Evaluation of Hybrid Scalable Video Coding for HTTP-based Adaptive Media Streaming with High-Definition Content

Michael Grafl, Christian Timmerer, Hermann Hellwagner
Alpen-Adria-Universität (AAU)
Klagenfurt, Austria
e-mail: *{firstname.lastname}*@itec.aau.at

Wael Cherif
Viotech Communications
France
e-mail: wael.cherif@viotech.net

Adlen Ksentini
Université de Rennes 1
France
e-mail: adlen.ksentini@irisa.fr

***Abstract*—Scalable Video Coding (SVC) in media streaming enables dynamic adaptation based on device capabilities and network conditions. In this paper, we investigate deployment options of SVC for Dynamic Adaptive Streaming over HTTP (DASH) with a special focus on scalability options, which are relevant for dynamic adaptation, especially in wireless and mobile environments. We evaluate the performance of SVC with respect to spatial and quality scalability options and compare it to non-scalable Advanced Video Coding (AVC). Performance evaluations are performed for various encoder implementations with high-definition (1080p) content. We show that a hybrid approach with multiple independent SVC bitstreams can have advantages in storage requirements at comparable rate-distortion performance.**

Keywords-scalable video coding; HTTP streaming; adaptation; high-definition; hybrid SVC-DASH

I. INTRODUCTION

Dynamic Adaptive Streaming over HTTP (DASH) [1] enables the client to select and adjust characteristics of a stream (e.g., spatial resolution and bitrate) on the fly while benefitting from existing HTTP infrastructures. While DASH is traditionally used with single-layer coding formats such as Advanced Video Coding (AVC), the usage of Scalable Video Coding (SVC) can offer further advantages in terms of adaptation capabilities and optimization of resource utilization [2][3].

The successful deployment of SVC in DASH strongly depends on proper and educated encoding configurations to facilitate adaptive streaming. This includes intelligent choices of scalability options, number of SVC layers, as well as spatial resolutions and bitrates. In particular, the increasing demand for media streaming over wireless networks and to mobile devices calls for bandwidth efficient streaming and adaptation to a heterogeneous context.

In this paper, we propose a hybrid SVC framework for DASH and high-definition (HD) content, comprising encoding guidelines and quality evaluations for various scalability options with a special focus on multiple resolutions. Therefore, we suggest using multiple independent SVC bitstreams, each bitstream's base-layer providing a given resolution corresponding to a certain device class (e.g., mobile, stationary, high-end) and allowing for signal-to-noise (SNR) adaptation through multiple enhancement layers which is aligned with today's SVC industry deployments, e.g., in video conferencing systems [4].

The remainder of this paper is structured as follows. In Section II, we give an overview of DASH and its deployment with SVC. Section III establishes encoding recommendations based on industry solutions for HTTP streaming. We validate those coding recommendations and investigate scalability options in Section IV, providing quality evaluations for major encoders. Section V concludes the paper and gives an outlook on future work.

II. DEPLOYMENT OF SVC IN DASH

With DASH, a server may offer multiple representations of the same content where each representation is typically characterized by – but not limited to – a specific resolution and bitrate. Those representations are described in an XML-based manifest file, called Media Presentation Description (MPD), which the client retrieves before starting the streaming session. The client picks the representation that is best suited for its current context (e.g., display resolution and available bandwidth). Each representation is split into temporal segments (e.g., 2-10 sec. each). The client can adapt to fluctuating network conditions by switching to lower bitrate representations at segment boundaries. Traditionally, representations are encoded as separate/independent (AVC) bitstreams. The deployment of SVC can bring some advantages in terms of storage (alleviating the need for multiple bitstreams of the same content to be stored at the server), cache performance [2][5], and adaptation [3]. With AVC, if the download of a segment cannot be completed before playout time, e.g., due to a sudden bandwidth decrease, the client has to decide whether to continue the download and risk stalling as segments need to be downloaded before decoding can start. Alternatively, the client may discard the current segment and switch to a lower representation which also increases the risk of stalling. The downloaded bits of the discarded segment are wasted.

