curve. Hence we employ the same step size of 833 bytes and $\pi_{\max}$ in the recursive fixing solution of our dynamic program.

The computational effort required for resolving our problem depends largely on the number of decision variables $\pi_s, \ s = 1, \ldots, S$, i.e. on the number S of streaming sequences in a given allocation segment. In particular, note that S depends on the aggregation case which is considered, namely video frame, GoP, scene, constant, or total, as well as on the length of an allocation segment (and also on the number of scenes within a given allocation segment for the scene and constant aggregation cases). For a given fixed length of the allocation segment (which is typically dictated by the bandwidth negotiation mechanisms in the network), the computational effort depends primarily on the aggregation case and scene length characteristics. In particular, we observe from Table II that the average scene length $\bar{N}$ is typically on the order of hundreds to thousands of frames. In other words, the scene aggregation approximately reduces the computational effort for the streaming optimization by a factor of $\bar{N}$, i.e. by two to three orders of magnitude compared to video frame-by-video frame optimized streaming.

Aside from these computational savings, the streaming with aggregation has the side effect of smoothing the network traffic, which is beneficial for a variety of networking mechanisms, such as link load probing, buffer management, and congestion control.

5. COMPARISON RESULTS

Figure 2 gives plots of the maximum quality Q^* for each individual allocation segment, for an average target rate of $C_{\max} = 1000$ kbps for the first four videos listed in Table II. Not

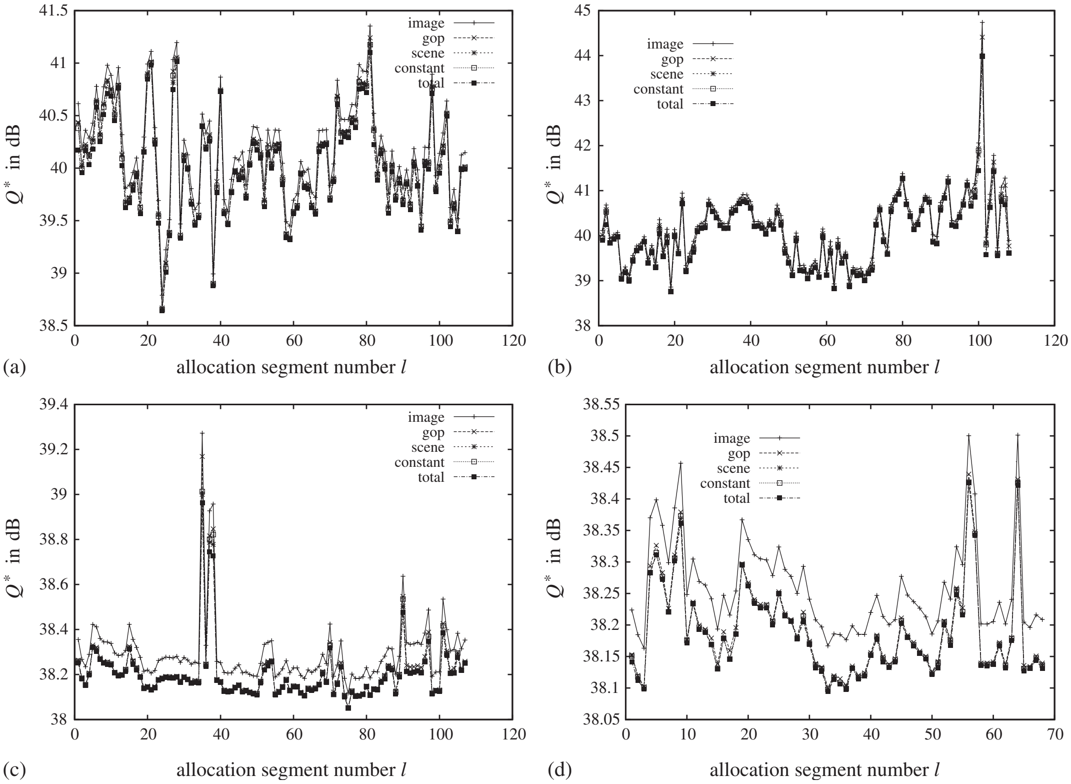


Figure 2. Maximum (PSNR) quality Q^* for each individual allocation segment for $C_{\max} = 1000$ kbps for (a) The Firm, (b) News, (c) Oprah with Commercials, and (d) Oprah.

surprisingly, for all allocation segments, the quality with video frame-by-video frame streaming optimization is higher than that for the other aggregation cases. This is because the video frame-by-video frame optimization is finer. However, the quality achieved by the other aggregation cases is very close to that of video frame-by-video frame streaming: the difference is usually only approximately 0.1 dB in PSNR. This is due to the high correlation between the enhancement layer rate-distortion curves of successive frames.

