Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Vis. Commun. Image R. 16 (2005) 397–411
JOURNAL OF VISUAL Communication & IMAGE Representation
www.elsevier.com/locate/jvci

Benefits and costs of scalable video coding for internet streaming

Matthias Narroschke

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, Germany

Received 25 February 2004; accepted 25 November 2004
Available online 23 May 2005

Abstract

Benefits and costs of scalable hybrid video coding techniques are analyzed with respect to internet streaming. Temporal, spatial, amplitude scalability, and combinations as described in MPEG-4 are considered. Benefits are a reduction of the server storage capacity, a reduction of the netload for multicast delivery and a graceful degradation in case of transmission errors. Costs are an increasing netload for unicast delivery and an increasing computational expense in the decoder. The result of an evaluation shows that temporal scalability has minimum costs among all analyzed techniques. It increases the netload for unicast only marginally with no additional computational expense in the decoder. Temporal scalability provides a reduction of the server storage capacity and netload for multicast by about 30% and two steps of graceful degradation. All other known standardized and nonstandardized techniques of spatial and amplitude scalability are associated with costs that appear too high to be attractive for internet streaming. Therefore, only temporal scalability is used at the present. Some of the scalable video coding techniques may become of interest for other applications where the investigated costs are less relevant.

Keywords: Video coding; Scalable video coding; Scalability; MPEG-4

E-mail address: narrosch@tnt.uni-hannover.de.

1047-3203/$ - see front matter
doi:10.1016/j.jvcir.2004.11.010

1. Introduction

As a result of the increasing multimedia communication via the internet, streaming of audiovisual content over IP-based heterogeneous networks becomes more and more important. In this scenario, which is illustrated in Fig. 1, content is provided by a streaming server and can be streamed to one or various clients.

Streaming services can be classified by the transmission mode. Each transmission mode is described by the delivery mode and by the transmission direction. The delivery mode can be either unicast or multicast. Unicast means that the server has a separate point-to-point connection to each participating client. Multicast means that the server has one point-to-multipoint connection to all participating clients (see Fig. 2).

The transmission direction can be either unidirectional or bidirectional.

The variety of individual clients causes the challenge to provide bitstreams of different data rates simultaneously for the same content because of different connections to the network and different processing speeds of these clients. Due to the heterogeneity of the network, beside various different available channel capacities, also different transmission error behaviors have to be handled, as well as fast variations of the available channel capacity.

Scalable video coding allows the decoding of only parts of the whole bitstream. Scalable encoded data contains one so-called base layer bitstream and one or more so-called enhancement layer bitstreams. A video of low resolution can be received by

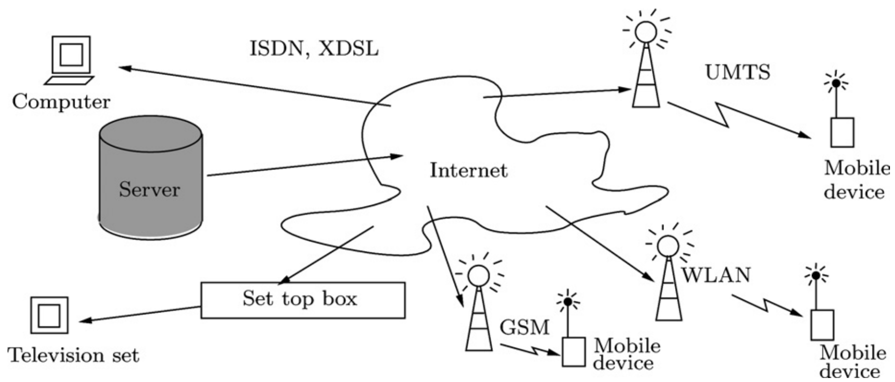

Fig. 1. Scenario of internet streaming.

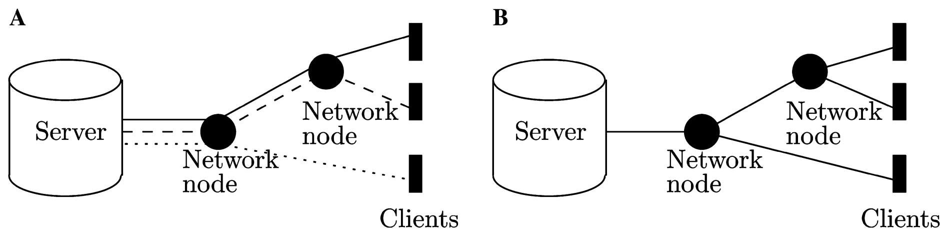

Fig. 2. Delivery modes: (A) Unicast, (B) multicast.

decoding just the base layer bitstream. The resolution of this video is reduced in amplitude, spatial, or temporal dimension. The enhancement layer bitstreams contain all additional information that is necessary for the decoding of the higher resolutions. Dependant on the transmission mode, scalable video coding offers different benefits and different costs, which are analyzed in detail in the following section.

All standardized scalable video coding techniques are based on hybrid coding. Amplitude scalability is standardized in the MPEG-2 SNR scalable profile [1]. This technique uses a two step quantization of the DCT coefficients. Coarse quantized DCT coefficients are transmitted in the base layer. Finer requantized quantization errors of these coefficients are transmitted in one single enhancement layer. In H.263 [5], the amplitude scalability is achieved by enhancement layers using hybrid coding techniques. The coding efficiency of these techniques is low [2,6]. MPEG FGS (Fine Granularity Scalability), a technique standardized in MPEG-4 [4], achieves a higher coding efficiency due to the use of bitplane coding in the enhancement layer [7].

