

Fine-Grained Scalable Streaming from Coarse-Grained Videos

Pengpeng Ni^{1,2}, Alexander Eichhorn¹, Carsten Griwodz^{1,2}, Pål Halvorsen^{1,2}

¹Simula Research Laboratory, Norway

²Department of Informatics, University of Oslo, Norway
{pengpeng, echa, griff, paalh}@simula.no

ABSTRACT

Scalable video is an attractive option for adapting the bandwidth consumption of streaming video to the available bandwidth. Fine-grained scalability can adapt most closely to the available bandwidth, but this comes at the cost of a high compression penalty. In the context of VoD streaming to mobile end systems, we have therefore explored whether a similar adaptation to the available bandwidth can be achieved by performing layer switching in coarse-grained scalable videos. In this approach, enhancement layers of a video stream are switched on and off to achieve any desired longer-term bandwidth. We performed user studies to evaluate the idea, and came to the far-from-obvious conclusion that layer switching is viable way for bit-rate savings and fine-grained bit-rate adaptation even for rather short times between layer switches.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Video

General Terms

Human Factors, Measurement, Experimentation

Keywords

Quality of Experience, Scalable Video, Layer Switching

1. INTRODUCTION

Streaming stored video to a large number of heterogeneous receivers over various networks introduces several challenges with respect to delivered rate and quality. Various layered video approaches that address this exist, including coarse-grained and fine-grained scalable video and multiple description coding. They can be used to choose a quality level whose bandwidth can be delivered to and consumed by a receiver with a limited amount of prefetching and buffering. They can also be used to adapt over time the amount

of bandwidth that is delivered to a single receiver. Fine-grained scalable video is apparently meant for the latter approach in particular. However, since both fine-grained scalable video and multiple description coding suffer from a very high overhead, the question arises whether switching between the number of layers of a coarse-grained video that are delivered to a receiver could also be used to exploit an arbitrary available bandwidth for quality improvement.

In this context, we look at the H.264 Scalable Video Coding (SVC), an international video coding standard that defines multi-dimensional scalability [8]. It supports different temporal resolutions, spatial resolutions and qualities of a video sequence using enhancement layers. SVC is designed for efficient and network-friendly operation [9]; heterogeneity and bandwidth variations can be supported by tuning the bit rate and using adaptation of the compressed bit stream to individual device capabilities.

The granularity of different scaling options is limited by the bit rates of contained operation points, i.e., between the different encoded quality layers. Scaling options are predetermined at encoding time and their number is limited by the SVC standard which supports 8 enhancement layers in total [8]. SVC's mid-grain scalability feature (MGS) is supposed to introduce higher adaptation granularity, but this comes at the cost of increased signaling overhead. For better bit-rate efficiency, it is thus desirable to limit the number layers and also the number of MGS partitions.

In this paper, we propose and investigate the usefulness of frequent layer switching techniques as a method for fine-grained bit-rate adaptation of SVC streams with few operation points. In contrast to static downscaling approaches, layer switching alternates between two or multiple operation points in order to meet a given bit-rate constraint over a short time-window.

Layer switching can achieve amounts of bandwidth consumption different from the long-term average of any layer of the coarse-grained scalable video. This ability can be interpreted and used in two ways:

- It can be interpreted as a means of exploiting a long-term average bandwidth that is unlike the average bandwidth of every operation point of the coarse-grained scalable video. Every average target bandwidth above the base layer's bandwidth demand can be achieved by switching enhancement layers on and off repeatedly. If the streaming itself is congestion-controlled, reaching the long-term average in this way would require prefetching.
- It can be seen as an alternative means to exploit the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NOSSDAV'09, June 3–5, 2009, Williamsburg, Virginia, USA.
Copyright 2009 ACM 978-1-60558-433-1/09/06 ...\$5.00.

