Functionalities and Costs of Scalable Video Coding for Streaming Services

Matthias Narroschke
Institut fuer Theoretische Nachrichtentechnik und Informationsverarbeitung
Universitaet Hannover, Appelstr. 9a, 30167 Hannover, Germany
narrosch@tnt.uni-hannover.de

Abstract

Functionalities and costs of scalable video coding techniques are analyzed for streaming services. Temporal, spatial, amplitude scalability and combinations are considered. Functionalities are reduction of netload for multicast transmission, reduction of the server storage capacity and graceful degradation in case of transmission errors. Costs are an increase of netload for unicast transmission and an increase of computational expense in the decoder. The result is that presently only temporal scalability has acceptable costs. Costs of known spatial scalability techniques are too large to be economically attractive. If costs of amplitude scalability can be reduced by future research the use of this technique combined with temporal scalability will allow improved functionalities.

1 Introduction

As a result of the growth of the multimedia communication on the internet, streaming of audiovisual content over IP-based heterogeneous networks becomes more and more important. In this scenario content is provided by a streaming server and can be streamed by one or more clients. The transmission mode can be either unicast or multicast. Unicast means that the server has a separate point-to-point connection to each client. Multicast means that the server has one point-to-multipoint connection to all participating clients (Figure 1). 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, besides various different available channel capacities, also different transmission error behaviors have to be handled, as well as fast variations of the available channel capacity. Therefore scalable video coding is useful for streaming services because it allows the decoding of only parts of the whole bitstream.

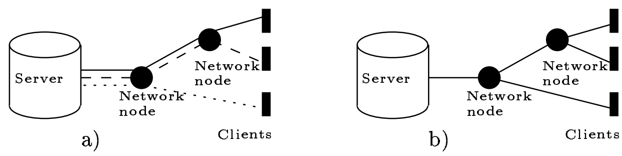

Figure 1: Transmission modes: a) Unicast, b) Multicast.

Scalable encoded data contains one so-called base layer bitstream and one or more so-called enhancement layer bitstreams. By decoding just the base layer bitstream a video of low resolution can be received. The resolution of this video is reduced in amplitude, spatial or temporal direction. The enhancement layer bitstreams contain all additional information that is necessary for decoding higher resolutions. Thus, scalable video coding enables the following functionalities: The server storage capacity and the netload for the transmission mode multicast can be reduced, because bitstreams of lower data rates are implicitly included in bitstreams of higher data rates. Furthermore, in combination with unequal error protection, scalable video coding enables a graceful degradation in case of transmission errors. Without scalable coding, transmission errors can lead to complete image loss.
But scalable video coding causes costs as well. Generally, the netload for a single stream in transmission mode unicast is increased due to the additional data rate that is needed for providing scalability. Also computational complexity is increased in decoders, which process one or more enhancement layers in addition to the base layer for higher resolutions.
Scalable coding algorithms are mainly based on standardized hybrid coding techniques. Amplitude scalability (AS) is standardized in the MPEG-2 SNR scalable profile [1]. This technique uses a two step quan-

tization of the DCT coefficients. In the base layer, coarse quantized DCT coefficients are transmitted. Finer requantized quantization errors of these coefficients are transmitted in one single enhancement layer. The coding efficiency of this technique is low [2]. Improved coding efficiency can be achieved by MPEG FGS (Fine Granularity Scalability) [2], a technique standardized in MPEG-4 [3]. This technique uses embedded quantization [4] in the enhancement layer to encode the quantization error, which allows more scalability steps as well.
In the "spatial scalable profiles" of MPEG-2 [1] and MPEG-4 [5], a technique for spatial scalability (SS) is standardized. Two spatial resolutions are provided using a so-called pyramid coding. The coding efficiency of this technique is low because the number of transmitted DCT coefficients and motion vectors is higher compared to non scalable coding. Spatial scalability can be realized more efficiently using subband coding [6].
Temporal scalability (TS) can be efficiently realized using B-Frames, which are not used as reference pictures for further predictions.
There are also algorithms based on three dimensional wavelet coding [7, 8], which are not considered in this paper.
In this paper, functionalities and costs of scalable video coding are analyzed for different streaming services. Afterwards the functionalities and costs are evaluated for the scalability techniques MPEG FGS [5], spatial scalability using subband coding [6], temporal scalability using B-Frames and all combinations.
The remainder of this paper is organised as follows. In Section 2, functionalities and costs of scalable video coding are analyzed for various streaming services. Section 3 evaluates these funcionalities and costs for different scalability techniques. Section 4 lists the results. The paper closes with the conclusions.