In MPEG-2 [1], MPEG-4 [8], and H.263 [5], techniques for spatial scalability are standardized. Two or several spatial resolutions can be provided by using pyramid coding.

Temporal scalability is standardized in MPEG-2 [1], MPEG-4 [8], and H.263 [5]. It is efficiently realized by using B-Frames in the enhancement layer. They allow a hierarchical prediction order and can be omitted without any drift effect on the prediction loop.

In this paper, the benefits and costs of scalable video coding are analyzed for different transmission modes. They are evaluated for the scalability techniques as described in MPEG-4:

- AS: MPEG FGS, standardized in MPEG-4 [4]
- SS: MPEG SSP, standardized in MPEG-4 [8]
- TS: Temporal scalability using B-Frames in the enhancement layer, standardized in MPEG-4 [8].

In addition, also all combinations of these techniques are evaluated of which some are standardized. The combination SS + TS is standardized in MPEG-4 [8]. TS + AS is standardized as FGST in MPEG-4 [4]. The combinations SS + AS and SS + TS + AS are both not standardized in MPEG-4.

The paper is organized as follows: In Section 2, the benefits and costs of scalable video coding are determined for different transmission modes. Section 3 evaluates these benefits and costs for different scalability techniques. The paper closes with the conclusions.

2. Benefits and costs for different transmission modes

In this section, the benefits and costs of scalable video coding are determined for different transmission modes.

2.1. Unidirectional transmission using unicast delivery

A streaming service demanding a unidirectional transmission and the delivery mode unicast is, for instance, Video on Demand. Each client can request a video individually with an arbitrary starting time. Due to the variety of different client terminals and different networks, it is necessary to provide a multitude of bitstreams with different data rates on the server simultaneously. Using scalable video coding, bitstreams of lower data rates are implicitly included in bitstreams of higher data rates. This allows a reduction of the server storage capacity if the sum of all scalable encoded bitstreams requires less data than the sum of all nonscalable encoded bitstreams.

Temporal variations of the available channel capacity in IP networks due to packet delay and packet loss can be faster than a server's reaction. In combination with unequal error protection or routers which can assign a higher priority to packets containing the base layer than to ones containing the enhancement layers scalable video coding enables to balance these variations generating only graceful degradations. This is achieved by protecting the base layer more than the enhancement layers and by omitting enhancement layers. Without scalable coding these variations can lead to a complete image loss at the decoder. This benefit is of significant importance for clients which have very small buffers to balance these variations, e.g., some mobile devices.

The benefits are associated with costs. The introduction of scalability into a coding system increases the output data rate of the encoder. In this paper, the additional data rate is called overhead.

Furthermore, additional computational expense is needed in decoders, which process one or more enhancement layers in addition to the base layer for higher resolutions. Especially in mobile devices, low computational expense is important to save power consumption.

The increase of the computational expense in the decoder is associated with all transmission modes considered here. Therefore it is not mentioned again for each mode.

2.2. Unidirectional transmission using multicast delivery

A streaming service demanding unidirectional transmission and the delivery mode multicast is, for instance, Broadcast. The starting time of the program is the same for all bitstreams and fixed for all clients. Due to the variety of different client terminals and different networks, it is necessary to provide a multitude of bitstreams with different data rates on the server simultaneously. All bitstreams have to be sent once and at the same time. Therefore scalable video coding offers two benefits: A reduction of the server storage capacity and a reduction of the netload if the sum of all scalable encoded bitstreams requires less data rate than the sum of all nonscalable encoded bitstreams.

As well as for a unidirectional transmission using unicast delivery fast variations of the available channel capacity can be balanced by scalable video coding using a graceful degradation.

2.3. Bidirectional transmission

For bidirectional transmission unicast is generally used at the present. Streaming services demanding a bidirectional transmission are, for instance, Video Communication and Video Conferencing using an additional central conference server. The server encodes the content in real time and sends one bitstream of a specific data rate. This data rate can be continuously adapted in a fine grained way in consultation with the client. Since bidirectional communication has a low end-to-end delay, variations of the available channel capacity in IP networks can be balanced by a continuous data rate adaptation by the encoder. For this transmission mode, a benefit of scalable video coding is to enable graceful degradation if very fast variations of the available channel capacity occur. However, the introduction of scalability into a coding system is associated with an increase of the required data rate.

In the future it may be possible that also multicast delivery is used in Video Conferencing. This mode will not be considered in this paper.