We show in Figure 3 the maximum quality averaged over all allocation segments for the entire video as a function of the target rate constraint $C_{\max}$. We observe again that the frame-by-frame streaming optimization gives slightly better quality. Importantly, we also observe that transmitting the same number of bits for each frame in a given allocation segment (case of total aggregation) tends to give quite significantly lower quality than the other aggregation levels for some ranges of the bit rate allocation $C_{\max}$. For the target bit rates of 400 and 600 kbps, for instance, the total aggregation gives a up to 1 dB less in average quality for *The Firm* and *News*, which is typically considered a significant difference in quality. We observed similar behaviours for the other videos and also for allocation segments that contain more than 1000 frames. The observed relatively large quality degradation with total aggregation is primarily due to the relatively steep rate-distortion curves of the individual video frames in that bit rate range.

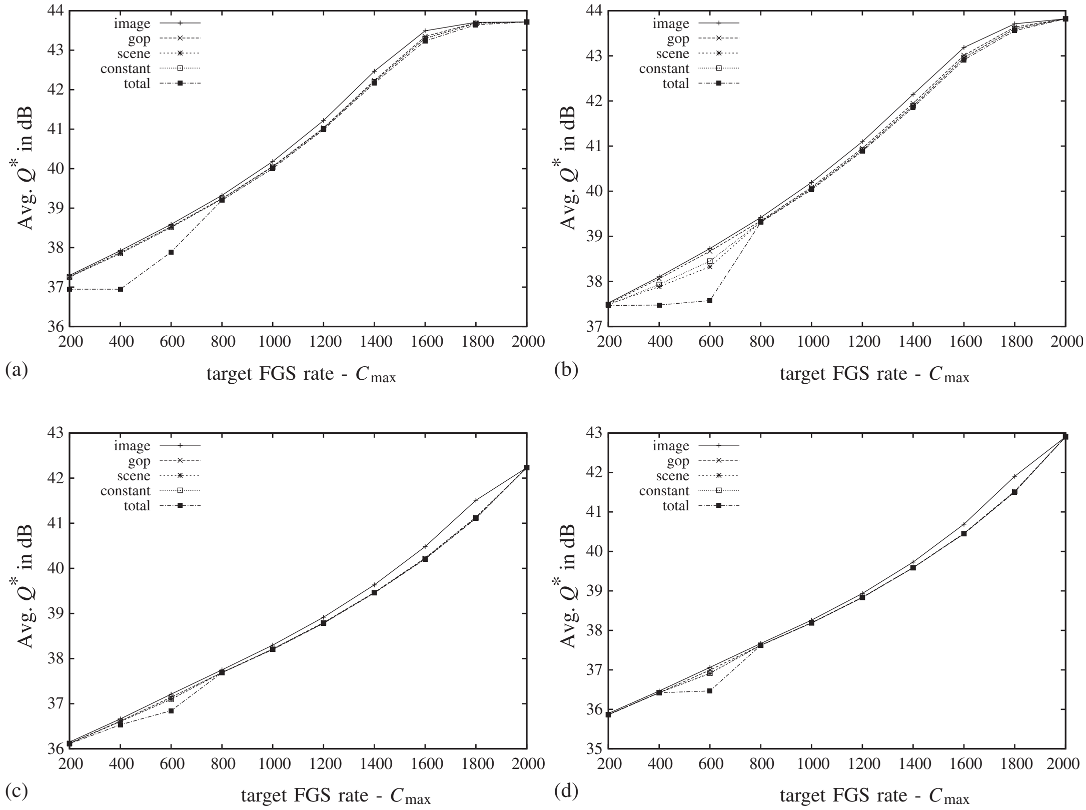


Figure 3. Maximum (PSNR) quality Q^* averaged over the individual allocation segments as a function of bit rate allocation $C_{\max}$ in kbps for (a) The Firm, (b) News, (c) Oprah with Commercials, and (d) Oprah.