Genre	Content	Detail	Motion	Audio	CGS Bitrate	MGS Bitrate	Overhead
Animation	BigBuckBunny	3.65	1.83	sound	419.8	520.1	23.9%
Cartoon	South Park	2.75	0.90	speech	433.1	539.4	24.5%
Docu	Earth 2007	3.64	1.61	sound	739.8	789.0	6.6%
Movie	Dunkler See	1.85	0.58	sound	192.1	263.1	37.0%
News	BBC News	2.92	0.69	speech	213.9	300.4	40.4%
Sports	Free Ride	3.32	1.90	music	472.9	484.8	2.5%

Table 1: Sequences used in the experiments. Detail is the average of MPEG-7 edge histogram values over all frames [7] and Motion is the MPEG-7 Motion Activity [4], i.e., the standard deviation of all motion vector magnitudes. Bit-rates are given in kbit for the total (global) SVC bitstream at the highest enhancement layer operation point. The overhead shows the increase in bit-rate when using MGS encoding instead of CGS with an otherwise similar encoder configuration.

temporary availability of bandwidth that exceeds the demands of the base layer but does not suffice the bandwidth demands of an enhancement layer. Through variations of the retrieval speed (implicitly in pull mode, explicitly in push mode), receivers can use the excess bandwidth during a period of base-layer playout to prefetch data for a period of enhanced-quality playout. The duration of these periods should be short to keep the prefetching buffer small, but perceived quality forbids an arbitrary choice.

However, layer switching can also influence the perceived quality of video content. To identify the feasibility of switching techniques and advice design constraints, we conduct a subjective quality assessment study. We are interested in the following research questions:

- Do people perceive a difference in quality between scaling and switching techniques?
- Is there a general preference of one technique over the other?
- Does a preference depend on genre, period length of switching or the scaling/switching dimension?
- Which periods and dimensions are perceived as less disturbing?

Furthermore, in our previous work [2], we showed that with a fixed set of operation points (6), low rate (<200 Kbps) streams have sufficient granularity using our tested scenario. However, increasing the bit rate, and thus supporting a wider range of scenarios, the granularity becomes coarse and the diversity is reduced. This may result in a lack of alternative scaling options, severely hurting the quality of experience (QoE) and wasting resources.

Therefore, we look at possibilities to make fine-grained scalable streaming from coarse-grained videos. In particular, we have investigated how the rate can be changed in a more fine-granular manner by frequently switching between predetermined layers - hopefully without affecting the QoE. Thus, the coarse-grained adaptation possibilities are enhanced by introducing intermediate scaling options by layer switching.

We restrict ourselves to on-demand and broadcast delivery of pre-encoded content at bit rates offered by existing wireless networks. Because we are interested in QoE perception on real mobile devices in natural environments, we conduct a field study rather than a synthetic laboratory experiment. We have tested several layer switching patterns giving rates in-between the precoded SVC layers.

Our subjective tests indicate that switching patterns with sufficient perceptual quality exist, i.e., if the switching period is long enough to avoid flickering, the layer switching is preferred compared to downscaling to the lower encoded SVC layer. Thus, achieving fine-grained scalable streaming from coarse-grained videos is possible.

The paper is organized as follows: Section 2 lists some related work. The experiments are described in section 3, and the results are given in section 4. We discuss our findings in section 5, and in section 6, we summarize the paper.

2. RELATED WORK

Scalable video coding increases perceptual uncertainty dramatically because of its multi-dimensional scaling possibility. There are a few published studies investigating the quality influence of different scaling options. In [1], a set of experiments was carried out to discover the Optimal Adaptation Trajectory that maximizes the user perceived quality in the adaptation space defined by frame rate and spatial resolution. It was shown that two-dimensional adaptation strategy outperformed one-dimensional adaptation. Meanwhile, according to an objective video quality model [10] that multiplicatively combines the quantization distortion and frame loss, it was claimed that quality scaling worked better than temporal scaling under most circumstances. Additionally, subjective tests introduced in [6] showed that high frame rate is not always more preferable than high fidelity quality for high motion video. Probably closest to our work, Zink et al.'s evaluation has been performed to investigate quality degradation caused by variations in the amount of transmitted layers during streaming sessions [11]. The authors' results showed that the perceived quality of video is influenced by the amplitude and the frequency of layer switchings. In contrast to our work, the relationship between quality impairment and layer switching at temporal and quality dimensions are not treated separately.

The subjective quality assessments by Cranley et al. [1] and Zink et al. [11] were conducted with regular monitors under lab conditions, which is different from our testing scenario defined for mobile video applications using iPods. Further, very few of previous studies performed subjective evaluation of the H.264 scalable extension. To the best of our knowledge, only in [2], a subjective field study about the H.264/SVC is introduced which also grounded our investigation presented in this paper.