2 Functionalities and costs of scalable video coding for major streaming services

Video streaming services can be clustered into two main groups, demanding either uni- or bidirectional streaming. Services for unidirectional streaming use either unicast or multicast transmission. For bidirectional streaming, unicast transmission is commonly used. In the following subsections, functionalities and costs of scalable videocoding are analyzed for each service.

2.1 Unidirectional streaming

Services with unicast transmission (e.g. Video on Demand) enable every client to request a video individually with an arbitrary starting time. Due to the variety of different end devices and networks, it is necessary to provide a multitude of bitstreams with different data rates on the server. Thus, a functionality of scalable video coding concerning these services is to reduce the server storage capacity, if the sum of all scalable encoded bitstreams requires less data than the sum of all non scalable encoded bitstreams.
Variations of available channel capacity in IP networks due to packet delay and packet loss can be faster than a server's reaction. Another functionality of scalable video coding is to balance these variations using a graceful degradation.
But scalable video coding also causes an additional data rate needed to provide scalability. This results in an increased netload for individual bitstreams, which include one or more enhancement layers. Also, additional computational expense is needed in all decoders, which process one or more enhancement layers beside the base layer. Especially in mobile devices, low computational expense is important because of the limited battery capacity and limited computational power.
Additional computational expense in the decoder is mirrored in all of the following analyzed services.
For services with multicast transmission (e.g. Broadcast) the starting time of the streaming for all bitstreams of all data rates is fixed. The client is told when to receive the content. Due to the variety of different end devices and networks, it is necessary to provide a multitude of bitstreams with different data rates on the server. All bitstreams must be sent only once and at the same time. Therefore scalable video coding offers two functionalities: Reduction of the server storage capacity and reduction of the netload, if the sum of all scalable encoded bitstreams requires less data rate than the sum of all non scalable encoded bitstreams.
Also, fast variations of available channel capacity can be balanced by scalable video coding using a graceful degradation.

2.2 Bidirectional streaming

Services demanding bidirectional streaming are, for instance, Video Communication and Video Conferencing. A fundamental requirement for bidirectional communication is a low end to end delay.
The server encodes the content in real time and sends one bitstream of a specific data rate in transmission mode unicast. This data rate is continuously adapted

in consultation with the client. This adaptation is possible due to an existing feedback channel. Because of the required low end to end delay, variations of available channel capacity in IP networks can be balanced by permanent data rate adaptation by the encoder. A functionality of scalable video coding is to enable a graceful degradation if very fast variations of available channel capacity occur. However, scalable video coding is associated with an increase of the data rate.

2.3 Lineup of functionalities and costs

A lineup of functionalities and costs is shown in Tables 1 and 2. Scalable video coding is especially useful for services demanding unidirectional streaming. The use of scalable video coding can reduce the server storage capacity. For transmission mode multicast, netload can be reduced as well. For all investigated services, scalable video coding enables a graceful degradation if very fast variations of available channel capacity occur, which can not be handled fast enough by arranging new parameters between the server and the client. Scalable video coding is not implicitly necessary for bidirectional services except for a graceful degradation. Concerning costs, computational expense in the decoder is increased for all services by using scalable video coding. For investigated services associated with unicast transmission, scalable video coding causes an increase of netload.

Streaming direction	Transmission mode	Functionalities		
		Reduction of server storage capacity	Reduction of netload	Graceful degradation
Uni-directional	Unicast	yes	no	yes
	Multicast	yes	yes	yes
Bi-directional	Unicast	no	no	yes

Table 1: Functionalities of scalable video coding for different services.

Streaming direction	Transmission mode	Costs	
		Increase of computational expense in the decoder	Increase of netload
Uni-directional	Unicast	yes	yes
	Multicast	yes	no
Bi-directional	Unicast	yes	yes

Table 2: Costs of scalable video coding for different services.

3 Evaluation of functionalities and costs

In the previous section it was pointed out that scalable video coding is not implicitly necessary for services demanding bidirectional streaming. Therefore this evaluation concentrates on services demanding unidirectional streaming.
The intention is to provide encoded audiovisual content for a multitude of different data rates simultaneously. For access to this audiovisual content 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 as listed in Table 3 should be supported. Low data rates are related to multiples of 8 kbits/s for UMTS and to multiples of 9.6 / 14.4 kbit/s for GSM with GPRS/ HSCSD. For each target data rate, the video has to be encoded in a specific temporal, spatial and amplitude resolution. The resolution assignment, which this evaluation is based on, is illustrated in Table 3. For example, it is assumed that a video of 32 kbit/s is encoded in QCIF, 12.5 Hz and medium amplitude resolution. This assignment is liable to subjective picture quality and might be slightly different in special applications. In the following subsections, a detailed evaluation is presented that concerns the functionalities and costs of each scalability technique according to Table 1 and Table 2 if encoded bitstreams for all data rates listed in Table 3 are provided. The reference is non scalable video coding.