2.4. Lineup of benefits and costs

Tables 1 and 2 show the lineup of the benefits and costs for all transmission modes. It can be seen that scalable video coding is especially useful for unidirectional transmission. Scalable video coding offers a reduction of the server storage capacity. If the delivery mode is multicast, also the netload can be reduced. For all investigated transmission modes, scalable video coding enables a graceful degradation if very fast variations of available channel capacity occur, which cannot be handled fast enough by automatic adaptation. Scalable video coding provides no benefits for bidirectional transmission except graceful degradation. There are two kinds of costs. The

Table 1
Benefits of scalable video coding for different transmission modes

Transmission mode		Benefits		
Transmission direction	Delivery mode	Reduction of the server storage capacity	Reduction of the netload	Graceful degradation
Unidirectional	Unicast	Yes	No	Yes
	Multicast	Yes	Yes	Yes
Bidirectional	Unicast	No	No	Yes

Table 2
Costs of scalable video coding for different transmission modes

Transmission mode		Costs	
Transmission direction	Delivery mode	Increase of the computational expense in the decoder	Increase of the netload
Unidirectional	Unicast	Yes	Yes
	Multicast	Yes	No
Bidirectional	Unicast	Yes	Yes

computational expense in the decoder is increased for all transmission modes by using scalable video coding. For transmission modes associated with the delivery mode unicast, scalable video coding causes an increase of the netload.

3. Evaluation of benefits and costs for different scalability techniques

In the previous section, it was pointed out that for bidirectional transmission the only benefit of scalable video coding is graceful degradation. Therefore the evaluation concentrates on unidirectional transmission where scalable video coding has more benefits.

For internet streaming, an encoded video must be provided at a multitude of different data rates by the server simultaneously. For access to videos over heterogeneous networks such as GSM, ISDN, XDSL, and UMTS, a multitude of data rates in the range of 9.6 kbit/s up to 2048 kbit/s should be supported. For this evaluation the target data rates listed in Table 3 are considered. The low data rates are related to multiples of 8 kbit/s for UMTS and to multiples of 9.6/14.4 kbit/s for GSM with GPRS or HSCSD. To each target data rate a specific temporal, spatial, and amplitude resolution is assigned which is illustrated in Table 3. Considered resolutions are: QCIF, CIF, and ITU-R 601 [13] as spatial resolutions, 8.3, 12.5, 25, and 50 Hz as temporal resolutions, "very low," "low," "medium," and "high" are amplitude resolutions. A high amplitude resolution means that the reconstructed samples of the video sequence contain only small quantization noise. A low amplitude resolution means that they contain high quantization noise. This assignment is based on subjective picture quality and might be slightly different in special applications. In the case of a 60 Hz time base, the temporal resolutions would be 10, 15, 30, and 60 Hz.

Table 3
Assignment of resolution levels to target data rates

Data rate [kbit/s]	Spatial resolution	Temporal resolution [Hz]	Amplitude resolution
9.6	QCIF	8.3	Low
14.4	QCIF	8.3	Medium
16	QCIF	8.3	High
19.2	QCIF	12.5	Very low
28.8	QCIF	12.5	Low
32	QCIF	12.5	Medium
43.2	QCIF	12.5	High
64	QCIF	25	Low
128	QCIF	25	High
256	CIF	12.5	Low
384	CIF	12.5	High
768	CIF	25	Low
1024	CIF	25	High
1536	ITU-R 601	25	High
2048	ITU-R 601	50	High

For example: It is assumed that the spatial resolution QCIF, the temporal resolution of 12.5 Hz, and a medium amplitude resolution are assigned to the data rate 32 kbit/s.

In the following subsections, a detailed evaluation of the benefits and costs of each scalability technique according to Tables 1 and 2 is presented assuming that encoded bitstreams for all considered data rates as listed in Table 3 are provided. The reference for this evaluation is nonscalable video coding.

3.1. Server storage capacity

Table 4 shows the calculation of the server storage capacity that is needed to provide an encoded bitstream of a video with a duration T_{video} for all data rates shown in Table 3 using different scalability techniques.

This calculation includes all overheads O_{AS}, O_{SS}, and O_{TS} in form of additional data rate for the same quality caused by the introduction of scalability into the coding system. Experimental results in [2,3,11] show that AS requires an overhead of approximately $O_{\text{AS}} \approx 80\%$ and SS of about $O_{\text{SS}} \approx 68\%$ [10] for the same PSNR. In a nonscalable coder, the data rate is depending on the number of consecutive B-Frames between I- or P-Frames. It can be drawn from Fig. 3 that the data rate is minimum if two B-Frames are applied. The corresponding data rate is used as a reference for calculating the overhead O_{TS}. For TS providing full and half frame rate an odd number of consecutive B-Frames is necessary. This enables to omit every second frame. If one B-Frame is used the overhead $O_{\text{TS},2}$ is approximately