Name	FPS	Layer	Bitrate	PSNR
G	25.0	EL+BL	100%	37.85
QD	25.0	BL	28%	33.18
TD	12.5	EL+BL	85%	33.14

Table 2: Operation points used in the experiments. All resolutions are QVGA. The PSNR is an average of all genres.

Name	Layer	Pattern	Period	Bitrate	PSNR
QS ₁	quality	G & QD	80ms	89%	35.57
QS ₂	quality	G & QD	1s	89%	35.61
QS ₃	quality	G & QD	2s	89%	35.62
MS	multi	G & QD	320ms	73%	33.07
TS ₁	temporal	G & TD	1s	92%	35.60
TS ₂	temporal	G & TD	2s	92%	35.49

Table 3: Switching patterns used in the experiments. All patterns are at QVGA resolution and were upscaled to 25fps if necessary. PSNR is an average of all genres.

3. EXPERIMENT DESIGN

3.1 Material

To test different kinds of content with varying detail and motion, we selected six sequences from different genres (see table 1). We obtained the content from a previous study on scalable coding [2] which allowed for a comparison with earlier results. All sequences were downsampled and eventually cropped from their original resolution to QVGA (320x240) in order to fit the screen size of our display devices. From each sequence, we extracted an 8 second clip (200 frames) without scene cuts. After extraction, the texture complexity and motion activity are measured according to MPEG-7 specification. We encoded the SVC bitstreams with version 9.16 of the JSVM reference software¹. The encoder was configured to generate streams in the scalable high profile with one base layer and one coarse-grained scalable (CGS) or MGS enhancement layer, a GOP-size of 4 frames with hierarchical B-frames, an intra period of 12 frames, inter-layer prediction and CABAC encoding. Note, SVC defines the set of pictures anchored by two successive key pictures together with the first key picture as a group of picture, where key pictures are usually encoded as P-frames within an intra period, see [8]. Due to the lack of rate-control for quality enhancement layers in JSVM, we used fixed quantization parameters (QP). Based on our previous experience and in order to obtain a perceivable quality difference, we selected QP 36 for the base-layer and QP 28 for the enhancement layer.

Table 1 further displays how the bit-rate of a CGS encoding compares to a MGS version of the same stream, generated with otherwise similar encoding parameters. As we can see, the higher flexibility of MGS for bit-rate adaptation comes at a considerable cost.

¹Available at http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm.

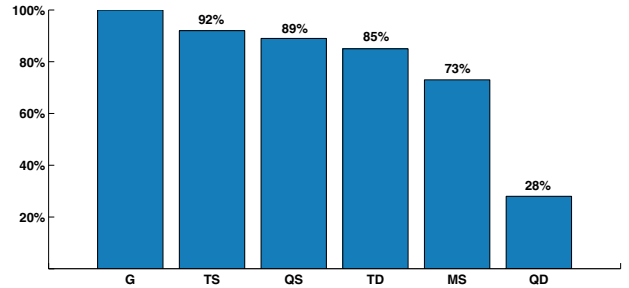


Figure 1: Bit-rate of different operation points and switching patterns.

3.2 Scalable Operation Points and Switching Patterns

In this study, we focused on quality scalability and temporal scalability only. We did not consider spatial scalability because it is undesirable for frequent layer switching due to the large decrease in perceived quality as shown in previous subjective studies [2].

From the encoded SVC bitstreams we extracted three scalable operation points (OP) with a high variability in bit-rate (see table 2 and figure 2a–2c). OP 'G' contains the full bitstream including the base-layer (Q₀) and the quality enhancement layer (Q₁) at the original frame-rate, while the two other OPs are each downsampled in a single dimension to the low-quality base-layer at full temporal resolution (QD) or a lower temporal resolution T₁ (12fps), but with quality enhancement (TD). These operation points were then used to generate streams with different switching patterns and to compare the switched streams' quality. In figure 2, we illustrate the two dimensional layer switching and downscaling. Figure 2b shows temporally downsampled bitstream in which frames belonged to higher temporal layer such as T₂ are skipped. Respectively, in figure 2c, the downscaling is only made in quality enhancement layer but higher temporal resolution is maintained.