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	TV	50	low
2048	TV	50	high

Table 3: Assignment of resolution levels to target data rates.

3.1 Server storage capacity

Table 4 shows the calculation of the server storage capacity in kbit per second of encoded video con-

tent that is needed to provide an encoded bitstream for all data rates of Table 3 using different scalability techniques. Differing from Table 3, it is assumed for this calculation that if TS is involved, a temporal resolution of 25 Hz and a high amplitude resolution is assigned to the data rate of 1536 kbit/s. Table 4 includes all data rate increases caused by scalable coding. AS requires an overhead of approximately $O_A \approx 80\%$ [2], SS supplying two resolutions requires about $O_{S,2} \approx 15\%$ [6], supplying three resolutions about $O_{S,3} \approx 25\%$ in form of additional data rate. With respect to the number of B-Frames, the lowest data rate for non scalable coding can be achieved by using two B-Frames. For TS providing a full and a half frame rate, an odd number of B-Frames is necessary. Thus, if only one B-Frame is used an additional data rate of approximately $O_{T,2} \approx (346/333)-1 \approx 4\%$ is needed, as can be seen in Figure 2. For providing a full, a half and one quarter frame rate, a number of $3 + n \cdot 4$ (with $n \geq 0$) B-Frames between two consecutive I- or P-Frames is required. The use of three B-Frames ($n = 0$) requires an additional data rate of approximately $O_{T,3} \approx (366/333)-1 \approx 10\%$, as can be seen in Figure 2 as well.
Using AS, video has to be encoded for target bit rates of 2048, 1024, 384, 128, 43.2 and 16 kbit/s. Bitstreams of all other data rates are implicitly available due to scalable coding. But an additional data rate of O_A is needed for AS.

The server storage capacity and netload for multicast transmission in kbit per second		
no scalab.	$9.6 + 14.4 + 16 + \ldots + 2048 = 6371$	$= S$
AS	$(2048 + 1024 + 384 + 128 + 43.2 + 16)(1 + O_A) = 6558$	$= 1.03S$
SS	$2048 + 1536 + (1024 + 768 + 384 + 256)(1 + O_{S,2}) + 32 + 19.2 + \ldots + 9.6 = 6472$	$= 1.02S$
TS	$(2048 + 1024 + 768 + 128 + 64)(1 + O_{T,2}) + 32 + 19.2 + \ldots + 9.6 = 4284$	$= 0.67S$
TS+ AS	$2048(1 + O_{T,2}) + (1024 + 128)(1 + O_{T,2})(1 + O_A) + 16 + 14.4 + 9.6 = 4326$	$= 0.68S$
SS+ AS	$2048(1 + O_A) + (1024 + 384)(1 + O_{S,2}) \cdot (1 + O_A) + 16 + \ldots + 9.6 = 6641$	$= 1.04S$
SS+ TS	$(2048 + 768)(1 + O_{S,2})(1 + O_{T,2}) + 384(1 + O_{S,2}) + 128 + 32 + 19.2 + \ldots + 9.6 = 4029$	$= 0.63S$
SS+TS+ AS	$2048(1 + O_{S,3})(1 + O_A)(1 + O_{T,3}) + 16 + 14.4 + 9.6 = 5109$	$= 0.80S$

Table 4: The server storage capacity and netload for multicast transmission for various scalability techniques.

3.2 Netload for multicast transmission

The netload for multicast transmission is proportional to the server storage capacity, because bitstreams of all data rates must be transmitted simultaneously.

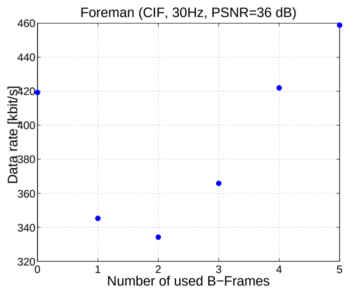

Figure 2: Number of used B-Frames versus data rate for fixed PSNR.

3.3 Graceful degradation

Each scalability technique behaves differently in terms of graceful degradation (GD). Table 5 shows the number of steps of a graceful degradation provided by each of the investigated scalability techniques. It is assumed that AS allows at least four steps. For example, if SS and AS are used, at least eight steps of a graceful degradation are provided. More than four steps of AS in each of the two spatial resolutions.