Table 4
Server storage capacity and netload for the delivery mode multicast

	Server storage capacity in kbit		Netload for delivery mode multicast in kbit per second
No scalability	$(9.6 + 14.4 + 16 + \cdots + 2048) \cdot T_{\text{video}} =$ $6371 \cdot T_{\text{video}}$	$= S$	$R_{\text{M}} = \frac{S}{T_{\text{video}}}$
AS	$[2048 + 1536 + (1024 + 384 + 128 + 43.2 + 16)$ $(1 + O_{\text{AS}})] \cdot T_{\text{video}} = 6455 \cdot T_{\text{video}}$	$= 1.01S$	$1.01R_{\text{M}}$
SS	$(2048 + 1536 + (1024 + 768 + 384 + 256)\,(1 + O_{\text{SS}}) +$ $32 + 19.2 + 16 + 14.4 + 9.6) \cdot T_{\text{video}} = 7761 \cdot T_{\text{video}}$	$= 1.21S$	$1.21R_{\text{M}}$
TS	$[(2048 + 1024 + 768 + 128 + 64)\,(1 + O_{\text{TS},2}) + 32 +$ $19.2 + 16 + 14.4 + 9.6] \cdot T_{\text{video}} = 4204 \cdot T_{\text{video}}$	$= 0.66S$	$0.66R_{\text{M}}$
TS + AS	$(2048\,(1 + O_{\text{TS},2}) + (1024 + 128)\,(1 + O_{\text{TS},2})$ $(1 + O_{\text{AS}}) + 16 + 14.4 + 9.6) \cdot T_{\text{video}} = 4244 \cdot T_{\text{video}}$	$= 0.67S$	$0.67R_{\text{M}}$
SS + AS	$(2048 + 1536 + (1024 + 384)\,(1 + O_{\text{SS}})\,(1 + O_{\text{AS}}) +$ $16 + 14.4 + 9.6) \cdot T_{\text{video}} = 7882 \cdot T_{\text{video}}$	$= 1.24S$	$1.24R_{\text{M}}$
SS + TS	$[(2048 + 768)\,(1 + O_{\text{SS}})\,(1 + O_{\text{TS},2}) + 384\,(1 + O_{\text{SS}}) +$ $256 + 128 + 32 + 19.2 + 16 + 14.4 + 9.6] \cdot$ $T_{\text{video}} = 5946 \cdot T_{\text{video}}$	$= 0.93S$	$0.93R_{\text{M}}$
SS + TS + AS	$(2048\,(1 + O_{\text{SS}})\,(1 + O_{\text{AS}})\,(1 + O_{\text{TS},3}) + 128(1 + O_{\text{AS}})$ $(1 + O_{\text{TS},2}) + 16 + 14.4 + 9.6) \cdot T_{\text{video}} = 6840 \cdot T_{\text{video}}$	$= 1.07S$	$1.07R_{\text{M}}$

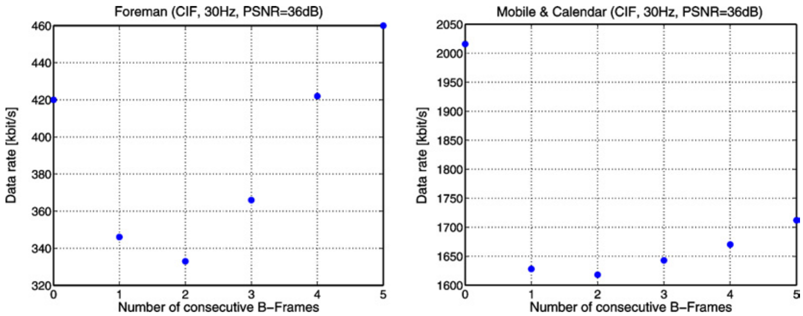

Fig. 3. Data rate versus the number of consecutive B-Frames for fixed PSNR for the testsequences *Foreman* and *Mobile* & *Calendar*.

2% in average. For providing full, half, and one quarter frame rate at least three consecutive B-Frames are required. This enables to drop every second B-Frame and also all B-Frames. The use of three consecutive B-Frames requires an additional data rate of approximately $O_{\mathrm{TS},3} \approx 6\%$ in average. It is assumed that TS provides three temporal resolutions in the combination TS + SS + AS, otherwise two.

For example, to provide encoded videos for all data rates and resolution levels as shown in Table 3 by using AS only, the video has to be encoded separately for the target data rates of 2048, 1536, 1024, 384, 128, 43.2, and 16 kbit/s requiring an overhead of O_{AS} for each scalable encoded bitstream. All other data rates are provided by AS.

3.2. Netload for the delivery mode multicast

The netload for the delivery mode multicast is proportional to the server storage capacity, because bitstreams of all data rates must be transmitted simultaneously. Thus, the netload equals the server storage capacity divided by the duration T_{video} of the video. Table 4 shows this netload as well.

3.3. Graceful degradation

Each scalability technique behaves differently in terms of graceful degradation. Table 5 shows the maximum number of steps of a graceful degradation. This maximum number of steps is only provided for clients which receive, in the case of an error free transmission, all possible enhancement layers. All other clients have fewer steps of graceful degradation. AS has the ability to allow many steps which are fine grained. In this paper, it is assumed that AS allows at least four steps. The restrictions for the number of steps result from the specifications of the evaluated scalability techniques as defined in Section 1.

For example, if SS and AS are used at least eight steps of a graceful degradation are enabled, four or more steps of AS in each of the two spatial resolutions.

