Adaptive Motion-Compensation Fine-Granular-Scalability (AMC-FGS) for Wireless Video

Mihaela van der Schaar and Hayder Radha, *Senior Member, IEEE*

***Abstract*—Transmission of video over wireless and mobile networks requires a scalable solution that is capable of adapting to the varying channel conditions in real-time (bit-rate scalability). Furthermore, video content needs to be coded in a scalable fashion to match the capabilities of a variety of devices (complexity scalability). These two properties—bit rate and complexity scalability—provide the flexibility that is necessary to satisfy the "Anywhere, Anytime, and Anyone" network paradigm of wireless systems. Meanwhile, MPEG-4 fine-granular-scalability (FGS) has been introduced as a flexible low-complexity solution for video streaming over heterogeneous networks (e.g., the Internet and wireless networks). FGS is also highly resilient to packet losses. However, the flexibility and packet-loss resilience associated with the FGS framework come at the expense of decreased coding efficiency compared with nonscalable coding. In this paper, a novel scalable video-coding framework and corresponding compression methods for wireless video streaming is introduced. Building on the FGS approach, the proposed framework, which we refer to as adaptive motion-compensation FGS (AMC-FGS), provides improved video quality of up to 2 dB. Furthermore, the new scalability structures provide the FGS framework with the flexibility to provide tradeoffs between resilience, higher coding efficiency and terminal complexity for more efficient wireless transmission.**

***Index Terms*—Bit-rate scalability, complexity scalability, FGS, scalable video, universal multimedia access, wireless video.**

I. Introduction

IN EMERGING wireless communication applications, multimedia data will be streamed over various access networks (GPRS, UMTS, WLANs, etc.) to a multitude of devices (PCs, TVs, PDAs, cellular phones, etc.) having different resource capabilities (display size, processing power, hardware support, memory, etc.). Hence, the transmission of multimedia data over wireless channels will need to cope with unpredictable bandwidth variations due to heterogeneous access-technologies of the receivers (e.g., 3G, 802.11a, 802.11b, etc.) or due to dynamic changes in network conditions (e.g., due to interference, co-existence). Moreover, the same multimedia data can be accessed by a large number of users/clients at any time, and from anywhere. This networking access paradigm is often referred to as Universal Multimedia Access (UMA). In the UMA framework, multimedia data is streamed from the network depending on the following three parameters: user preferences, communication channel characteristics, and device capabilities.

Recently, several scalable coding methods have been successfully proposed for video transmission through heterogeneous networks (see for example [1]–[8]). One of these techniques is the MPEG-4 fine-granular scalability (FGS) scheme [5], [8], that can adapt in real-time (i.e., at transmission time) to the bandwidth variations over heterogeneous networks and to the terminal capabilities, while using the same pre-encoded stream. Some key advantages of the MPEG-4 FGS framework are its packet-loss resilience and flexibility in supporting streaming applications [9]. Naturally, these properties come, in general, at the expense of video quality. In [8], FGS performance was compared with that of nonscalable streams coded at discrete bit rates covering the same bandwidth range. The results obtained showed up to 2–3-dB reduction in video quality (when comparing FGS with nonscalable coding) for certain sequences that exhibit high temporal correlations among successive frames. This is mainly due to the lack of motion compensation in the FGS enhancement layer (EL). Hence, for sequences that have a high-degree of motion and large number of scene cuts (e.g., "MTV" like sequences or high-action scenes), FGS performance is comparable to the performance of nonscalable coding. Moreover, in [8], FGS has also been compared to "traditional" SNR coding with multiple layers (without motion-compensation within the EL) [10], as employed in MPEG-2 or MPEG-4, and the results indicated that FGS outperforms multilayer (discrete) SNR coding with several decibels over a wide range of bit rates due to its adaptive and effective bitplane coding technique and the lack of overhead associated with introducing a new EL.

Consequently, to adapt the FGS scheme that was originally designed for Internet video transmission to wireless UMA, both 1) the performance of the MPEG-4 FGS coding method needs to be improved to allow for more efficient transmission over bandwidth-limited wireless networks and 2) a framework needs to be developed that allows addressing the bandwidth- and complexity-scalability requirements for UMA in a joint manner. In this paper, such an *adaptive* video-coding framework that addresses both the quality and scalability issues is presented. Building upon the FGS approach, the proposed framework provides improved coding efficiency of up to 2 dB. This improvement is based on incorporating some prediction within the original FGS structure in an adaptive manner. Our proposed solution is based on new FGS-based scalability struc-

Manuscript received August 20, 2001; revised April 15, 2002.

M. van der Schaar is with Philips Research Labs USA, Briarcliff Manor, NY 10510 USA (e-mail: Mihaela.vanderschaar@philips.com).

H. Radha is with the College of Engineering, Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: radha@egr.msu.edu).

Publisher Item Identifier 10.1109/TCSVT.2002.800319.

tures. These new structures provide the FGS framework with the flexibility to provide easy tradeoffs between higher coding efficiency, bandwidth scalability, and terminal complexity for more efficient wireless transmission.

Several mechanisms for improving "traditional" SNR scalability in predictive coding by exploiting EL information have already been proposed. For instance, in [2], the current EL frame is predicted from the motion-compensated reconstruction of the previous EL frames. However, in this case, the EL does not exploit the current base-layer residual information. An improved technique is proposed in [11] that uses an estimation-theoretic framework to optimally compute the prediction for the current frame given the past EL reconstruction (like in [2]) *and* the base-layer parameters and variables, including the base-layer reconstruction and quantization interval. By optimally selecting the prediction to be used for the current frame, the scalable coding performance can be improved up to several decibels [11] compared with the methods using predictions from only the base-layer [10] or the EL [2]. This method was also successfully employed in [11] for multilayer scalable coders and can be easily extended to FGS coding. Nevertheless, the main disadvantage of this method resides in its complexity, since multiple motion-compensation loops are necessary for the EL coding.

Several alternative multiple-layer techniques have also been proposed to exploit further temporal redundancies within the FGS framework [12], [13]. One such technique is the progressive FGS (PFGS) method introduced in [12]. PFGS employs additional motion-compensation loop(s) for the P and B EL frames in order to improve the performance of the FGS framework. Our solution differs from PFGS in two ways. First, we present two simplified scalability structures that are suitable for low-complexity wireless devices, and second, we propose an adaptive framework that switches between the original FGS and the newly proposed structures based on the sequence characteristics, channel conditions and/or allowed device complexity. Although one of our scalability structures can be considered as a simplified version of the PFGS method proposed in [12], we believe that the combination of complexity-scalability/complexity-reduction with quality improvement, which characterizes our proposed solution, is more flexible for new and emerging wireless and mobile networks. Hence, the novelty of our solution resides in its adaptability to the sequence characteristics, channel conditions, and/or device complexity.

The scalability structures proposed here can considerably improve the performance of the FGS framework while preserving most of the flexibility and attractive characteristics typical to the "basic" FGS scheme.[1] We refer to our proposed scalability structures as motion-compensation FGS (MC-FGS) [14], [15]. Moreover, our MC-FGS based structures can serve two classes of wireless receivers. One is tailored for relatively powerful devices (e.g., laptops) connected to wireless LANs. The second MC-FGS solution is for a lower complexity solution that is more suitable for "thin" devices. Each of these solutions has their own advantages and disadvantages as described below. For the remainder of this paper, we assume that our proposed adaptive solution can be supported by a) an FGS-based encoder, which has access to some information regarding the bandwidth variation and complexity of devices served by a particular wireless network (e.g., a wireless LAN), or b) a server residing at the interface between the wired and wireless segment of an end-to-end streaming service and which has access to a "basic" FGS stream. Consequently, our proposed solution could be realized by modifying a standard FGS stream at the "interface" server. This server can either real-time stream or store the modified (still scalable) stream locally for further viewing by the different devices that are capable of accessing the server over the wireless network.