3.4 Computational expense in the decoder

Table 5 shows the results of the evaluation for the worst case. Each shown computational expense is normalized to the one of non scalable coding. Worst case means that the decoder must perform all subtasks and process all enhancement layers. This evaluation is based on static performance estimation results concerning the main subtasks of a decoder, which are inverse cosine transform (IDCT), motion compensation and reconstruction of the decoded image, as worked out in [9]. In addition to the mentioned subtasks, subband filtering is required for spatial scalable coding. It is assumed that this operation requires a computational expense, which is comparable to 8-tap filtering [9]. The computational expense is calculated by the cycle rate required to process the decoding. A non scalable decoder requires 10 MHz for the IDCT, 5.3 MHz for the reconstruction and 43,6 MHz for the motion compensation [9], a total of 58.9 MHz. A decoder for two spatial resolutions additionally requires a quarter of 43.6 MHz for the motion compensation in the low resolution and 2 times 25.27 MHz for subbandfiltering. This is a total of 120.34 MHz, which is 2.04 times higher compared to the computational expense of a non scalable decoder.

	Number of steps of a graceful degradation		Normalized computational expense in the decoder (worst case)	
no scalability		1		1
AS	$\geq$	4		1.18
SS		2		2.04
TS		2		1
TS + AS	$\geq 2 \cdot 4 =$	8	$1 \cdot 1.18 =$	1.18
SS + AS	$\geq 2 \cdot 4 =$	8	$2.04 \cdot 1.18 =$	2.41
SS + TS	$2 \cdot 2 =$	4	$2.04 \cdot 1 =$	2.04
TS +SS + AS	$\geq 3 \cdot 3 \cdot 4 =$	36	$2.35 \cdot 1.18 \cdot 1 =$	2.54

Table 5: Number of steps of a graceful degradation and normalized computational expense in the decoder for the worst case.

3.5 Netload for unicast transmission

According to the additional data rate that is needed to provide scalability, the data rate of each individual stream in unicast transmission mode can be increased. Thus, the whole netload for unicast streaming is increased. Table 6 shows the netload for unicast transmission for the worst case. The worst case concerns bitstreams that have a maximum additional data rate due to scalable encoding compared to a bitstream of non scalable coding. R_U is the netload that is needed if no scalability is used. For example if SS and AS are used, additional data rates of $O_{S,2}$ and O_A are needed. Netload for individual unicast streams is not increased if the client requests just the base layer.

Netload for transmission mode unicast (worst case)		
no scalability	R_U	$= R_U$
AS	$R_U(1 + O_A)$	$= 1.8R_U$
SS	$R_U(1 + O_{S,2})$	$= 1.15R_U$
TS	$R_U(1 + O_{T,2})$	$= 1.04R_U$
TS + AS	$R_U(1 + O_{T,2})(1 + O_A)$	$= 1.87R_U$
SS + AS	$R_U(1 + O_{S,2})(1 + O_A)$	$= 2.07R_U$
SS + TS	$R_U(1 + O_{S,2})(1 + O_{T,2})$	$= 1.20R_U$
TS +SS +AS	$R_U(1 + O_{S,3})(1 + O_A)(1 + O_{T,3})$	$= 2.48R_U$

Table 6: Netload for transmission mode unicast for the worst case.

4 Results

The evaluated functionalities and costs are shown in Table 7. Bold numbers mark costs that are considered not acceptable for the use in services. Compared to non scalable coding, AS has low costs concerning the increase of computational expense in the decoder (18%) but large costs concerning the increase of the netload for unicast transmission (80%). SS is associated with large costs regarding the increase of computational expense in the decoder (104%) and with low costs concerning the increase of netload for unicast transmission (15%). These large costs are mirrored in all combinations in which AS and SS are involved. Presently, only TS has acceptable costs for services (gray shaded in the Table). This scalability technique leads to 33% reduction of the server storage capacity and 33% reduction of the netload for multicast transmission. But with respect to graceful degradation, TS only provides two steps.

	Functionalities			Costs	
	Reduction of the server storage requirement	Reduction of netload for multicast	Steps of GD	Increase of comput. expense decoder	Increase of netload for unicast
no scalability	0%	0%	1	0%	0%
AS	-3%	-3%	$\geq$4	18%	**80%**
SS	-2%	-2%	2	**104%**	15%
TS	33%	33%	2	0%	4%
TS +AS	32%	32%	$\geq$8	18%	**87%**
SS +AS	-4%	-4%	$\geq$8	**141%**	**107%**
SS +TS	37%	37%	4	**104%**	20%
TS +SS +AS	20%	20%	$\geq$36	**154%**	**148%**

Table 7: Functionalities and costs of scalable video coding.