SVC can be deployed in DASH as follows. Each representation contains an SVC layer and describes the dependencies between layers as shown in Listing 1. The `dependencyId` attribute indicates which other representations (i.e., lower SVC layers) are required for decoding a given representation. As long as the client has downloaded the SVC base layer, it can decode at least a basic representation

978-1-4673-5828-6/13/$31.00 ©2013 IEEE

```xml
<AdaptationSet>
  <Representation id="0" width="960" height="540"
        bandwidth="1200000">
    <SegmentList> <SegmentURL media="540p-BL-seg1.264"/>
    </SegmentList>
  </Representation>
  <Representation id="1" dependencyId="0" width="960"
        height="540" bandwidth="1975000">
    <SegmentList> <SegmentURL media="540p-EL1-seg1.264"/>
    </SegmentList>
  </Representation>  <!-- Further representations... -->
  <Representation id="4" dependencyId="0 1 2 3" width="1920"
        height="1080" bandwidth="4000000">
    <SegmentList> <SegmentURL media="1080p-EL4-seg1.264"/>
    </SegmentList>
  </Representation>  <!-- Further representations... -->
</AdaptationSet>
```

Listing 1. Simplified MPD for SVC streaming of multiple resolutions with a single bitstream featuring spatial scalability.

of the content, thus avoiding the risk of stalling. Each additional enhancement layer increases the video quality.

Related evaluations on SVC performance [8] and SVC-DASH [2][5] have not considered the bitrates and resolutions typically used by industry solutions. We argue that well-chosen SVC configurations are an important aspect towards a successful deployment of SVC for DASH-based services. Throughout this paper, we discuss and evaluate several deployment options for SVC in DASH. One option is to use a single SVC bitstream comprising all representations. The advantage of such a configuration is that the redundancy of having multiple similar bitstreams for a single content is removed. Furthermore, caching performance can be increased as all clients use the same SVC base layer. The downside of this approach is that the coding overhead increases with the number of SVC layers, specifically when covering a wide range of spatial resolutions. If the coding overhead becomes too high, it will outweigh the advantages of SVC.

Our proposal is to encode the content into multiple independent SVC bitstreams, one per resolution (e.g., representing certain device classes), and only relying on SVC quality scalability. The approach is referred to as *hybrid SVC-DASH* and the idea behind this approach is to confine the coding overhead by avoiding spatial scalability while benefitting from SVC's advantages. Provided a sufficient bitrate range for each bitstream for the purpose of dynamic adaptation, a client will try to maintain one resolution during the entire streaming session as resolution switches are more disturbing for the viewer than mere bitrate changes [6].

SVC offers two modes for quality scalability: coarse grain scalability (CGS), which uses mechanisms for spatial scalability but for a single resolution, and medium grain scalability (MGS), which offers a finer granularity for frame-based quality adaptation. In order to obtain a higher number of SVC quality layers for covering a higher range of bitrates, these two modes could be combined. However, the issue arises that not all of these layers are actually useful for a client as we will discuss later. In the following sections, we establish and validate encoding recommendations for SVC streaming.

TABLE I. BITRATE RECOMMENDATIONS FOR SVC STREAMING.

Resolution	Bitrate suggestions (4 bitrates) [kbps]	
	AVC streaming	SVC streaming
1920x1080	8000, 6000, 5000, 4000	10400, 7200, 5500, 4000
1280x720	6000, 4000, 2500, 1500	7800, 4800, 2750, 1500
960x540	2700, 2250, 1800, 1200	3500, 2700, 1975, 1200
640x360	1600, 1250, 900, 600	2075, 1500, 990, 600

III. SVC ENCODING RECOMMENDATIONS

Several industry solutions for HTTP streaming (Apple HTTP Live Streaming, Adobe HTTP Dynamic Streaming, Microsoft Smooth Streaming, YouTube, and MTV) provide guidelines for AVC-based deployment as discussed in [7].