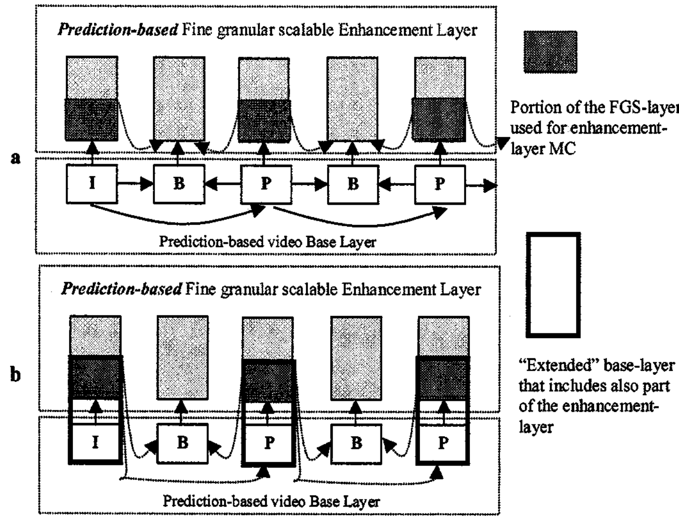

Fig. 1. MC-FGS scalability structures. (a) Two-loop MC-FGS. (b) Single-loop MC-FGS.

The remainder of the paper is organized as follows. In Section II, the MC-FGS scalability structures are described with the corresponding encoder algorithms and architectures. We also describe a high-level (heuristic) adaptive MC-FGS based algorithm that employs the MC-FGS scalability structures. In Section III, simulation results for the improvements in video quality obtained with the newly proposed algorithms are presented. In Section IV, we outline the conclusions and discuss the suitability of adaptive MC-FGS for wireless networks. Finally, in the Appendix, an analysis of the FGS coding penalty associated with the high flexibility in adapting to the bandwidth variations is provided. In particular, the Appendix shows results demonstrating a clear relationship between the temporal correlation in a sequence and the amount of penalty that standard FGS suffers in video quality.

II. MC-FGS Structure

In this section, two extensions to the "basic" FGS scheme are presented (see Fig. 1), both of which introduce MC within the FGS EL. Below, we describe these two approaches and highlight their advantages and drawbacks.

A. Two-Loop MC-FGS for B Frames

Algorithmic Description: The proposed MC-FGS framework portrayed in Fig. 1(a) introduces a MC loop within the FGS EL to exploit the remaining temporal correlation within this layer. For simplicity and to limit the bit-rate overhead associated with a second MC loop, the MC loop within the FGS-layer re-uses the base-layer motion-vectors and prediction

[1]Subsequently, we refer to the MPEG-4 FGS scheme, which has no motion-compensation within the FGS EL, as the "basic" FGS scheme.

modes, while the resulting EL residual is compressed using the same embedded codec as in the "basic" FGS scheme. As illustrated in Fig. 1(a), not all (only up to $R_{\max}$[2]) FGS bitplanes are part of the EL MC-loop. The number of bitplanes, M, included in the FGS MC-loop is chosen by trading off:

1) coding gain (i.e., a large M value potentially leads to improved temporal decorrelation within the FGS-layer)
2) prediction drift[3] occurring within the FGS-layer at low bit rates, when fewer bitplanes than M are transmitted/received, (i.e., a small value of M limits prediction drift).

Hence, to limit the drift incurred at low bit rates, preferably only a few FGS bitplanes are included in the EL MC-loop, e.g., $M = 2$–3 bitplanes. (For improved performance, a different value of M can be chosen for each frame.) The MC prediction within the FGS-layer is restricted to B frames because a relatively high coding gain is obtained while preserving many of the "basic" FGS structure benefits. For instance, under this framework, if a number of bitplanes lower than M is transmitted for the I and/or P EL frames, the prediction drift is confined to the B enhancement frames.

Furthermore, this MC-FGS structure achieves a higher coding gain (see Section III) than the "basic" FGS scheme due to its superior temporal decorrelation for the B enhancement frames. Since B frames account for 66% of the total EL bit rate[4] in an IBBP GOP structure, the loss in quality associated with restricting the EL MC to the B frames is limited for most sequences. Another reason why eliminating the MC between the P enhancement frames has only a limited effect on coding efficiency resides in the less accurate MC prediction of the P enhancement frames, which have a larger distance to their reference frames than the B frames. Hence, we include in the two-loop MC-FGS only the B enhancement frames.

From the resilience perspective, since the two-loop MC-FGS is restricted to the B enhancement frames, packet-losses occurring in I or P EL frames will not propagate beyond the EL B frames. Thus, subsequent EL P frames within the GOP remain unaffected. Moreover, since the base-layer B frames remains unaffected, the EL B frames can be easily interpolated from the correctly received EL reference frames or can be entirely discarded. This represents one of the advantages of this simplified MC-FGS structure when compared with a more elaborate scheme such as the one proposed in [12].

The number of B frames between two P frame in the described two-loop MC-FGS structure strongly influences both its coding (PSNR) performance and resilience. At low transmission bit rates, a large number of B frames results in a decreased PSNR performance due to prediction drift and reduced resilience since multiple B EL frames are affected and the concealment is not very effective since the used EL reference frames are at a larger distance from the current frames. At high transmission bit rates, a large number of B frames results in increased performance since more frames take advantage of the EL prediction.

[2]$R_{\max}$ and $R_{\min}$ are the maximum and minimum bit rate available over the network at all times.

[3]Prediction drift occurs within the FGS EL if $K < M$ bitplanes are transmitted/received because the references used for MC prediction within the FGS-layer are different at the encoder and decoder. At the encoder prediction is based on M-bitplanes of the reference, while at the decoder it is based on only K-bitplanes.

[4]In the "basic" FGS framework, a simple, yet efficient rate-control can be performed by allocating an equal number of bits to all EL frames ($I/P/B$), since there is no MC in this layer [7].

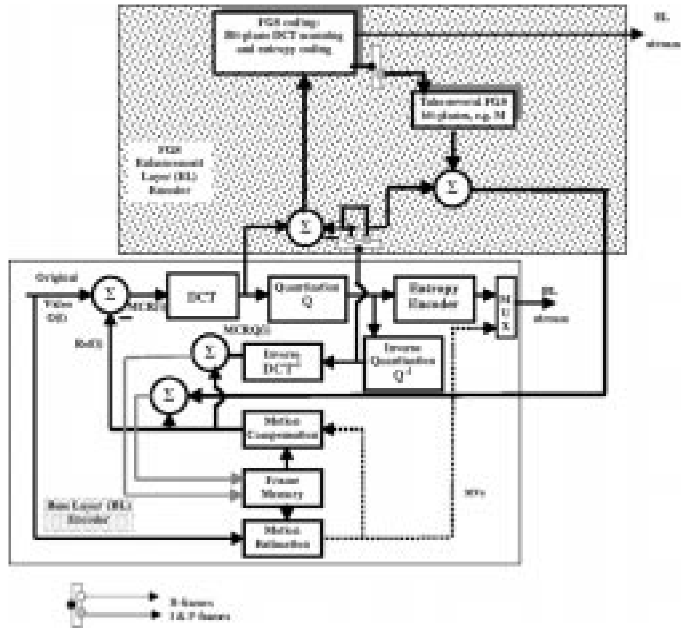

Fig. 2. Block diagram of two-loop MC-FGS encoder.

Complexity Discussion: The encoder block-diagram of the two-loop MC-FGS is depicted in Fig. 2. From a complexity perspective, it is important to notice that the base-layer structure remains unchanged. For the I and P frames, there is no MC within the EL and hence, for these frames, the complexity is also the same as for the "basic" FGS framework. An additional EL MC loop is introduced for the B frames, at both the encoder and decoder. For the EL B frames, the signal coded in the FGS-layer of the ith frame is

$$\begin{aligned}&\mathrm{MCFGS}(i,\, M)\\&\quad= \mathrm{FGSR}(i) - \mathrm{MCFGSR}(i,\, M)\\&\quad= \mathrm{MCR}(i) - \mathrm{MCRQ}(i) - \mathrm{MCFGSR}(i, M)\end{aligned}$$

where

$\mathrm{FGSR}(i)$	$= \mathrm{MCR}(i) - \mathrm{MCRQ}(i)$, as in the "basic" FGS;
$\mathrm{MCFGSR}(i,\, M)$	ith frame MC-prediction based on the M most significant bitplanes of the FGS-layer reference frame(s);
$MCR(i)$	MC-residual of the ith frame;
$MCRQ(i)$	reconstructed MC residual of the ith frame (after quantization and dequantization).