The steep rate-distortion curves make the quality relatively more sensitive to changes in the bit budget, resulting in a more degraded video quality for the relatively coarse optimization over the entire allocation segment.

From the results considered so far we may draw two main conclusions. First, streaming optimization on the basis of individual video frames gives typically only insignificantly higher quality than optimizing on the basis of streaming sequences that have the length of video scenes. Secondly, allocating the same number of bits for each frame in an allocation segment may give significantly lower quality than optimizing the streaming on the basis of streaming sequences that have the length of video scenes. Consequently it appears that both the scene aggregation and the constant aggregation are equally well suited for video streaming. In other words, it appears that there is essentially no difference between the maximum quality achieved when aggregating over scenes or arbitrary sequences. We now proceed to examine these two aggregation cases more closely.

Despite the essentially same average quality for the scene and constant aggregation, the actual perceived quality may be somewhat different. This is because the average PSNR quality does not account for temporal effects, such as the variations in quality between consecutive video frames: two sequences with a same average quality may have different variations in video frame MSE, and thus different perceived quality. To illustrate this phenomenon, we monitor the maximum quality variation between consecutive video frames $\max_{i=2,\ldots,N_s}\{|q_{s,i}(\pi_s) - q_{s,i-1}(\pi_s)|\}$ of a given streaming sequence.

Table III gives the maximum variation in quality averaged over all streaming sequences in the video for different $C_{\max}$ rates. We observe that the average maximum variation in quality for a given FGS rate is always smaller with scene aggregation than with constant aggregation. The difference is insignificant for the *Lecture* video which is characterized by slow motion and long scenes, but reaches close to 0.3 dB for the *News*, *Toy Story*, and *Football* videos which are characterized by high motion as well as relatively short and highly variable scene lengths. Overall, the results in the table indicate that selecting a constant number of bits for the enhancement layer of all video frames within a given video scene yields on average a smaller maximum variation in video frame quality than selecting a constant number of bits for fixed-length sequences with a sequence length equal to the average scene length. Note that both approaches have the same computational effort for the streaming optimization. Therefore, it is preferable to choose streaming sequences that correspond to visual scene shots rather than segmenting the video arbitrarily. This result is intuitive since frames within a given scene shot are more likely to have similar visual complexity, and thus similar rate-distortion characteristics, than frames from different scenes.

To further examine this effect, we plot in Figure 4 the minimum value of the maximum variation in quality over all streaming sequences of a given allocation segment. We observe that the min–max variation in video frame quality is typically larger for arbitrary segmentation. This indicates that the minimum jump in quality in the streaming sequences is larger for arbitrary segmentation (constant aggregation). In the case of scene segmentation the minimum jumps in quality are smaller; when the scene shot consists of one homogeneous video scene without any significant changes in motion or visual content, the maximum variation is close to 0 dB. As shown in Figure 4, for some allocation segments, the difference with arbitrary segmentation can be more than 1 dB.

More generally, we expect the difference in rendered quality between scene-based segmentation and arbitrary segmentation to be more pronounced with a scene segmentation that is finer than scene shot-based segmentation. A finer segmentation would further segment

Table III. Average of maximum quality variation (in dB) for scene aggregation and constant aggregation.

	$C_{\max} = 800$ kbps		$C_{\max} = 1600$ kbps	
	Scene	Const.	Scene	Const.
The Firm	1.84	1.99	0.81	0.92
Oprah with Commercials	2.64	2.77	2.68	2.76
Oprah	2.43	2.47	2.60	2.64
News	2.22	2.55	1.43	1.64
Star Wars	1.90	2.11	0.85	0.97
Silence of the Lambs	1.33	1.37	1.37	1.40
Toy Story	2.34	2.54	1.46	1.74
Football	2.21	2.56	1.09	1.38
Lecture	2.64	2.69	2.06	2.08