Performance evaluations of SVC typically suggest a coding overhead of 10% per quality layer compared to AVC [8]. For SVC streaming with 4 layers per resolution, we adjust the bitrate suggestions as follows based on [7]. The bitrate for the base layer remains the same in order to provide at least a basic quality at low bandwidths. The first enhancement layer is increased by 10%, the second by 20% and the third by 30%, as compared to the corresponding AVC encoding bitrates. Table I provides bitrate recommendations for AVC and SVC streaming with 4 bitrates at resolutions from 1920x1080 (1080p) down to 640x360. Quality evaluations based on these recommendations are given in the following section.

IV. SVC ENCODING PERFORMANCE

In this section, we validate the aforementioned SVC coding recommendations for various encoders and provide rate-distortion (RD) performance evaluations for several scalability options.

A. Test-bed Setup

Four high-definition (1080p) test sequences were selected based on their Spatial Information (SI) – i.e., amount of spatial detail – and Temporal Information (TI) – i.e., amount of motion – defined in [9] to cover different characteristics of video content: *PedestrianArea* (low SI, low TI), *Dinner* (low SI, high TI), *DucksTakeOff* (high SI, low TI), and *CrowdRun* (high SI, high TI). The *Dinner* sequence has a frame rate of 30 fps, the other sequences have 25 fps. The first 250 frames of each sequence were encoded.

We tested the AVC encoder *x264*[1] and the following major SVC encoders: SVC reference software *Joint Scalable Video Model* (JSVM)[2], *MainConcept*[3], *VSS*[4], and *bSoft*[5]. While the MainConcept and VSS encoders use requantization for MGS layers, the bSoft encoder distributes transform coefficients automatically across layers (also known as MGS vectors). The JSVM encoder supports both

---

1 http://www.videolan.org/developers/x264.html
2 Joint Scalable Video Model (JSVM), Version 9.19.15, 2011
3 http://mainconcept.com/
4 http://www.vsofts.com/technology/scalable-video-coding.html
5 http://bsoft.net/