The EL MC loop for the B frames re-uses the base-layer motion vectors and prediction modes. Furthermore, the additional complexity associated with the MC within the EL is lower than for the base-layer decoding, since the motion vectors are already decoded by the base layer and the MC is reduced to fetching the reference data and adding the residual data.

The proposed two-loop MC-FGS is less complex than the PFGS-scheme in [12], since in MC-FGS the MC is inserted within the EL only for the B frames. Furthermore, if the decoder does not have enough processing power to decode the additional EL MC loop, with MC-FGS, the EL B frames or even all of the B frames can be discarded without affecting the succeeding frames. Hence, the proposed two-loop MC-FGS allows for graceful complexity scalability in addition to bit-rate scalability.

Summarizing, the two-loop MC-FGS characteristics are as follows.

Advantages:

- improved coding efficiency compared with the "basic" FGS scheme;
- unmodified base-layer structure and complexity;
- prediction drift occurring at low bit rates is confined to the EL B frames;
- packet-losses within the EL do not propagate beyond the EL B frames.

Disadvantages:

- the coding gain is limited since the base-layer remains unchanged independent of the transmission bit rate, and thus, MC-FGS does not take full advantage of the temporal correlation among successive frames;
- higher complexity, since an additional MC-loop is added in the EL for the B frames. However, this disadvantage is compensated for by providing a device the option of resorting to single-loop decoding when needed, as explained above.

B. Single-Loop MC-FGS

Algorithmic Description: In the single-loop MC-FGS depicted in Fig. 1(b), both the base layer *and* EL are used for the *base-layer* prediction. Thus, unlike the two-loop MC-FGS, this new structure *does modify* the base-layer performance. While the base-layer coding process remains unaltered, the coding parameters (e.g., quantization-step) change due to the improved reference frames resulting from the introduction of EL data in the MC prediction loop. The fine-granular method employed by the "basic" FGS coding scheme is also used for coding the EL residual.

For the single-loop MC-FGS, we introduce the notion of an "extended base-layer" which includes integrated BL/EL data. Hence, if the transmission bandwidth drops below the rate necessary for transmitting this "extended" base-layer, the truncated "extended base-layer" data will suffer from drift until the next I frame (in the case where MC-FGS is employed for all frames). Consequently, with the proposed approach, even though prediction drift occurs at low transmission bit rates, the fine-granular scalability property is still preserved, and a decodable stream can be generated at all bit rates between the base-layer bit rate R_{BL} and the maximum bit rate $R_{\max}$. Two implementations of the proposed single-loop MC-FGS method can be envisaged.

- MC-FGS is applied for all frames, thereby achieving improved temporal decorrelation for all frames, and hence, higher coding efficiency.
- MC-FGS is restricted to the B frames, to ensure that even at low bit rates, the prediction drift does not propagate beyond subsequent base-layer P frames.

It is important to note that for the single-loop MC-FGS, there is significant coding gain that can be obtained by including all frames (i.e., not only the B frames) in the EL MC, unlike the two-loop MC-FGS frames.

From a packet-loss resilience perspective, if the one loop MC-FGS is applied only on the B frames, then losses occurring within an I or P EL frame are confined to the B frames for which this frame is used as a reference. However, compared with the two-loop MC-FGS presented above, it is important to notice that in this case, the base-layer of the B frame will also be affected. Alternatively, if all frames are included in the single-loop MC-FGS, a loss within an I or P EL frame will propagate to all BL and EL frames until the end of the GOP. This represents one of the drawbacks of the single-loop approach when compared with the two-loop framework described above.

Consequently, the GOP structure employed for coding the single-loop MC-FGS structures strongly influences both their coding performance and their resilience and resulting drift at low transmission bit rates. For the single-loop MC-FGS, the tradeoff is between

- a large GOP size, which leads to a good performance at high transmission bit rates since bit-rate costly I frames are not frequently inserted;
- a small GOP size, which leads to reduced drift effects that are accumulated at low transmission bit rates.

The effect of the GOP sizes on the single-loop MC-FGS quality can be reduced if intra-coded macroblocks are regularly coded to limit the drift.

Complexity Discussion: The single-loop MC-FGS encoder is depicted in Fig. 3 for the case MC-prediction is limited to B frames. An important advantage of the single-loop MC-FGS framework is its low implementation complexity: only a set of logical "and" operations are added to the "basic" FGS encoder and decoder. If the "extended" base-layer is used only for the prediction of B frames, an additional frame memory is necessary for both the encoder and decoder, since the references for the P and B frames prediction are different.

Summarizing, the single-loop MC-FGS characteristics are:

Advantages:

- high coding efficiency since the base-layer performance is modified to take advantage of the improved temporal decorrelation at transmission bit rates higher than the "extended" base-layer bit rate;
- low-complexity due to the single-loop structure.

Disadvantages:

- prediction drift can result at bit rates lower than the "extended base-layer" bit rate;
- packet losses occurring in the EL may affect the base-layer performance.

As mentioned above, to limit prediction drift or packet-loss propagation, the single-loop MC-FGS can be restricted to B frames at the expense of lower coding gain. Consequently, single-loop MC-FGS is especially suitable for efficient transmission through channels with few packet-losses and for

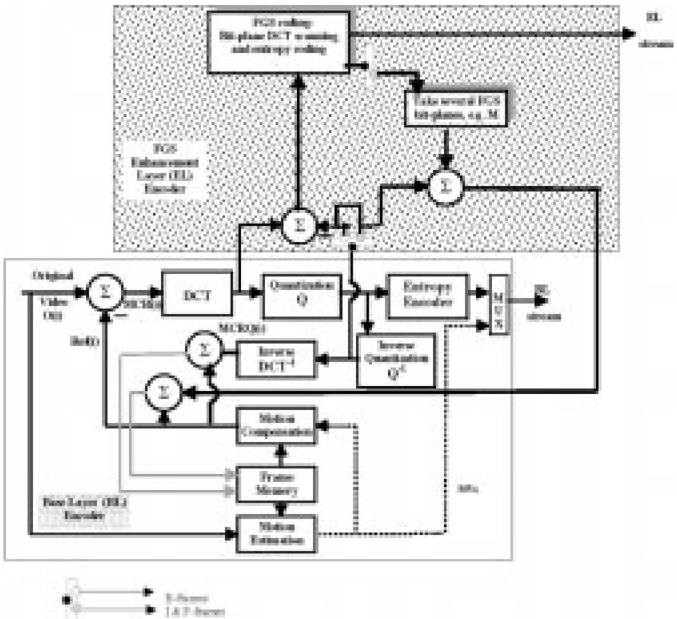

Fig. 3. Block-diagram of the single-loop MC-FGS encoder applied only for B frames.

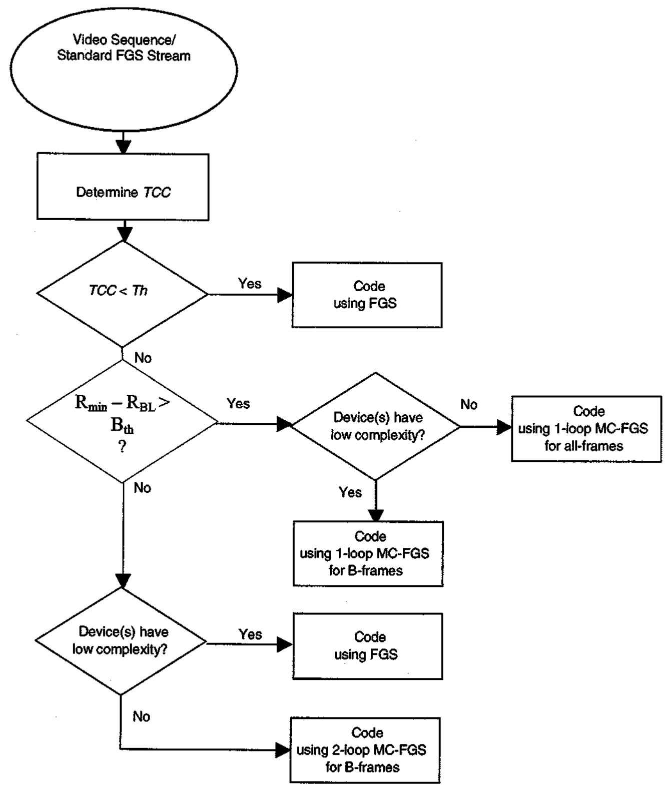

Fig. 4. Adaptive MC-FGS decision algorithm for switching between the various FGS structures.

applications requiring low-complexity decoders (or codecs in general).