Table 5
Maximum number of steps of a graceful degradation

	Maximum number of steps of a graceful degradation
No scalability	1
AS	$\geqslant 4$
SS	2
TS	2
TS + AS	$\geqslant 2 \cdot 4 = 8$
SS + AS	$\geqslant 2 \cdot 4 = 8$
SS + TS	$2 \cdot 2 = 4$
TS + SS + AS	$\geqslant 3 \cdot 2 \cdot 4 = 24$

3.4. Computational expense in the decoder

In this subsection, the costs due to the increase of the computational expense in the decoder are estimated. The costs are derived from results as worked out in [18] for the MPEG-4 ASP decoder when decoding videos in ITU-R 601 resolution encoded at data rates of 1.5 and 3.0 Mbit/s. The computational expense is measured by the required cycles per picture element (pel) related to the high resolution which is achieved by decoding all enhancement layers. It is measured in the unit [$cycles/pel_{hr}$]. Considered are the main components of each decoder, which are inverse cosine transform, motion compensation, and reconstruction of the decoded image. For spatial scalable coding, also the interpolation filtering and the weighting of the prediction signals are considered in addition to the above mentioned components. Figs. 4–6 show the estimated cycles/pel needed for the processing by the considered components of a nonscalable MPEG-4 decoder, the MPEG-FGS decoder and the MPEG-SSP decoder. The required total cycles/pel is the sum of the required cycles/pel of the components.

For example: a nonscalable decoder requires approximately 40 $cycles/pel_{hr}$ for the IDCT, 22 $cycles/pel_{hr}$ for the reconstruction including copy, add, and average operations, and 180 $cycles/pel_{hr}$ for the motion compensation, which is a total of 242 $cycles/pel_{hr}$. This total computational expense of the nonscalable decoder is referred to as C. The spatial scalable MPEG-SSP decoder requires C $cycles/pel_{hr}$ for the decoding of the high spatial resolution and in addition 10 $cycles/pel_{hr}$ for the IDCT, 45 cy-

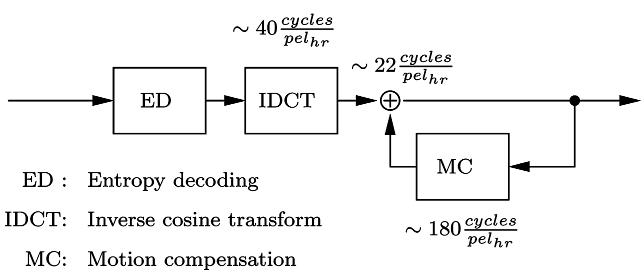

Fig. 4. Estimated cycles/pel of the components of a nonscalable hybrid decoder (MPEG-4).

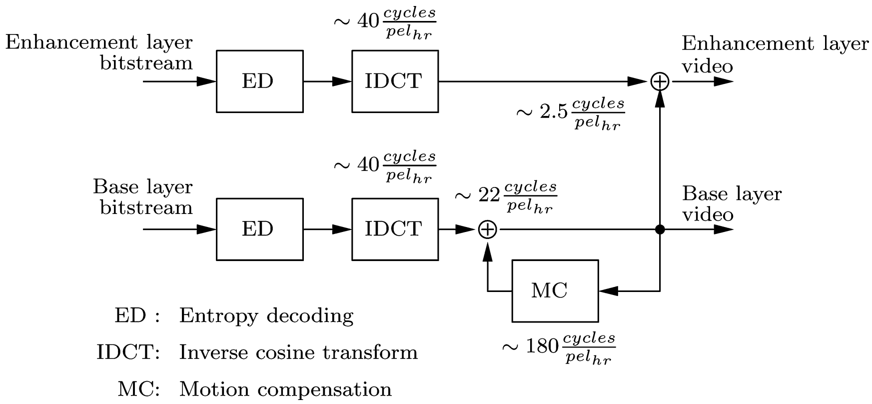

Fig. 5. Estimated cycles/pel of the components of an amplitude scalable MPEG-FGS decoder.

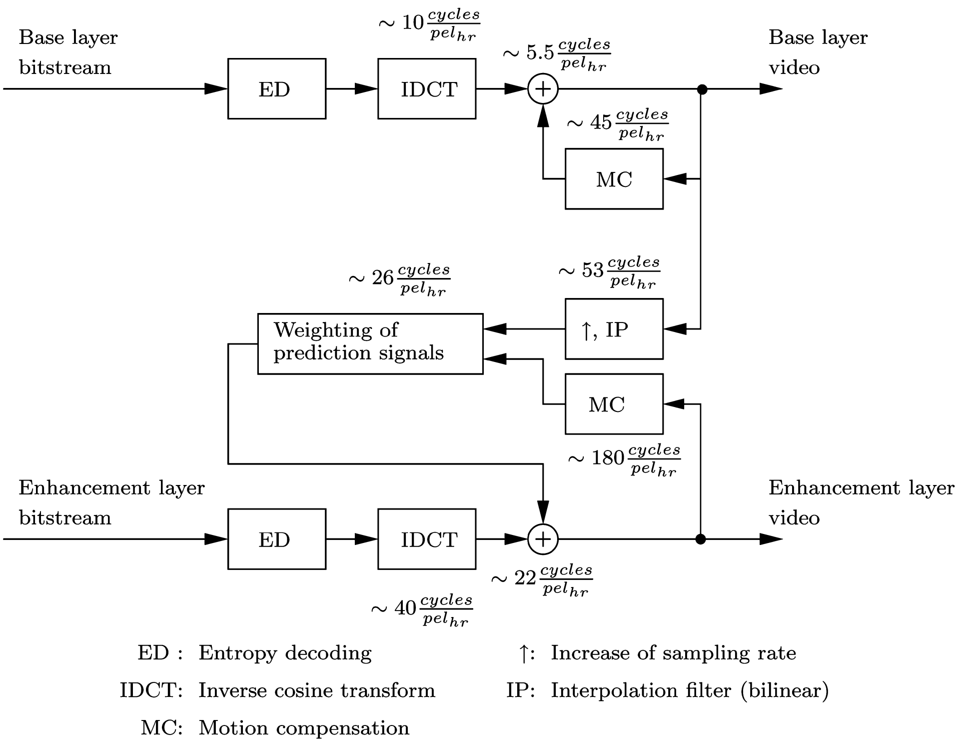

Fig. 6. Estimated cycles/pel of a spatial scalable MPEG-SSP decoder.