We generated three different switching patterns (figure 2d–2f), two that perform switching in a single dimension (temporal or quality) to observe the effects of switching dimensions isolated and one pattern that performs multi-dimensional switching to observe the combined effects. The patterns alternate between exactly two operation points (three in the multi-dimensional case) at different periods. A period means a round turn of layer switching, for instance, for a 8 seconds video sequence with period of 2s, there will be at least 8 times change of scaling direction (up and down). In our experiment, periods have been selected based on anchor frame distance (4 frames) and common Intra-refresh intervals (1s and 2s). We expect that periods will influence the visual quality in different ways. Similar switching patterns that just differ in period length consume almost the same long-term average bit-rate.

The average bit-rate consumption ratios for our operation points and switching patterns are illustrated in figure 1. The bit-rates are averaged for the six types of content. Figure 1 gives the percentage of bits consumed by the respective sub-stream compared to the global bitstream. Higher scaling granularity is preferable for video streaming [2]. As shown by figure 1, our switching patterns increase the granular-

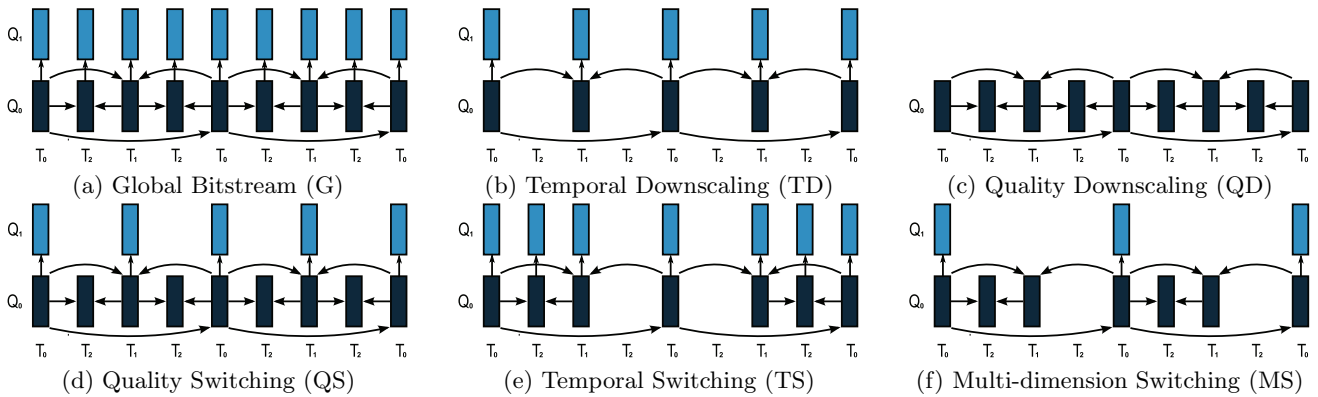


Figure 2: Bitstream layout for downscaling and layer switching options used in the experiments. Q and T denote the quality and temporal dimension respectively.

ity beyond the scaling opportunities given by the operation points without introducing extra overhead.

3.3 Procedure

We conducted a field study at different locations (office, library, cafeteria) because this research method allows us to observe natural user experience under familiar viewing conditions rather than quality perception in a single synthetic laboratory environment. As mobile display devices, we used the iPod classic and the iPod touch from 2008. The two iPod models contain respectively a 2.5-inch and 3.5-inch display and have pixel resolutions of 320x240 and 480x320 at 163 pixel per inch. The selected display size is sufficient for depicting content at QVGA resolution according to [5]. All videos had an undistorted audio track to decrease the exhaustion of test participants.

We were interested if and under which conditions users would prefer frequent layer switching over layer scaling. Furthermore, we required a simple, robust and reliable test that allowed us to conduct a large number of comparisons. Hence, we based our test design on the Pair Comparison method as defined by the ITU-T recommendations for subjective video quality assessment methods for multimedia applications P.910 [3]. The Pair Comparison method is a preferential test, where pairs of clips are presented in rapid succession. We always compared one layer switching pattern against one static operation point. Each pair of patterns was presented twice during a test sequence, once in each possible order to assess the reliability of votes from each participant and detect inconsistent ratings. The order of all the pairs of a test sequence was a random permutation. Between subsequent pairs, there was a 6-second break, displaying a mid-grey image with black text that called for voting and announced the following clip. The participants were asked to judge whether they preferred the first or the second clip in the pair or whether they did not perceive a difference.