Note also that since the introduction of the single-loop MC-FGS in [14], several papers [16]–[17] have reported improvements by carefully controlling the amount of drift introduced at low bit rates.

C. Adaptive MC-FGS (AMC-FGS)

The performance of the MC structures described in the previous two sections depends on the sequence characteristics as well as the network characteristics or device capabilities. Hence, for optimal streaming performance over wireless networks, an adaptive streaming system that chooses the most suitable FGS structure based on the bandwidth variations or device capabilities should be employed. The adaptation could take place either at encoding time or at a proxy within the network (e.g., at the base-station). For instance, in the case of a live broadcast, the bandwidth range of the active receivers can be determined (e.g., based on RTP reports). The bit rate of the "extended" base-layer can then be set to be lower than that of most clients, while the base-layer bit rate can be set to equal the lowest receiver bit rate. If the content is coded off-line, the switch between the various structures can take place at the proxy by transcoding between the various FGS formats. Alternatively, if multiple versions of the same content are coded using the various FGS structures, the stream that has the best performance for a particular network condition or device capability can be transmitted. Switching between the various streams can for instance be performed at an I frame.

Fig. 4 portrays a proposed decision mechanism for choosing between the various FGS structures. In the figure, TCC is a measure of the temporal correlation among successive FGS frames[5] and Th is some (TCC) threshold. (See the Appendix for more details.) These parameters (TCC and Th) are used to determine if the standard FGS structure is sufficient (from a quality perspective). If that is the case, then FGS is the clear choice due to its advantages (low complexity, high-scalability, and packet-loss resilience). Otherwise, one of the MC-FGS structures can be employed if there is a high-temporal correlation among successive frames.

The decision for using single-loop versus double-loop MC-FGS can be based on the bandwidth variation. In the figure we show a very simple example of making this decision by evaluating the difference between the minimum bit rate $R_{\rm min}$ (for all the connections) and the base-layer bit rate $R_{\rm BL}$ (e.g., of an already coded FGS stream). If the difference in bit rate is higher than some threshold $B_{\rm th}$, then single-loop MC-FGS is a viable option. Otherwise, to improve quality for higher bit-rate receivers while avoiding the disadvantages of single-loop MC-FGS, double-loop MC-FGS should be considered. It is important to note that the proposed algorithm can be made adaptive per GOP (or frame or even macroblock). It should also be noted that for more complex receivers, two-loop MC-FGS could even be replaced with the PFGS structure proposed in [12].

III. Simulation Results

In this section, the results for both proposed MC-FGS methods are given[6] for the three MPEG-4 sequences *Coastguard*,

[5]Note that TCC can be replaced by simpler methods of computing the temporal correlation like motion-vectors and texture information that are already computed for the base-layer rate-control.

[6]It is important to note that a comparison between the two-loop MC-FGS and PFGS could not performed, since neither the implementation details, nor the PFGS software are available.

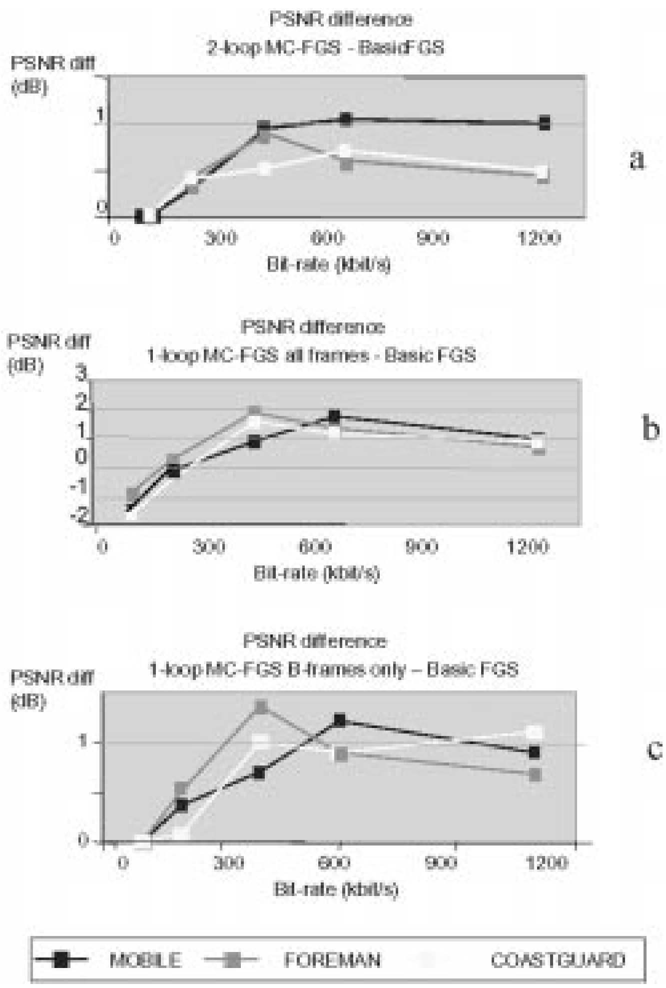

Fig. 5. Comparison between the "basic" FGS framework and proposed MC FGS structures.[7]

Foreman and *Mobile* at CIF-resolution, 10 Hz. As mentioned previously, the GOP structure employed for coding the three MC-FGS structures strongly influences their coding performance. In the results shown in this section, the same GOP structure with $N_{\rm GOP} = 21$ and $M_{\rm GOP} = 3$ has been used for all structures. The results have been generated for a base-layer bit rate $R_{\rm BL}$ equal to 100 kbits/s and $R_{\max}$ equal to 1 Mbits/s. For the base-layer encoding, the TM5 rate control has been used, while the bit-rate allocation for the "basic" FGS scheme allocated a fixed number of bits to each EL frame. However, more sophisticated algorithms can be employed for the FGS EL. Note that the EL rate-control is especially important in optimizing the performance of the new MC-FGS structures. However, even very simple rate-control mechanisms can result in good rate-distortion performance. For example, in this section, a simple yet efficient approach is used, which allocates the total bit-budget $B_{\rm tot}$ of a GOP according to

$$Btot = bI * No_{I\_\rm frames} + bP * No_{P\_\rm frames} + bB * No_{B\_\rm frames}$$

[7]To compare the performance gap between single-layer and FGS with that between single-layer and the new MC-FGS structures, the reader is referred to the Appendix, where such results are given for various sequences or to [8].

TABLE I
PSNR IN DECIBELS AS A FUNCTION OF THE BIT RATE AND NUMBER OF BITPLANES M INCLUDED IN THE MC OF THE SINGLE-LOOP MC-FGS FOR THE "FOREMAN" SEQUENCES. (THE HIGHEST PSNR VALUE FOR A PARTICULAR BIT RATE IS HIGHLIGHTED)

Bit-rate (kbit/s)	*M=0*	*M=1*	*M=2*	*M=3*
100	**29.1**	28.6	28.2	27.6
200	30.6	**30.9**	30.8	30.1
400	32.7	33.1	**34.6**	33.5
600	34.7	35.0	**36.2**	35.7
1100	38.2	38.2	38.9	**39.7**

with $No_{I\_\rm frames}$, $No_{P\_\rm frames}$, $No_{B\_\rm frames}$ being the number of I, P and B frames, respectively, within a GOP and bI, bP, bB being the bit budgets for the various frame types with $bI > bP > bB$. As mentioned above, the overall quality can be varied by employing a different number of EL bitplanes within the MC loop. In Fig. 5(a), the PSNR differences between the proposed two-loop MC-FGS and the "basic" FGS scheme are plotted. Similarly, Fig. 5(b) and (c) portray the PSNR difference between the single-loop MC-FGS structures for all frames and B frames, respectively.