cles/*pel*$_{hr}$ for the motion compensation, and 5.5 *cycles*/*pel*$_{hr}$ for the reconstruction of the low spatial resolution. Furthermore, it requires 53 *cycles*/*pel*$_{hr}$ for the interpolation filtering of the reconstructed base layer signal and about 26 *cycles*/*pel*$_{hr}$ for the

Table 6
Approximated computational expense in the decoder for the worst case

	Computational expense in the decoder (worst case)
No scalability	C
AS	$1.18 \cdot C$
SS	$1.58 \cdot C$
TS	$1.00 \cdot C$
TS + AS	$1.00 \cdot 1.18 \cdot C = 1.18 \cdot C$
SS + AS	$1.58 \cdot 1.18 \cdot C = 1.86 \cdot C$
SS + TS	$1.58 \cdot 1.00 \cdot C = 1.58 \cdot C$
TS + SS + AS	$1.00 \cdot 1.58 \cdot 1.18 \cdot C = 1.86 \cdot C$

weighting of the prediction signals. Thus, the MPEG-SSP decoder requires a total of 381.5 $cycles/pel_{hr}$ which is 1.58 times that needed by the nonscalable decoder.

Table 6 shows the results of the evaluation of the computational expense in the decoder for the worst case, which means that the decoder must process all components and all enhancement layers.

3.5. Netload for the delivery mode unicast

Due to the additional data rate needed by a scalable coder compared to a nonscalable one, the data rate of each individual stream in the delivery mode unicast is increased, except of the base layer stream. Table 7 shows the netload for the delivery mode unicast for the worst case when all enhancement layers are transmitted. The reference netload R_U is the netload of a nonscalable MPEG-4 coder. For example, if SS and AS are used, overheads of O_{SS} and O_{AS} are needed. This results in a netload of $3.02R_U$.

3.6. Evaluation results

The evaluated benefits and costs are listed in Table 8. Bold numbers mark costs that are considered as not acceptable. Compared to nonscalable coding, AS has low costs concerning the increase of computational expense in the decoder (18%)

Table 7
Netload for delivery mode unicast for the worst case

Netload for delivery mode unicast (worst case)	
No scalability	R_U
AS	$R_U(1 + O_{AS}) = 1.8R_U$
SS	$R_U(1 + O_{SS}) = 1.68R_U$
TS	$R_U(1 + O_{TS,2}) = 1.02R_U$
TS + AS	$R_U(1 + O_{TS,2})\ (1 + O_{AS}) = 1.84R_U$
SS + AS	$R_U(1 + O_{SS})\ (1 + O_{AS}) = 3.02R_U$
SS + TS	$R_U(1 + O_{SS})\ (1 + O_{TS,2}) = 1.71R_U$
TS + SS + AS	$R_U(1 + O_{SS})\ (1 + O_{AS})\ (1 + O_{TS,3}) = 3.21R_U$

Table 8
Benefits and costs of scalable video coding

	Benefits			Costs	
	Reduction of		Steps of graceful degradation	Increase of	
	server storage capacity (%)	netload for multicast (%)		computational expense in the decoder (%)	netload for unicast (%)
No scalability	0	0	1	0	0
AS	−1	−1	⩾4	18	**80**
SS	−21	−21	2	**58**	**68**
TS	34	34	2	0	2
TS + AS	33	33	⩾8	18	**84**
SS + AS	−24	−24	⩾8	**86**	**202**
SS + TS	7	7	4	**58**	**71**
TS + SS + AS	−7	−7	⩾24	**86**	**221**

but high costs concerning the increase of the netload for unicast delivery (80%). SS is associated with high costs regarding to both the increase of computational expense in the decoder (58%) and to the increase of netload for unicast delivery (68%). These high costs are mirrored in all combinations in which AS and SS are involved. In Fig. 7, the increase of the computational expense in the decoder and of the increase of the netload for unicast are represented in one common diagram. It can be observed that presently TS has minimum costs among all analyzed techniques. This scalability technique allows a reduction of the server storage capacity of 34% and a reduction of the netload for multicast delivery of also 34% compared to nonscalable coding. But TS provides only two steps of graceful degradation.

To reduce the increase of netload for unicast of SS + TS, Benzler developed a spatial and temporal scalability technique based on subband coding [9]. He reduced the increase from 68% to less than 20%. But this has been achieved with an increased computational expense in the decoder of about 100%. Thus, the costs were only shifted.

Compared to TS and SS, AS offers a graceful degradation which is much finer grained due to a significant higher number of steps. Stimulated by this advantage, Benzler and Narroschke reduced the increase of the netload for unicast of AS + TS [11,12]. Benzler achieved a reduction from 84% to about 66% without additional computational expense in the decoder. Narroschke achieved a further reduction to about 50% (see Fig. 7). But his technique is associated with an increased computational expense in the decoder of 51%. Similar results are achieved for AS + TS by Wu et al. [19] and van der Schaar et al.[20].