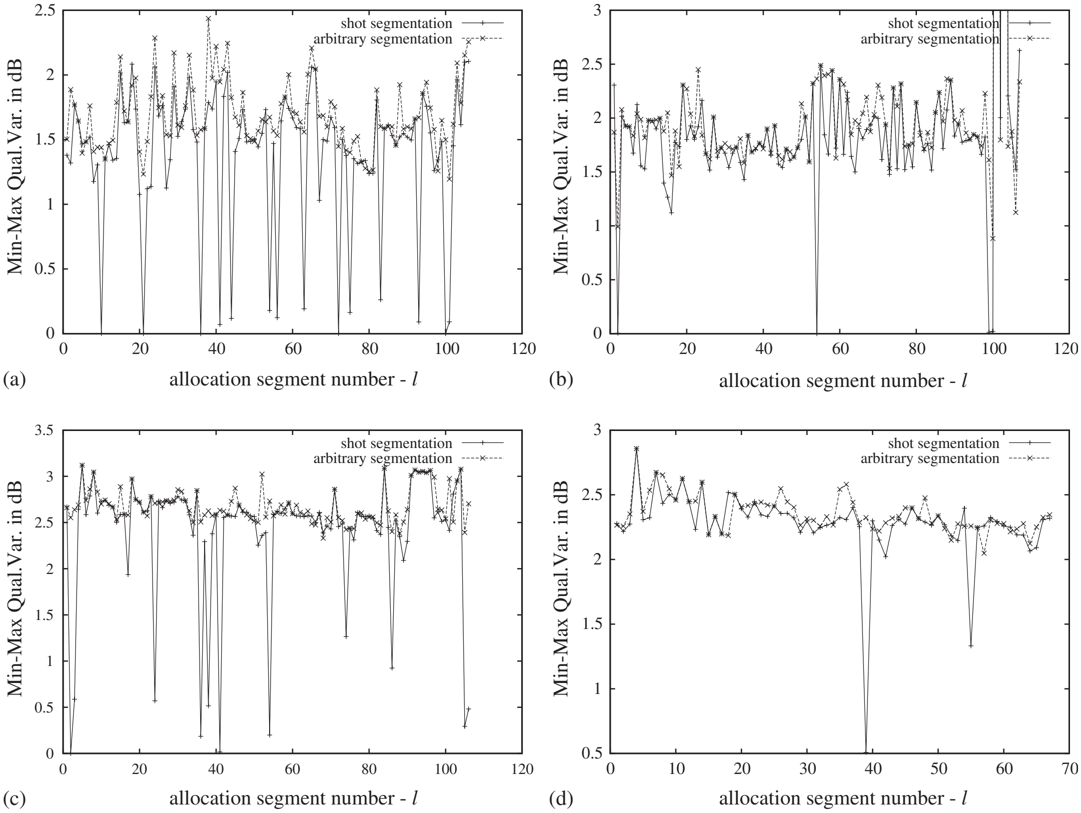


Figure 4. Min–max quality variations for the individual allocation segments for $C_{\max} = 800$ kbps for (a) The Firm, (b) News, (c) Oprah with Commercials, and (d) Oprah.

sequences with varying rate-distortion characteristics, e.g. sequences with changes in motion or visual content other than director's cuts. This would increase the correlation between the qualities of the frames within a scene, which would further reduce the quality degradation due to scene-based streaming optimization over video frame-based streaming optimization.

6. CONCLUSIONS

We have investigated the rate-distortion optimized streaming at different video frame aggregation levels. We have found that the optimal scene-by-scene adjustment of the FGS enhancement layer rate reduces the computational complexity of the optimization significantly compared to video frame-by-video frame optimization, while having only a very minor impact on the video quality. We also found that reducing the computational optimization effort by aggregating the video frames arbitrarily (without paying attention to the scene structure) tends to result in significant quality deteriorations.

ACKNOWLEDGEMENTS

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. This work has been supported in part by the National Science Foundation under Grant No. Career ANI-0133252 and Grant No. ANI-0136774 as well as the state of Arizona through the IT301 initiative, and two matching grants from Sun Microsystems.