At bit rates higher than the "extended" base-layer rates, the results of the single-loop MC-FGS structure outperform that of the two-loop MC-FGS. This is mainly because in the single-loop MC-FGS, the reference frames are also improved and thus the base-layer frames take advantage of the improved temporal decorrelation. However, it is important to notice that at very low bit rates (up to 1.5 $R_{\rm BL}$ dependent on the sequence), single-loop MC-FGS for all frames results in relatively poor quality due to prediction drift. For single-loop MC-FGS restricted to B frames, the drift is limited since in this case only the B frames are affected. To limit the drift occurring at low bit rates, fewer bitplanes can be inserted in the MC loop. However, this comes at the expense of reduced coding efficiency for the higher bit rates.

From the previous plots, it can be noticed that the PSNR gain obtained by the various MC-FGS methods depends on the sequence characteristics: for sequences with a high temporal correlation between frames (e.g., "Mobile") the gain is higher than for those with less temporal correlation (e.g., "Coastguard"). Fortunately, the sequences for which the proposed MC-FGS schemes have a high coding efficiency are precisely those sequences for which the "basic" FGS had a poor R-D performance (see the results in the Appendix). Consequently, the proposed MC-FGS structures reduce the FGS coding gain penalty compared to the nonscalable codecs to less than 1 dB for most sequences.

In the results portrayed in Fig. 5, the number of bitplanes M included in the MC-loop of the single-loop MC-FGS equals 2. To better understand the relationship between M and the resulting prediction-drift at low bit rates versus the quality improvement at high bit rates, the PSNR results are given in Table I for the "Foreman" sequence at various bit rates. Note that at low bit rates, the "basic" FGS ($M = 0$) gives the best PSNR performance, while the single-loop MC-FGS exhibits drift that increases with M. As the bit rate increases, the single-loop MC-FGS starts to outperform the "basic" FGS performance and thus a larger value of M can be chosen. The results shown in Table I indicate that indeed M should be chosen to allow for the best overall performance across all transmission bit rates.

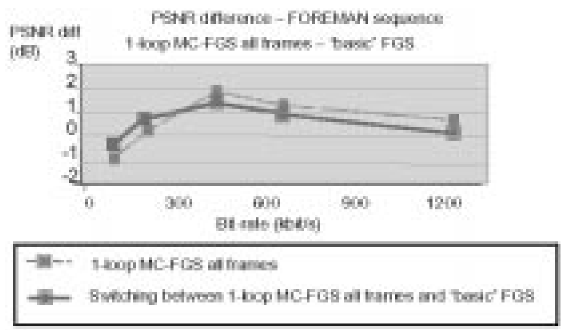

Fig. 6. Drift reduction by alternating the "basic" FGS framework and single-loop MC FGS for all frames based on temporal correlation.

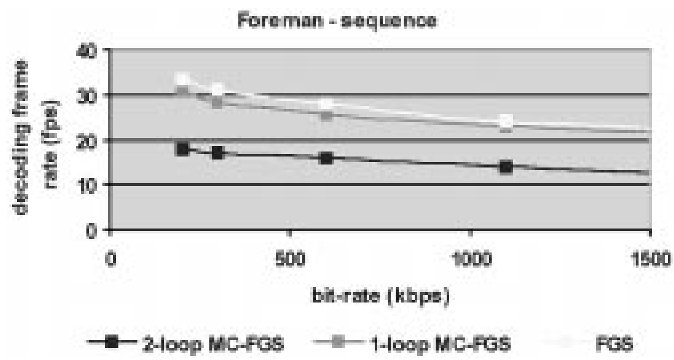

Fig. 7. Complexity of the various FGS structures.

To limit the drift of single-loop MC-FGS for all frames while preserving its gain at higher bit rates, a different number of EL bitplanes should be incorporated in the MC prediction loop of single-loop MC-FGS coding based on the temporal correlation within the sequence (see Section II). For sequences with a high temporal correlation, the single-loop MC-FGS scheme should used to exploit this redundancy, while for sequences with limited temporal correlation, the "basic" FGS scheme with no EL MC should be employed. This mechanism has been employed for the "Foreman" sequence that exhibits moderate temporal correlation in the beginning and high temporal correlation toward the end of the sequence. The results are portrayed in Fig. 6. It can be seen that by switching between the single-loop MC-FGS scheme for all frames and the "basic" FGS scheme, the best tradeoff between high performance gain at high bit rates and reduced drift at low bit rates can be achieved.

Based on the previously described analysis, it can be determined that MC-FGS should be employed only for sequences with a high temporal correlation. For these sequences, the coding penalty for FGS is high, and employing MC can significantly improve the coding efficiency. Moreover, since the sequences exhibit high temporal correlation, temporal error-concealment techniques become more efficient, and thus, the reduced resilience of MC-FGS is of lesser importance, since it can be compensated by efficient concealment. Furthermore, for these sequences, the prediction drift introduced at lower bit rates is also less visible. For sequences with low temporal correlation, MC-FGS is not necessary because the temporal correlation has already been exploited at the base-layer level and the FGS coding penalty is very limited. Hence, the tradeoff between the FGS and MC-FGS structures can be made dependent on the temporal correlation of the sequence to be coded.

From a complexity perspective, two-loop MC-FGS is the most complex, followed by single-loop MC-FGS as depicted by Fig. 7, where the complexity is expressed as the number of frames that can be decoded using the various structures on a 300-MHz Pentium PC.[8] Thus, depending on the mobile-terminal's capabilities, different streams can be transmitted for optimal performance. Also, from Fig. 7 it can be established that FGS structures exhibit complexity scalability as well as bit-rate scalability, i.e., the decoding complexity decreases with the transmitted bit rate.

[8]While the performance analysis was obtained using a high complexity PC, the results obtained are indicative for both the complexity scalability and the difference in complexity between the FGS structures. However, to determine the exact complexity levels of the FGS structures for various wireless terminals, a more in-depth complexity analysis for specific platforms is required.

IV. Conclusions and Discussion

In this paper, we presented a new adaptive MC-FGS coding scheme that is able to fulfill the "Anywhere, Anytime, and Anyone" wireless transmission paradigm, due to its ability to adapt to large variations in network characteristics and to the resource heterogeneity between devices. The proposed scheme adapts among the basic FGS structure and two novel MC-FGS approaches based on the sequence characteristics, network characteristics and device capabilities and is thereby able to achieve a large gain in coding efficiency (up to 2-dB improvement in quality) compared with the basic FGS scheme. The proposed MC-FGS coding structures obtain an improved coding efficiency by introducing MC within the FGS EL. Furthermore, it should be mentioned that the performances of the proposed MC-FGS schemes could be further improved by employing more sophisticated rate-control mechanisms. Alternatively, the results can also be improved by switching between the various AMC-FGS structures on a macroblock basis, rather than on a frame level. For instance, in [18] it has been shown that the performance of PFGS can be improved if the references used for prediction in the EL are selectively determined on a macroblock basis. Nevertheless, switching at the macroblock level requires transmitting the prediction structure for each macroblock unless the decision can be solely based on base-layer information known at both the encoder and decoder before the transmission of the FGS EL. Also, switching the prediction structure at the macroblock level increases the complexity of the AMC-FGS structure significantly. Moreover, the macroblock-adaptive decision mechanism leads to a (significantly) more complex encoder. The adaptive MC-FGS scheme provides increased flexibility in customizing the FGS framework for a particular application. For example, with the addition of these MC-FGS structures, tradeoffs can be easily made between coding efficiency, robustness to packet-losses and computational complexity. For improved coding gain, single-loop MC-FGS for all frames can be used for higher