There is a chance to reduce the netload for unicast transmission by future research. Then AS combined with TS would become interesting for various services. Table 8 shows functionalities and costs for the presumption, that only 30% instead of 80% additional data rate is needed for AS. It can be seen that using

	Functionalities			Costs	
	Reduction of the server storage requirement	Reduction of netload for multicast	Steps of GD	Increase of comput. expense decoder	Increase of netload for unicast
no scalability	0%	0%	1	0%	0%
AS	26%	26%	$\geq$4	>18%	30%
SS	-2%	-2%	2	**104%**	15%
TS	33%	33%	2	0%	4%
TS +AS	42%	42%	$\geq$8	>18%	35%
SS +AS	25%	25%	$\geq$8	**>141%**	**50%**
SS +TS	37%	37%	4	**104%**	20%
TS +SS +AS	42%	42%	$\geq$36	**>154%**	**79%**

Table 8: Functionalities and costs of scalable video coding with the assumption that the increase of netload for unicast transmission of AS is 30% instead of 80%.

improved AS combined with TS, the server storage capacity and the netload for multicast transmission can be reduced by an additional 9% compared to single TS and by 42% compared to non scalable coding. This technique offers improved graceful degradation

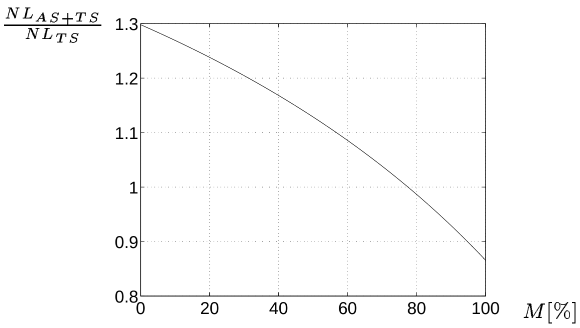

Figure 3: Normalized total netload versus multicast traffic M for the worst case.

due to an increased number of steps as well.
Since the combination of improved AS and TS will still result in an increase of 35% netload for unicast transmission in the worst case, the relation between the traffic amount of multicast M and unicast transmission has significant influence on the choice of the scalability technique. The total netload for combined AS and TS or just single TS is:

$$\begin{aligned} NL_{AS+TS} &= NL_{ns} \cdot (M \cdot 58\% + (1-M) \cdot 135\%) \\ NL_{TS} &= NL_{ns} \cdot (M \cdot 67\% + (1-M) \cdot 104\%). \end{aligned} \tag{1}$$

NL_{ns} is the netload if no scalable coding is used. Figure 3 shows that only if the multicast traffic exceeds 77.5% the use of combined AS and TS will be efficient, because the total netload can be reduced compared to the total netload using TS.

5 Conclusions

Functionalities and costs of scalable video coding are analyzed for streaming services. Functionalities are reduction of the server storage capacity, reduction of the netload for multicast transmission and graceful degradation. Costs are an increasing computational expense in the decoder and an increasing netload for unicast transmission. These functionalities and costs are evaluated for different scalability techniques.
The result is that presently only the costs of TS are acceptable. The server storage capacity and the netload for multicast can be reduced by 33%. But only two steps of graceful degradation are enabled by this technique. SS and AS are associated with costs too large to be economically attractive.
If the increase of the unicast netload of AS can be reduced in the future from currently 80% to 30% without extensive additional computational expense in the decoder, a combination of AS and TS would become economically interesting for services using multicast transmission because of improved functionalities compared to single TS. But since this combination would still result in an increase of 35% netload for unicast in the worst case, the use of this combination will reduce the total netload only if at least 77.5% of all traffic is multicast.

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] 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

[4] W. Li, "Bit-Plane Coding of DCT Coefficients for fine granularity scalablility", in *Contribution to 45th MPEG Meeting*, Atlantic City, NJ, October 1998, m3989

[5] 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

[6] U. Benzler, "Spatial Scalable Video Coding using a combined Subband-DCT Approach", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 7, pp. 1080–1087, October 2000.

[7] J.-R. Ohm, "Three-Dimensional Subband Coding with Motion Compensation," IEEE Transactions on Image Processing, Vol. IP-3, No. 5, pp. 559-571, September 1994

[8] S.-J. Choi, J. W. Woods, "Motion compensated 3-D subband coding of video," in *IEEE Transactions on Image Processing*, vol. 8, no. 2, February 1999

[9] H.-J. Stolberg, M. Bereković, P. Pirsch, "A platform-Independent Methodology for Performance Estimation of Streaming Media Applications", in *Proc. of IEEE International Conference on Multimedia and EXPO*, 2002