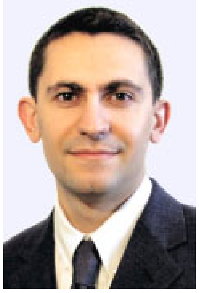
REFERENCES

1. Li W. Overview of fine granularity scalability in MPEG–4 video standard. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3):301–317.
2. Radha H, van der Schaar M, Chen Y. The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP. *IEEE Transactions on Multimedia* 2001; **3**(1):53–68.
3. Zhang Q, Zhu W, Zhang Y-Q. Resource allocation for multimedia streaming over the internet. *IEEE Transactions on Multimedia* 2001; **3**(3):339–355.
4. Bajaj S, Breslau L, Shenker S. Uniform versus priority dropping for layered video. *Proceedings of ACM SIGCOMM*, Vancouver, Canada, September 1998; 131–143.
5. Nelakuditi S, Harinath RR, Kusmierek E, Zhang Z-L. Providing smoother quality layered video stream. *Proceedings of the 10th International Workshop on Network and Operating System Support for Digital Audio and Video* (*NOSSDAV*), Chapel Hill, NC, June 2000.
6. Rejaie R, Handley M, Estrin D. Layered quality adaptation for internet video streaming. *IEEE Journal on Selected Areas in Communications* 2000; **18**(12):2530–2543.
7. Chou PA, Sehgal A. Rate-distortion optimized receiver-driven streaming over best-effort networks. *Proceedings of Packet Video Workshop*, Pittsburg, PA, April 2002.
8. Rejaie R, Reibman A. Design issues for layered quality—adaptive internet video playback. *Proceedings of the Workshop on Digital Communications*, Taormina, Italy, September 2001; 433–451.
9. Sun M-T, Reibman AR. *Compressed Video over Networks*. Marcel Dekker: New York, 2001.
10. Wu D, Hou YT, Zhu W, Zhang Y-Q, Peha JM. Streaming video over the internet: approaches and directions. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3):1–20.
11. Rejaie R, Estrin D, Handley M. Quality adaptation for congestion controlled video playback over the internet. *Proceedings of ACM SIGCOMM*, Cambridge, MA, September 1999; 189–200.
12. Saparilla D, Ross KW. Optimal streaming of layered video. *Proceedings of IEEE INFOCOM*. Israel: Tel Aviv, March 2000; 737–746.
13. Miao Z, Ortega A. Expected run-time distortion based scheduling for delivery of scalable media. *Proceedings of Packet Video Workshop*, Pittsburg, PA, April 2002.
14. Dai M, Loguinov D. Analysis of rate-distortion functions and congestion control in scalable internet video streaming. *Proceedings of the 13th International Workshop on Network and Operating Systems Support for Digital Audio and Video* (*NOSSDAV*), Monterey, CA, June 2003; 60–69.
15. Dai M, Loguinov D, Radha H. Statistical analysis and distortion modeling of MPEG-4 FGS. *Proceedings of IEEE International Conference on Image Processing*, September 2003.
16. Radha H, Chen Y. Fine–Granular–Scalable video for packet networks. *Proceedings of Packet Video Workshop*, New York, NY, April 1999.
17. Radha H, Chen Y, Parthasarathy K, Cohen R. Scalable internet video using MPEG-4. *Signal Processing*: *Image Communication* 1999; **15**(1,2):95–126.
18. Stuhlmuller KW, Link M, Girod B, Horn U. Scalable internet video streaming with unequal error protection. *Proceedings of Packet Video Workshop*, New York, NY, April 1999.
19. van der Schaar M, Radha H. Unequal packet loss resilience for fine-granular-scalability video. *IEEE Transactions on Multimedia* 2001; **3**(4):381–393.
20. Yang XK, Zhu C, Li Z, Feng GN, Wu S, Ling N. A degressive error protection algorithm for MPEG-4 FGS video streaming. *Proceedings of IEEE International Conference on Image Processing*, Rochester, NY, September 2002; 737–740.
21. Liu TM, Qi W, Zhang H, Qi F. Systematic rate controller for MPEG-4 FGS video streaming. *Proceedings of IEEE International Conference on Image Processing*, Thessaloniki, Greece, October 2001; 985–988.
22. de Cuetos P, Ross KW. Adaptive rate control for streaming stored fine–grained scalable video. *Proceedings of NOSSDAV*, Miami, FL, May 2002; 3–12.
23. de Cuetos P, Guillotel P, Ross KW, Thoreau D. Implementation of adaptive streaming of stored MPEG-4 FGS video. *Proceedings of IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002; 405–408.