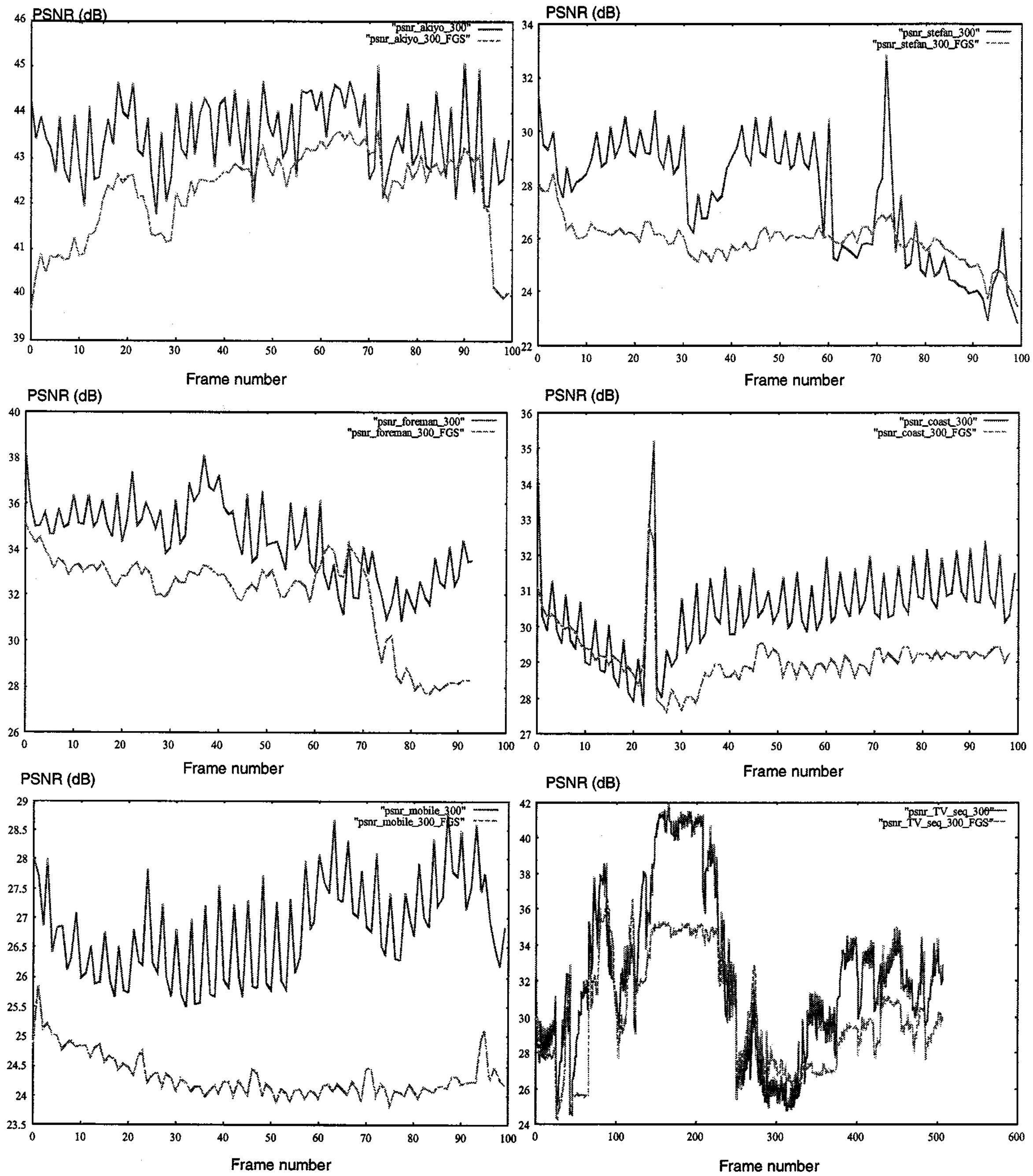

Fig. 8. PSNR values per frame for various sequences coded with a nonscalable coder at 300 kbits/s and coded with FGS with a 100-kbits/s base layer and a 200-kbits/s EL.

bandwidth wireless transmission to "thin" clients. Alternatively, for low-bandwidth wireless channels, either the basic FGS or the two-loop MC-FGS scheme for B frames can be employed depending on the sequence characteristics and device capabilities.

The adaptive MC-FGS framework has several advantages for wireless transmission besides the adaptation to bandwidth variations, robustness to losses in the wireless channel and complexity scalability to the various mobile devices.

1) *Adaptive QoS Management*: For video applications like real-time video streaming and video-conferencing, a very strict set of QoS requirements needs to be satisfied. Hence, a certain QoS level needs to be negotiated for each individual session (connection), called a service level agreement (SLA). A certain SLA can define service parameters like precedence, reliability, delay and throughput [19]. Using these SLA parameters, the mobile terminal (MT) negotiates a specific SLA with the network, or chooses from a set of available SLAs when initiating the session.

However, any number of conditions can cause the network to be unable to meet the QoS agreement. These conditions are primarily the prevailing channel conditions, mobility and transmission conditions that are external to the wireless network (i.e., congestion within the Internet when trying to access multimedia data). This is in sharp contrast with traditional distributed applications, where a stable presence and a consistently high network quality are possible. Hence, an unqualified QoS guarantee is impossible, and it is thus the variation in QoS that forms the difference between wireless and wired networks. This implies that for wireless networks, *adaptive QoS management* that specifies a range of acceptable QoS (SLAs) rather than trying to guarantee specific values is more suitable.

The previously mentioned QoS functions can be supported very easily by the adaptive MC-FGS framework that allows for prioritized transmission, different levels of resilience, different throughputs, etc. For instance, by switching between the "basic" FGS and the single-loop MC-FGS for all frames, an increased robustness to losses can be easily traded-off against a higher coding efficiency at higher bit rates. Hence, FGS is able to cooperate with the QoS management over wireless networks to support adaptation, thereby achieving a higher video transmission quality over wireless channels.

2) *"Scalable" Proxy Caching Strategies*. In the wireless network, caches are necessary at various intermediate nodes (proxies) to guarantee the negotiated QoS for a certain payload flow. For instance, caches containing "high-priority" (i.e., base-layer) packets can be employed to ensure that if data is lost e.g., due to environmental conditions, it can be retransmitted from these intermediate nodes instead of from the source, thereby reducing the end-to-end delay.

In the current implementation, if these caches overflow, the entire cache content is preempted. If adaptive MC-FGS is used, various priorities can be assigned to each packet according to its contribution to the visual quality (ranging from base-layer packets to "extended" base-layer packets to nonmotion compensated EL packets containing the less significant bitplanes). Consequently, for effective cache management, the stored packets can be discarded depending on their priority or time-dependencies (e.g., whenever the presentation time of that packet has passed).

3) *Adaptive Channel Adaptation/Allocation*. In wireless transmission, a different number of channels can be allocated to the same MT, depending for instance, on the number of MTs within the wireless cell. Adaptive MC-FGS bitstreams can be easily partitioned into prioritized sub-streams (classes) that contribute incrementally to the resultant visual quality. These sub-streams can then be transmitted through various "traffic" channels, with various QoS guarantees. Hence, as the number of channels allocated to an MT varies dynamically, the number of sub-streams transmitted to it is smoothly and instantaneously adapted. Such an adaptation cannot be achieved with nonscalable coding schemes, since the data cannot be easily partitioned into a number of discrete sub-streams whose transmission is decided "on the fly" depending on the available number of traffic channels.

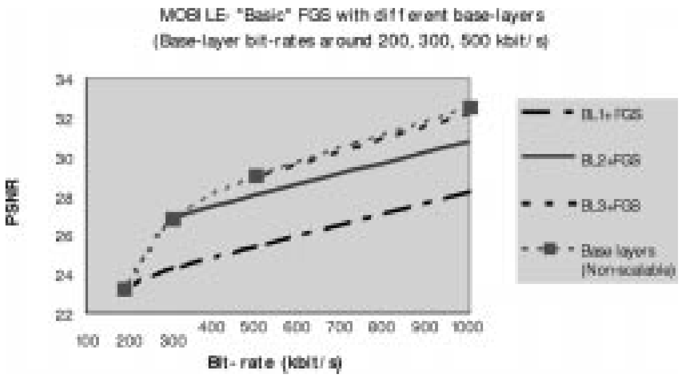

Fig. 9. "Basic" FGS performance in comparison with multiple nonscalable streams.