Beside the research activities in the field of hybrid coding there are also activities in the area of 3D-Wavelet coders which can provide scalability in all three dimensions inherently, see Ohm [14], Woods [15,16], and Secker and Taubman [17]. Present techniques require an increase of the netload for unicast of just about 10–15% as shown by Hanke and Wien [21,22]. But they are associated with an increase of the computational expense in the decoder of about 100% [23]. These costs are also shown in Fig. 7. Whereas for all investigated scalability techniques which are based

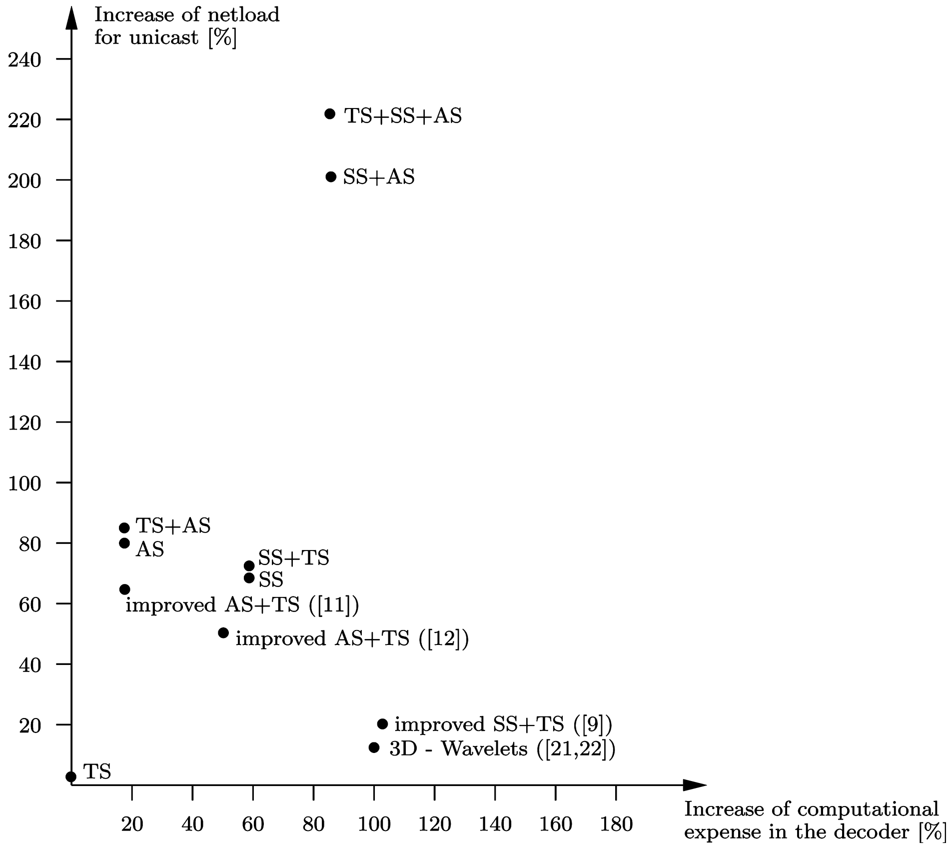

Fig. 7. Increase of computational expense in the decoder versus increase of netload for unicast for the evaluated scalability techniques.

on hybrid coding the increases of the data rate for unicast and of the computational expense in the decoder are assessed with respect to the nonscalable technique they were introduced in, the increases for the 3D-Wavelet coders are assessed with respect to the recent standard H.264/AVC [25] which is based on hybrid coding.

4. Conclusions

Benefits and costs of scalable video coding are analyzed with respect to internet streaming. Benefits are a reduction of the server storage capacity, a reduction of the netload for multicast transmission and graceful degradation in the case of transmission errors. Costs are an increasing computational expense in the decoder and an increasing netload for unicast transmission. These benefits and costs are evaluated for different standardized scalability techniques which are based on hybrid coding and for combinations of these techniques.

The result is that presently TS has minimum costs among all analyzed techniques. The server storage capacity and the netload for multicast delivery can be reduced by 34% compared to nonscalable coding. It increases the netload for unicast by just 2% with no additional computational expense in the decoder. Changes of the available channel capacity of about 35% can be balanced. But only two steps of graceful degradation are enabled by this technique. All other known standardized and nonstandardized techniques are associated with costs that are too high to be attractive for streaming applications. Thus, TS is the main technique being used at the present.

Beside scalable hybrid video coding techniques also scalable 3D-Wavelet coding techniques are developed. These techniques are associated with an increase of the netload for unicast of about 10–15% but they require an increase of the computational expense in the decoder of about 100%. These high costs are problematic for mobile receivers but might be acceptable for special applications. To encourage the development of scalable coding techniques with lower costs the ISO has recently launched a call for proposals [24].

References

[1] MPEG-2: ISO/IEC JTC1/SC29/WG11, Revised Text for ITU-T Recommendation H.262 - ISO/IEC 13818-2: Information technology—Generic Coding of Moving Pictures and Associated Audio Information: Video, ISO/IEC, Geneva, March 1995.