3.4 Participants

Twenty-eight paid assessors (25% female) at mean age of 28 participated in the test. Among the assessors, 90% are at an age between 18-34 while 10% are at an age between 35-39. All of the assessors are college students with different education but no one has majored in multimedia technologies. All of the assessors are familiar with concepts like digital TV and Internet video streaming while 75% of

them claimed that media consumption is part of their daily life. Although quite a few of assessors have previous experience in watching video on handheld devices like iPod, a brief introduction about how to operate the iPods during the experiments was given to the assessors prior to a test session. A whole test session lasted for about one hour, including two short breaks. Each participant watched in total 144 clip pairs. During the session the assessors were free to choose a comfortable watching position and to adjust the watching distance. For example, they could choose to sit on sofas or in front of a desk. They were also free to decide when they wanted to continue the test after a break. We obtained a total of 2016 preference ratings of which 44% indicated a clear preference (consistent ratings on both clip orders), 31% a tendency (one undecided rating) and 10% no difference (two undecided ratings). We observed 15% conflicting ratings, where participants gave opposite answers to a test pattern and its hidden check pattern. Participants with more than 1.5 times the inter-quartile range of conflicting ratings above the average were regarded as outliers. In total, we removed two outliers from our data set. Regardless of remaining conflicts we found statistically significant results.

4. RESULTS

For each clip pair, we obtained a single measure about which clip a participant preferred to watch. If undecided, participants could also select that they had no preference. This resembles a repeated measurement design with three rating categories. We used all ratings from both clip-pair orders (AB, BA) in our analysis. We also included conflicting ratings, because they would just decrease significance, but not invalidate our results. For statistical analysis, we ran Binomial tests to see if a significant majority of ratings for one of the preference categories existed.

Results are reported as preference for layer switching or layer scaling with 0.01 confidence intervals. Rate-savings are reported as percentage of the original bit-rate. If a preference was found as not significant we still give a weak tendency. Table 4 displays preferences between switching and scaling across genres, and table 5 shows results for different period lengths. The 'all' line in table 5 contains general results for all periods and all genres.

Switching	TS	TS	QS	QS	MS	MS
Downscaling	TD	QD	TD	QD	TD	QD
Animation	○	+	-	(+)	-	(-)
Cartoon	(○)	+	-	(+)	-	(○)
Documentary	(+)	+	-	(-)	-	-
Short Movie	○	+	-	(-)	-	-
News	○	+	-	(+)	-	-
Sports	(○)	+	-	(-)	-	(○)

Table 4: Quality preference per genre for layer switching vs. downscaling (+ switching preferred, - downscaling preferred, ○ no preference, (*) not significant).

Switching	TS	TS	QS	QS	MS	MS
Downscaling	TD	QD	TD	QD	TD	QD
80ms			-	-		
320ms					-	(-)
1s	(○)	+	-	(+)		
2s	(○)	+	-	+		
All	(○)	+	-	(+)	-	(-)

Table 5: Quality preference over different switching periods for layer switching vs. downscaling (+ switching preferred, - downscaling preferred, ○ no preference, (*) not significant). Empty cells are not covered by this study.

4.1 Temporal Layer Switching

Participant ratings indicate no clear preference when temporal switching (TS) is compared to temporal downscaling (TD). This is significant for all low-motion sequences where temporal resolution is less important to convey information, but also not significant for other genres. Besides a weak tendency towards an undecided rating, a general conclusion is not possible.

One possible reason for this observation is that temporal resolution changes between 25 fps and 12 fps have a minor impact on quality perception. This confirms results of previous studies as reported in [6] and [2]. Using more bandwidth for a temporally switched stream (92%) compared to a temporal downscaled stream (85%) is thus not justified by a significant increase in quality perception. We are currently investigating whether this observation also applies to switching to lower temporal resolutions (below 10 fps).