24. Cohen R, Radha H. Streaming fine–grained scalable video over packet-based networks. *Proceedings of IEEE Globecom*, San Francisco, CA, November 2000; 288–292.
25. Zhao L, Kim J-W, Kuo C-CJ. Constant quality rate control for streaming MPEG-4 FGS video. *Proceedings of IEEE International Symposium on Circuits and Systems* (*ISCAS*), Scottsdale, AZ, May 2002; 544–547.
26. Zhang XM, Vetro A, Shi YQ, Sun H. Constant quality constrained rate allocation for FGS-coded video. *IEEE Transactions on Circuits and Systems for Video Technology* 2003; **13**(2):121–130.
27. Hsiao H-F, Liu Q, Hwang J-N. Layered video over IP networks by using selective drop routers. *Proceedings of IEEE International Symposium on Circuits and Systems* (*ISCAS*), Scottsdale, AZ, May 2002; I-411–I-444.
28. Zhou J, Shao H-R, Shen C, Sun M-T. Multi-path transport of FGS video. *Proceedings of Packet Video Workshop*, Nantes, France, April 2003.
29. Tung Y-S, Wu J-L, Hsiao P-K, Huang K-L. An efficient streaming and decoding architecture for stored FGS video. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; **12**(8):730–735.
30. Parthasarathy S, Radha H. Optimal rate control methods for fine granularity scalable video. *Proceedings of the International Conference on Image Processing* (*ICIP*), Barcelona, Spain, September 2003; 805–808.
31. Zhou J, Shao H, Shen C, Sun M-T. FGS enhancement layer truncation with minimized intra-frame quality variation. *Proceedings of the IEEE International Conference on Multimedia and Expo* (*ICME*), Baltimore, MD, July 2003; 361–364.
32. Vieron J, Turletti T, Hjnocq X, Guillemot C, Salamatian K. TCP-compatible rate control for FGS layered multicast video transmission based on a clustering algorithm. *Proceedings of IEEE International Symposium on Circuits and Systems* (*ISCAS*), Scottsdale, AZ, May 2002; 453–456.
33. Chen TP-C, Chen T. Fine-grained rate shaping for video streaming over wireless networks. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP*), vol. 5, Bangkok, Thailand, April 2003; 688–691.
34. Choi K, Kim K, Pedram M. Energy-aware MPEG-4 FGS streaming. *Proceedings of the 40th Design Automation Conference* (*DAC*), Anaheim, CA, June 2003; 912–915.
35. Liu J, Li B, Li B, Cao X. Fine–grained scalable video broadcasting over cellular networks. *Proceedings of IEEE Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002; 417–420.
36. Stockhammer T, Jenkac H, Weiss C. Feedback and error protection strategies for wireless progressive video transmission. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; **12**(6):465–482.
37. van der Schaar M, Radha H. Motion-compensation fine–granular–scalability (MC-FGS) for wireless multimedia. *Proceedings of Fourth IEEE Workshop on Multimedia Signal Processing*, Cannes, France, October 2001; 453–458.
38. van der Schaar M, Radha H. Adaptive motion-compensation fine–granular–scalability (AMC-FGS) for wireless video. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; **12**(6):360–371.
39. Xu J, Zhang Q, Zhu W, Xia X-G, Zhang Y-Q. Optimal joint source-channel bit allocation for MPEG-4 fine granularity scalable video over OFDM system. *Proceedings of the International Symposium on Circuits and Systems* (*ISCAS*), vol. 2, Bangkok, Thailand, May 2003; 360–363.
40. Chen RY, van der Schaar M. Complexity-scalable MPEG-4 FGS streaming for UMA. *Proceedings of IEEE International Conference on Consumer Electronics* 2002; 270–271.
41. Chen R, van der Schaar M. Resource-driven MPEG-4 FGS for universal multimedia access. *Proceedings of IEEE International Conference on Multimedia and Expo* (*ICME*), Lausanne, Switzerland, August 2002; 421–424.
42. Kim T, Ammar MH. Optimal quality adaptation for MPEG-4 fine–grained scalable video. *Proceedings of 22nd IEEE Annual Joint Conference of the IEEE Computer and Communications Societies* (*INFOCOM*), vol. 1, San Francisco, CA, March/April 2003; 641–651.
43. Hung B-F, Huang C-L. Content-based FGS coding mode determination for video streaming over wireless networks. *IEEE Journal on Selected Areas in Communications* 2003; **21**(10):1595–1603.
44. Lie W-N, Tseng M-Y, Ting I-C. Constant-quality rate allocation for spectral fine granular scalable (SFGS) video coding. *Proceedings of the IEEE International Symposium on Circuits and Systems* (*ISCAS*), vol. 2, Bangkok, Thailand, May 2003; 880–883.
45. Ugur K, Nasiopoulos P. Combining bitstream switching and FGS for H.264 scalable video transmission over varying bandwidth networks. *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and signal Processing* (*PACRIM*), vol. 2, Victoria, BC, Canada, August 2003; 972–975.
46. Xu J, Wu F, Li S. Bit allocation for progressive fine granularity scalable video coding with temporal-SNR scalabilities. *Proceedings of the IEEE International Symposium on Circuits and Systems* (*ISCAS*), vol. 2, Bangkok, Thailand, May 2003; 876–879.
47. Feng W-C. On the efficacy of quality, frame rate, and buffer management for video streaming across best-effort networks. *Journal of High Speed Networks* 2002; **11**(3,4):199–214.
48. de Cuetos P, Seeling P, Reisslein M, Ross KW. Evaluating the streaming of FGS-encoded video with rate-distortion traces. *Technical Report 2004*, Arizona State University, traces available from http://trace.eas.asu.edu/indexfgs.html.
49. Seeling P, de Cuetos P, Reisslein M. Fine granularity scalable (FGS) video: implications for streaming and a trace-based evaluation methodology. *IEEE Communications Magazine* 2005; **43**(1).