[2] W. Li, Y. Chen, Experiment Result on Fine Granularity Scalability, in: Contribution to 47th MPEG Meeting, Seoul, Korea, March 1999, m4473.

[3] U. Benzler, U. Pestel-Schiller, Result of Core Experiment on Fine Granularity Scalability for Video (part 2: MC with drift), in: Contribution to 48th MPEG Meeting, Vancouver, Canada, July 1999, m4847.

[4] MPEG-4 FGS: ISO/IEC JTC1/SC29/WG11, ISO/IEC 14496-2 FDAM4: Information Technology—Coding of Audio-Visual Objects—Part 2: Visual, Amendment 4: Streaming Video Profile, ISO/IEC JTC1/SC29/WG11 N3904, Pisa, January 2001.

[5] ITU-T Recommendation H.263: Video Coding for Low bit rate Communication, Version 3, November 2000.

[6] G. Côté, B. Erol, M. Gallant, H.263+: video coding at low bit rates, IEEE Trans. Circ. Syst. Video Technol. 8 (7) (1998).

[7] W. Li, Bit-plane coding of DCT coefficients for fine granularity scalablility, in: Contribution to 45th MPEG Meeting, Atlantic City, NJ, October 1998, m3989.

[8] MPEG-4: ISO/IEC JTC1/SC29/WG11, ISO/IEC 14496:2000-2: Information technology—Coding of Audio-Visual Objects—Part 2: Visual, ISO/IEC, Geneva, December 2000.

[9] U. Benzler, Spatial scalable video coding using a combined subband-DCT approach, IEEE Trans. Circ. Syst. Video Technol. 10 (7) (2000) 1080–1087.

[10] U. Benzler, Scalable multi-resolution video coding using a combined subband-DCT approach, in: Proceedings of Picture Coding Symposium, PCS'99, Portland, USA, 21–23 April 1999.

[11] U. Benzler, Enhanced fine granular scalable video coding, in: Proceedings of the Systemics, Cybernetics and Informatics Conference, SCI/ISAS 2001, vol. 12.

[12] M. Narroschke, Improving the coding efficiency of MPEG-4 FGS by using hybrid coding in the enhancement layer, in: Proceedings of Picture Coding Symposium, PCS'03, Saint-Malo, France, April 2003.

[13] ITU-R 601: International Telecommunication Union, Recommendation ITU-R BT.601-5: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios, Geneva, 1995.

[14] J.-R. Ohm, Three-dimensional subband coding with motion compensation, IEEE Trans. Image Process. vol. IP-3 (5) (1994) 559–571.
[15] S.-T. Hsiang, J.W. Woods, Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank, Signal Process.: Image Commun. 16 (2001) 705–724.
[16] S.-J. Choi, J. Woods, Motion compensated 3-D subband coding of video, IEEE Trans. Image Process. 8 (2) (1999) 155–167.
[17] A. Secker, D. Taubman, Motion-compensated highly scalable video compression using an adaptive 3-D wavelet transform based on lifting, in: Proceedings of IEEE International Conference on Image Processing ICIP 2001, Thessaloniki, Greece, 2001, pp. 1029–1032.
[18] H.-J. Stolberg, M. Bereković, P. Pirsch, A platform-independent methodology for performance estimation of multimedia applications, in: Proceedings of International Conference on Multimedia and Expo, vol. 2, August 2002, pp. 105–108.
[19] Y. He, F. Wu, S. Li, Y. Zhong, S. Yang, H.26L-Based Fine Granularity Scalable Video Coding, in: Proceedings of ISCAS 2002, vol 4, Phoenix, USA, May 2002, pp. 819–822.
[20] M. van der Schaar, H. Radha, Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video, in: IEEE Trans. Circ. Syst. Video Technol., 12(6), 2002, pp. 360–371.
[21] K. Hanke, M. Wien, RD Performance of MC-EZBC compared to MPEG-4 AVC JM21, in: Contribution to 61th MPEG Meeting, Klagenfurt, Austria, 2002, m8656.
[22] K. Hanke, RD Performance of Fully Scalable MC-EZBC (Results of EE1), in: Contribution to 62th MPEG Meeting, Shanghai, China, 2002, m9000.
[23] J.-R. Ohm, Complexity and delay analysis of MCTF interframe wavelet structures, in: Contribution to 61th MPEG Meeting, Klagenfurt, Austria, 2002, m8520.
[24] ISO/IEC, Call for proposals on scalable video coding technology, ISO/IEC JTC1/SC29/WG11, N5958, Brisbane, October 2003.
[25] ISO/IEC 14496-10:2003, Coding of Audiovisual Objects—Part 10: Advanced Video Coding, 2003, also ITU-T Recommendation H.264 Advanced video coding for generic audiovisual services.

Matthias Narroschke received his Dipl.-Ing. degree in electrical engineering from the University of Hanover in 2001 (with highest honors). Since then he has been working toward the PhD degree at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung of the University of Hanover. His research interests are image processing, video coding und internet streaming. In 2003, he became a Senior Engineer at this institute. He received the Robert-Bosch-Prize for the best Dipl.-Ing. degree in electrical engineering in 2001. He is an active delegate to the Motion Picture Experts Group (MPEG) of ISO/IEC SC29.