When layer switching in the temporal (TS) or quality dimension (QS) is compared to downscaling in the other dimension (QD and TD, respectively), the results indicate a clear preference towards decreasing the temporal resolution rather than the quality of a video. With high significance our results are consistent across all genres and independent of the switching period. The result again confirms previous findings reported in [6]. People seem to be more sensitive to reductions in picture quality than to changes in frame rates when watching video on mobile devices. This clearly indicates that switching is a viable option for frequent temporal resolution changes. Although temporal base-layers consume the main bit-rate and potential savings are small, switching can still yield fine-grained adaptation in the upper bit-rate range of a stream.

For a fair comparison, it is noteworthy that the temporal switching pattern TS (92%) had a considerably higher bit-

rate than the low quality operation point QD (28%). However, the quality of switching pattern QS (89%) compared to the lower temporal resolution TD (85%) shows, that a lower bit-rate stream can yield a higher subjective quality regardless of the content.

4.2 Quality Layer Switching

When quality switching (QS) is compared to downscaling in the same dimension (QD), the combined results over all period sizes are not significant. There is also no general tendency towards a single adaptation technique that can be attributed to content characteristics alone. However, we observed a significant preference for quality-layer switching at long periods while for shorter periods a preference for quality scaling exists.

We attribute this observation to a flickering effect that was perceived as disturbing by almost all participants. Flickering is caused by fast switching between high- and low-quality encodings which leads to rapid iteration of high and low frequency textures. At longer switching periods, this effect becomes less annoying and disappears eventually. We call the limit at which flickering disappears the *flickering threshold*. Interestingly, long switching periods above the flickering threshold are also preferred to a constant low quality.

We just conducted tests with equally long intervals of high and low quality. Hence, the bit-rate demand of a quality switching scheme is still much higher than that of the low-quality operation point (89% vs. 28%). Asymmetric patterns with longer low-quality intervals will have a much lower bit-rate consumption and offer a wider range of bit-rate adaptivity. We will investigate whether such patterns can also yield a better visual quality. We assume, however, that the flickering threshold plays an important role for asymmetric patterns as well.

4.3 Multi-dimensional Layer Switching

Multi-dimensional layer switching (MS) is supposed to combine fine-granular bit-rate savings of temporal and quality switching while yielding increased perceptual quality over pure downscaling. Due to time constraints, we restricted our assessment to a single period length, which turned out to be shorter than the flickering threshold.

In direct comparison to temporal scaling, multi-dimensional switching performs significantly worse. The same tendency exists when compared to quality downscaling, although only a few results are statistically significant here. We account this effect to the short switching period and expect the situation to become similar to our observations about quality layer switching when the period lies above the flickering threshold. At the current period of 320ms (8 frames), the multi-dimensional switching pattern (73%) was perceived worse than a quality-downscaled bitstream (28%), which confirms the harmfulness of short switching periods.

5. DISCUSSION

Our experiments indicate that frequent layer switching is a valid technique for dynamic bitrate adaptation. It can complement static downscaling to achieve any desired long-term target bandwidth when streaming coarse-grained scalable videos. Both techniques are equally valid options to deal with bandwidth fluctuations, but certain design restrictions for layer switching apply.

5.1 Preferences for Switching and Scaling

Although all participants in our test perceived a visual difference between switching and downscaling in almost all cases, there is no general tendency to prefer one technique over the other. One exception is that temporal switching and also temporal downscaling down to 12fps seem to result in better subjective quality than any type of quality-layer reduction. In the observed range between 25 and 12 fps temporal adaptation can be employed at any desired ratio. An observation was made in previous studies as well [2].

For quality switching, the period length is a crucial design criteria. Very short periods should be avoided because they introduce flickering at edges and in high frequency textures. We identified a flickering threshold at 2 seconds, but this observation needs further justification in additional experiments.

5.2 Fine-grained Bit-rate Adaptation

As shown by our results, layer switching has a large potential for exploiting available bandwidth that is between coarse-grained scaling opportunities. In this study we have only investigated switching patterns with a 1:1 ratio between frames from upper and lower operation points. A finer adaptation granularity can be achieved by adaptively turning this ratio, but several factors impose practical limits on switching schemes.

Very long switching periods require large averaging windows over which the desired target bitrate is achieved. This may conflict with fast bandwidth fluctuations which are typical over wireless channels and in congested networks. Missing a higher available bandwidth for short periods is acceptable, but missing bandwidth reductions due to a too large averaging window is unacceptable.