Furthermore, employing adaptive MC-FGS allows for easy admission control whenever a new MT enters a cell, while preserving the SLA of the existing MTs. In this case, easy joint-quality control can be performed among the various video data streams that are transmitted to the various MTs within a cell, such that the quality of the previously existing MTs is degraded gracefully as a result of admitting a new data flow. Moreover, in this manner, various priorities can be assigned to different application classes. For instance, a higher priority should be assigned to a normal phone call than to multimedia streaming. Thus, the video coding scheme employed for multimedia streaming needs to be able to adapt on-the-fly to bandwidth variations that are necessary in order to accommodate the bandwidth required for the higher priority applications.

APPENDIX
FGS PERFORMANCE ANALYSIS

While providing high flexibility in adapting to bandwidth variations and robustness to packet-losses, the "basic" FGS scheme is less efficient than a nonscalable coder operating at the same transmission bit rate [8]. To illustrate this, in Fig. 8, we plotted the performance of MPEG-4 FGS and nonscalable coding for several well-known video test sequences—*Foreman, Coastguard, Mobile, Akiyo, Stefan*—at CIF-resolution and 10 Hz.[9] Also, a longer sequence from a television broadcast—*TV_seq*—that contains several scene cuts has been added to the test set to represent "typical" content.

From Fig. 8, it can be established that the coding penalty (i.e., difference in PSNR) between MPEG-4 nonscalable coding and the MPEG-4 FGS coding (i.e., "basic" FGS structure) varies depending on the characteristics of the coded content. The explanation for these results resides in the intrinsic characteristics of the FGS framework: a limited temporal correlation (e.g., in

[9]For the results in this section, the TM5 rate-control has been adopted for coding the base-layer with a bit rate equal to 100 kbits/s. The employed GOP structures used $N_{\mathrm{GOP}} = 21$ and $M_{\mathrm{GOP}} = 3$.

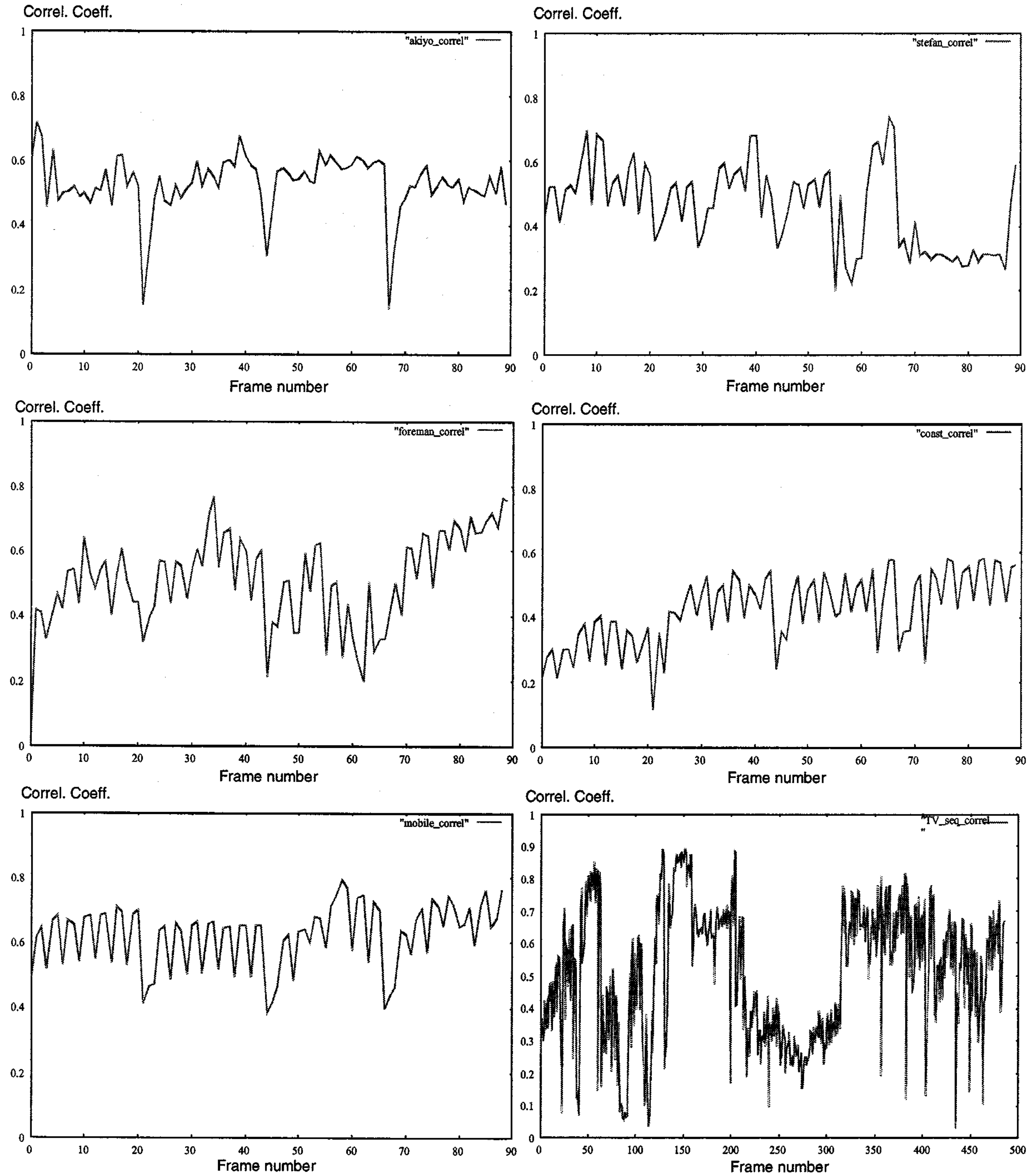

Fig. 10. TCC as a function of the frame number for various sequences.

high-motion sequences) can be easily eliminated by a base-layer coded at a low bit rate, leading to very limited coding penalty for "basic" FGS. However, for sequences with high temporal correlation, a low bit-rate base-layer cannot fully exploit this correlation, as can be seen for the *Mobile* sequence. In this case, a higher base-layer bit rate is necessary to reduce the "basic" FGS coding penalty as can be seen from Fig. 9. (Consequently, to improve the FGS performance for such sequences, the single-loop MC-FGS structure can be employed that results in a higher "extended" base-layer bit rate.)

Subsequently, we study the correspondence between the FGS coding efficiency penalty and the temporal correlation within the FGS EL in more detail. We compute the TCC, which measures the correlation coefficient between the current EL frame and the motion-compensated reference EL frame that is used as a predictor for the current frame. TCC can be determined at encoding time either in real-time or off-line, by using

$$\mathrm{TCC} = \frac{\mathrm{ABS}\left(\sum_{w=1}^{W}\sum_{h=1}^{H}(f(w,h)-\mathrm{Ave}_f)(r(w,h)-\mathrm{Ave}_r)\right)}{\sqrt{\sum_{w=1}^{W}\sum_{h=1}^{H}(f(w,h)-\mathrm{Ave}_f)^2\sum_{w=1}^{W}\sum_{h=1}^{H}(r(w,h)-\mathrm{Ave}_r)^2}}$$

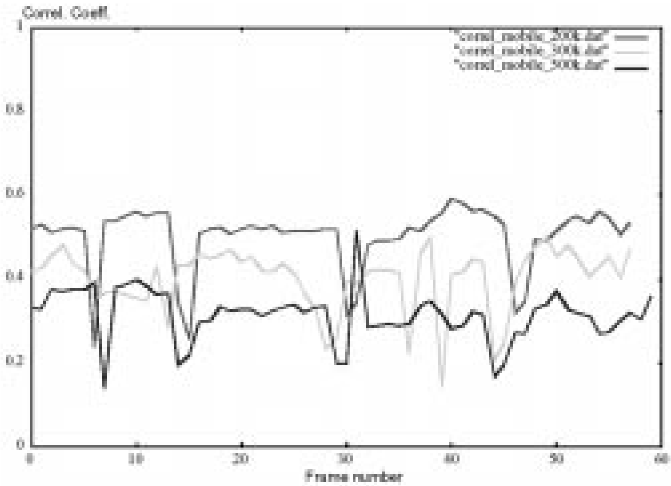

Fig. 11. Temporal correlation for the "Mobile" sequence at various bit rates.

where

ABS	absolute value function;
W, H	width and height of the frame/image, respectively;
f	current EL frame;
Ave_f	average pixel value of f;
r	motion-compensated reference EL frame for f;
Ave_r	average pixel value of r.