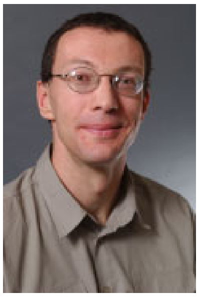
50. Seeling P, Reisslein M, Kulapala B. Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial. *IEEE Communications Surveys and Tutorials* 2004; **6**(3):58–78.
51. Rohaly AM *et al*. Video quality experts group: current results and future directions. *Proceedings of the SPIE Visual Communications and Image Processing*, vol. 4067, Perth, Australia, June 2000; 742–753.
52. Denardo EV. *Dynamic Programming*: *Models and Applications*. Prentice-Hall: Englewood Cliffs, NJ, 1982.

AUTHORS' BIOGRAPHIES



Philippe de Cuetos is a researcher with the Ecole Nationale Superieure des Telecommunications (ENST), Paris, France. He conducted his PhD research at the Institute Eurecom, Sophia-Antipolis, France, and received the PhD degree from the University of Niece, France, in 2003. His research interests are in the area of video communication, in particular in exploiting fine granular scalability for efficient network transport.

Patrick Seeling received the 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 PhD 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 IEEE and the ACM.



Martin Reissuing is an Assistant Professor in the Department of Electrical Engineering at Arizona State University, Tempe. He received the Dipl-Ing (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and the MSE degree from the University of Pennsylvania, Philadelphia, in 1996. Both in electrical engineering. He received his PhD in systems engineering from the University of Pennsylvania in 1998. During the academic year 1994–1995 he visited the University of Pennsylvania as a Fulbright scholar. From July 1998–October 2000 he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin and lecturer at the Technical University Berlin. He is Editor-in-Chief of the IEEE Communications Surveys and Tutorials and has served on the Technical Program Committees of IEEE Infocom, IEEE Globecom, and the IEEE International Symposium on Computer 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 Quality of Service, video traffic characterization, wireless networking, and optical networking.

Keith W. Ross joined Polytechnic University as the Leonard Shustek Chair Professor of Computer Science in January 2003. Before joining Polytechnic University, he was a professor for 5 years in the Multimedia Communications Department at Eurecom Institute in Sophia Antipolis, France. From 1985–1997, he was a professor in the Department of Systems Engineering at the University of Pennsylvania. He received a BSEE from Tufts University, a MSEE from Columbia University, and a PhD in Computer and Control Engineering from The University of Michigan. Professor Ross has worked in stochastic modelling, QoS in packet-switched networks, video streaming, video on demand, multi-service loss networks, web caching, content distribution networks, peer-to-peer networks, application-layer protocols, voice over IP, optimization, queuing theory, optimal control of queues, and Markov decision processes. He is an associate editor for IEEE/ACM Transactions on Networking. Professor Ross is co-author (with James F. Kurose) of the best-selling textbook, Computer Networking: A Top-Down Approach Featuring the Internet, published by Addison-Wesley. He is also the author of the research monograph, Multiservice Loss Models for Broadband Communication Networks, published by Springer. From July 1999 to July 2001, Professor Ross founded and led Wimba, an Internet startup which develops asynchronous voice products.