When prefetching, clients must provide sufficient buffer resources to store prefetched data at higher quality for later display. Layer switching can easily achieve a long-term average bitrate, but flickering can result when the averaging window is too short.

The layer switching pattern must be supported by the SVC encoding structure and synchronized to the decoder operation to avoid prediction errors. The switching patterns used in our study assumed short GOP sizes and frequent intra-updates to allow for short switching periods. Due to inter-frame prediction switching may not be possible at every frame boundary. Frequent layer switching points are usually in conflict with practical encoder setups that use multiple reference pictures, long GOPs and rare intra-updates for increased coding efficiency. This requires a trade-off at encoding time.

6. CONCLUSION

We have investigated whether we can achieve fine-grained video scalability using coarse-grained H.264 SVC without introducing the high overhead of MGS in a mobile device streaming scenario. This was tested by switching enhancement layers on and off to achieve the target bit-rate between CGS operation points. We tested different switching patterns against different downscaling patterns, and our subjective tests indicate that switching patterns with sufficient perceptual quality exist. For example, both the temporal switching pattern (TS, figure 2e) and quality switching pattern (QS, figure 2d), if above the flickering threshold, give

bit-rates between the encoded SVC base layer and the enhancement layer, and they were preferred compared to the quality-downscaled operation point (QD, figure 2c). Thus, we claim that such fine grained adaption is possible in different scenarios.

However, based on our preliminary tests, we cannot say which switching pattern will give the best possible result. This requires additional subjective studies. For example, we must further investigate the flickering threshold and the different ratios between high and low switching points. At this point, we have also only tested clips without scene changes. To further limit the perceived quality degradation of switching techniques, scene changes can for example be used as switching points.

7. REFERENCES

- [1] CRANLEY, N., PERRY, P., AND MURPHY, L. User Perception of adapting Video Quality. *International Journal of Human-Computer Studies* 64, 8 (2006), 637–647.
- [2] EICHHORN, A., AND NI, P. Pick your Layers wisely - A Quality Assessment of H.264 Scalable Video Coding for Mobile Devices. In *IEEE International Conference on Communications* (2009).
- [3] INTERNATIONAL TELECOMMUNICATIONS UNION. *ITU-T P.910. Subjective video quality assessment methods for multimedia applications*, 1999.
- [4] JEANNIN, S., AND DIVAKARAN, A. MPEG-7 Visual Motion Descriptors. *IEEE Trans. on Circuits and Systems for Video Technology* 11, 6 (Jun 2001), 720–724.
- [5] KNOCHE, H. O., AND SASSE, M. A. The sweet spot: how people trade off size and definition on mobile devices. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia* (New York, NY, USA, 2008), ACM, pp. 21–30.
- [6] MCCARTHY, J. D., SASSE, M. A., AND MIRAS, D. Sharp or Smooth?: Comparing the Effects of Quantization vs. Frame Rate for Streamed Video. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems* (2004), pp. 535–542.
- [7] PARK, D. K., JEON, Y. S., AND WON, C. S. Efficient use of Local Edge Histogram Descriptor. In *Proc. of ACM workshops on Multimedia* (2000), pp. 51–54.
- [8] SCHWARZ, H., MARPE, D., AND WIEGAND, T. Overview of the Scalable Extension of the H.264/AVC Video Coding Standard. *IEEE Trans. on Circuits and Systems for Video Technology* 17, 9 (2007), 1103–1120.
- [9] WENGER, S., YE-KUI, W., AND SCHIERL, T. Transport and Signaling of SVC in IP Networks. *IEEE Trans. on Circuits and Systems for Video Technology* 17, 9 (2007), 1164–1173.
- [10] WU, H., CLAYPOOL, M., AND KINICKI, R. On combining Temporal Scaling and Quality Scaling for Streaming MPEG. In *Proc. of NOSSDAV* (2006), pp. 1–6.
- [11] ZINK, M., KÜNZEL, O., SCHMITT, J., AND STEINMETZ, R. Subjective Impression of Variations in Layer Encoded Videos. In *Proc. of Intl. Workshop on Quality of Service* (2003), pp. 137–154.