Fig. 10 plots the values of TCC for each frame of the various video sequences.[10] By comparing the EL temporal correlation for a certain frame in Fig. 10 with the coding efficiency penalty for that particular frame between the "basic" FGS and the nonscalable coders in Fig. 9, the following conclusions can be drawn.

1) If the temporal correlation coefficient is below a certain threshold, e.g., 0.4, then FGS can outperform the nonscalable coder performance. This is because there is no advantage associated with employing MC for higher frequencies, and because the FGS entropy coder is more efficient than the nonscalable (run, amplitude) entropy coder for the compression of the high frequency SNR residual signal. Examples where FGS outperforms the nonscalable coder are the scene change in the *Foreman* sequence, the high-motion scene in the *TV_seq* sequence, etc.).
2) Low-motion sequences have high TCCs, leading to a large coding penalty gap between FGS and the nonscalable coder.

Another interesting observation can be made from Fig. 11, where the temporal correlation for the "Mobile" sequence at several bit rates is portrayed. From Fig. 11, it can be seen that, as expected, the temporal correlation between frames decreases with increasing bit rate. This also explains why the quality difference between the nonscalable coder and FGS decreases if a higher base-layer bit rate is employed for FGS. Hence, the FGS coding penalty is high only for low-motion sequences and it is caused by a reduced exploitation of the temporal correlation within the sequence.

[10]Note that the temporal correlation varies not only from frame to frame, but also from macroblock to macroblock. Consequently, as mentioned in the Conclusions, the results of AMC-FGS can be improved by switching between the various structures on a macroblock basis, rather than on a frame level.

The above analysis indicates that the temporal correlation measured in the direction of motion, TCC, is a good indication of the coding penalty gap between the nonscalable and the "basic" FGS coders. Hence, TCC can be successfully used for:

1) determining the number of bitplanes to be employed in the "extended" base-layer;
2) switching among the AMC-FGS structures introduced in this paper to reduce the FGS quality penalty when compared with nonscalable coding.

ACKNOWLEDGMENT

The authors would like to thank the three anonymous reviewers who provided excellent and thorough feedback that improved the quality of this paper.

REFERENCES

[1] B. Girod and N. Farber, *Wireless Video*. New York: Marcel Dekker, 2000, pp. 465–511.
[2] U. Horn, K. W. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Commun.*, Sept. 1999.
[3] W. Tan and A. Zakhor, "Real-time internet video using error resilient scalable compression and TCP-friendly transport protocol," *IEEE Trans. Multimedia*, vol. 1, pp. 172–186, June 1999.
[4] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional Set Partitioning in Hierarchical Trees (SPIHT)," in *Proc. IEEE Data Compression Conf.*, Mar. 1997, pp. 251–260.
[5] S. J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, pp. 155–167, Feb. 1999.
[6] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 426–435, Oct. 1996.
[7] *Output Document MPEG-Meeting*, Study of ISO/IEC 14 496-2: 1999/FPDAM4, N3670, Oct. 2000.
[8] 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, pp. 53–68, Mar. 2001.
[9] M. van der Schaar and H. Radha, "Unequal resilience for fine-granular-scalability video," *IEEE Trans. Multimedia*, vol. 3, pp. 381–394, Dec. 2001.
[10] D. Wilson and M. Ghanbari, "Transmission of SNR scalable two layer MPEG-2 coded video through ATM networks," in *Proc. 7th Int. Workshop Packet Video*, Mar. 1996, pp. 185–189.
[11] K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Image Processing*, vol. 10, pp. 965–976, July 2001.
[12] F. Wu, S. Li, and Y. Q. Zhang, "DCT-prediction based progressive fine granularity scalability," in *Proc. ICIP 2000*, vol. 3, Oct. 2000, pp. 556–559.
[13] U. Benzler *et al.*, "Result of core experiment on fine granularity scalability for video (part 2: MC with drift)," Contribution to the MPEG-standard, m4847, July 1999.
[14] M. van der Schaar and H. Radha, "Motion compensation based fine-granular scalability (MC-FGS)," Contribution to the MPEG-standard, m6475, Oct. 2000.
[15] M. van der Schaar and H. Radha, "Motion compensation fine-granular-scalability for wireless multimedia," in *Proc. IEEE Workshop Multimedia Signal Processing*, vol. 10, Oct. 2001, pp. 453–458.
[16] A. R. Reibman, L. Bottou, and A. Basso, "DCT-based scalable video coding with drift," in *Proc. ICIP 2001*, vol. 2, Oct. 2001, pp. 989–992.
[17] S.-E. Han and B. Girod, "SNR scalable coding with leaky prediction," in *Contribution to STUDY GROUP 16 Question 6, Video Coding Experts Group (VCEG), 14th Meeting*, Sept. 2001, VCEG-N53.
[18] X. Sun, F. Wu, S. Li, W. Gao, and Y. Q. Zhang, "Macroblock-based Progressive Fine Granularity Scalable (PFGS) video coding with flexible temporal-SNR scalabilities," in *Proc. ICIP 2001*, vol. 2, Oct. 2001, pp. 1025–1028.
[19] ETSI, "Universal Mobile Telecommunications System (UMTS); QoS concept and architecture," ETSI TS 123 107 V3.5.0, Release 1999, Dec. 2000.

Mihaela van der Schaar received the M.S. and Ph.D. degrees in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1996 and 2001, respectively.

In 1996, she joined Philips Research Laboratories, Eindhoven, The Netherlands, as a Research Scientist in the TV Systems Department. From 1996 to 1998, she was involved in several projects that investigated low-cost very high quality video-compression techniques and their implementation for TV, computers, and camera systems. Since November 1998, she has been with Philips Research, Briarcliff Manor, NY, where she is a Senior Member of Research Staff in the Wireless Communications and Networking Department. She is currently involved in the research of video-coding techniques for Internet and wireless video streaming and leads a team of researchers working on scalable video coding and streaming algorithms. Since 1999, she has been an active participant to the MPEG-4 video standard, contributing to the "Fine Granularity Scalability" tool. Her research interests include image and video coding, multimedia communications and networking, and the transmission of multimedia data over wireless and packet networks. She has co-authored more than 40 conference and journal papers in this field and holds several patents.

Hayder Radha (SM'01) received the B.S. degree (with honors) from Michigan State University at Ann Arbor in 1984 and the M.S. degree from Purdue University, Lafayette, IN, in 1986, both in electrical engineering, the Ph.M. degree, in 1991, and the Ph.D. degree from the Department of Electrical Engineering, Columbia University, New York, in 1993.

He joined the faculty at Michigan State University in 2000 as an Associate Professor in the Department of Electrical and Computer Engineering. He is also an Adjunct Professor at the City University of New York (CUNY). He was a Principal Member of Research Staff with Philips Research, Briarcliff Manor, NY, during 1996–2000, where worked in the areas of video communications and networking and high-definition television. Here, he initiated an Internet video research program at and led a team of researchers working on scalable video coding, networking, and streaming algorithms. Previously, he was with Bell Laboratories, Murray Hill, NJ, as a Member of Technical Staff, where he worked in the areas of digital communications, signal and image processing, and broadband multimedia communications (1986–1996). He served as a Co-Chair and Editor of the ATM and LAN Video Coding Experts Group of the ITU-T between 1994–1996. His research interests include image and video coding, multimedia communications and networking, and the transmission of multimedia data over wireless and packet networks. He has 25 patents in these areas (granted and pending).

Dr. Radha received the Bell Laboratories Distinguished Member of Technical Staff Award and a Research Fellow Appointment from Philips Research.