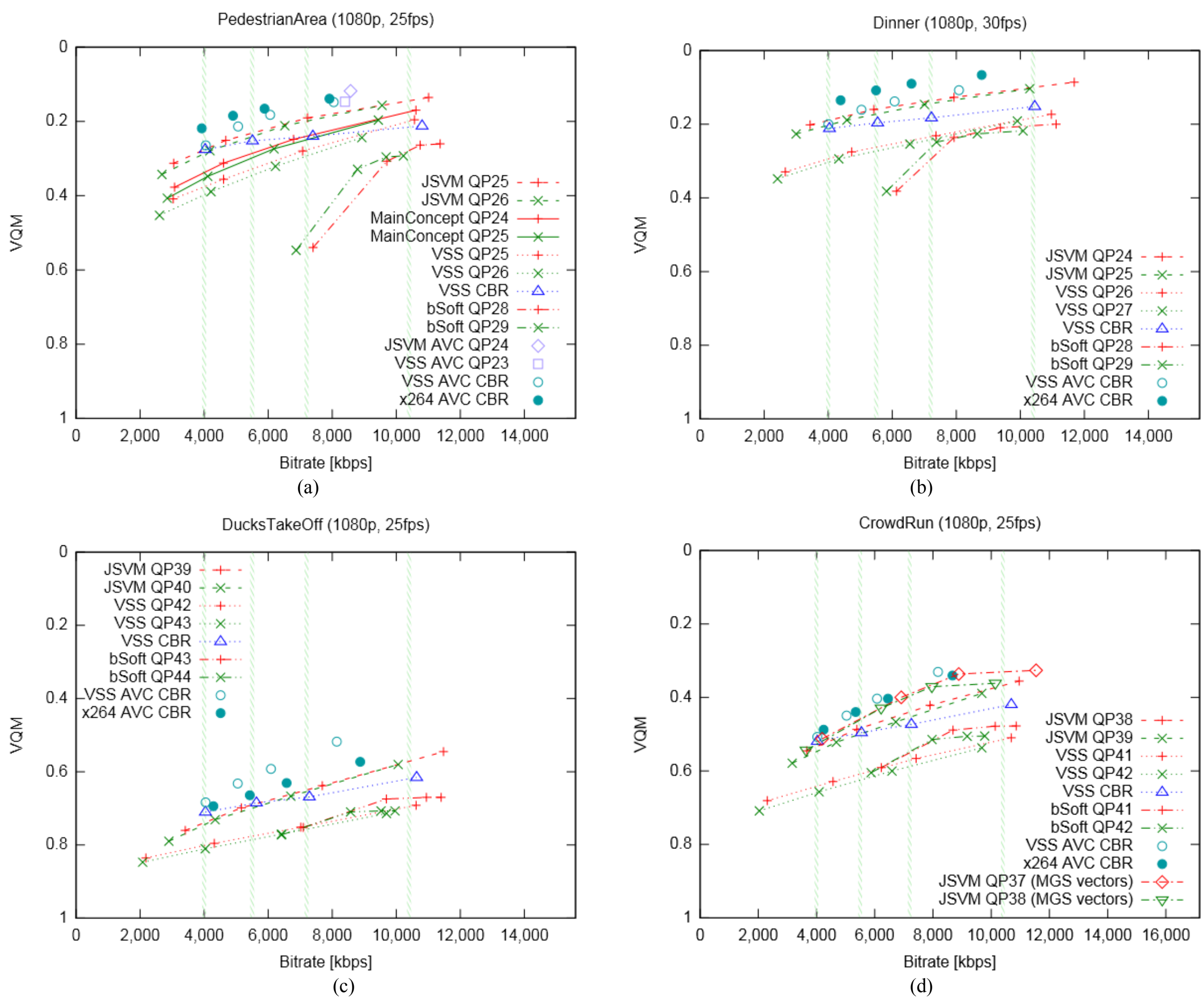


Figure 1. VQM results of AVC and SVC with 4 bitrates for (a) PedestrianArea, (b) Dinner, (c) DucksTakeOff, and (d) CrowdRun sequences.

behaviors (i.e., requantization and MGS vectors via manual distribution of transform coefficients) [10].

The encoders were configured with an intra coded picture period of 32 and the entropy coding mode set to context-adaptive binary arithmetic coding (CABAC). For SVC encoding with fixed QP rate control mode and requantization between MGS layers, the deltaQP was set to 2 as suggested in [7]. The deltaQP denotes the QP difference between two MGS layers. It controls the bitrate distance and the quality gap between those layers. The value was chosen to meet the bitrate suggestions of Table I.

Our RD performance evaluations are based on Peak Signal-to-Noise Ratio (PSNR) and the NTIA Video Quality Metric (VQM)[6]. While PSNR is a widely used metric for video quality evaluation, VQM yields a better correlation with the human visual system [11].

[6] http://www.its.bldrdoc.gov/vqm

B. Encoder Comparison and Bitrate Validation

We first compare the RD performance of the x264 encoder to SVC encoders in order to establish a base line for our further tests. For SVC we use a single-layer configuration (i.e., an AVC-compatible base layer) and a configuration with 4 MGS layers. Single-layer (AVC) bitstreams are encoded in CBR mode with target bitrates suggested for AVC streaming. SVC bitstreams with 4 MGS layers are encoded in fixed QP rate control mode for all SVC encoders. The JSVM, MainConcept, and VSS encoders were tested with requantization between MGS layers, the bSoft encoder with MGS vectors under automatic distribution of transform coefficients. We also tested the JSVM encoder using MGS vectors with a partitioning into three MGS slices containing 1, 2, and 13 transform coefficients. The partitioning was found through empirical testing to best match the recommended bitrates. The deltaQP between the base layer and the enhancement layer was set to 6, which amounts to the same as the 3 requantized enhancement layers

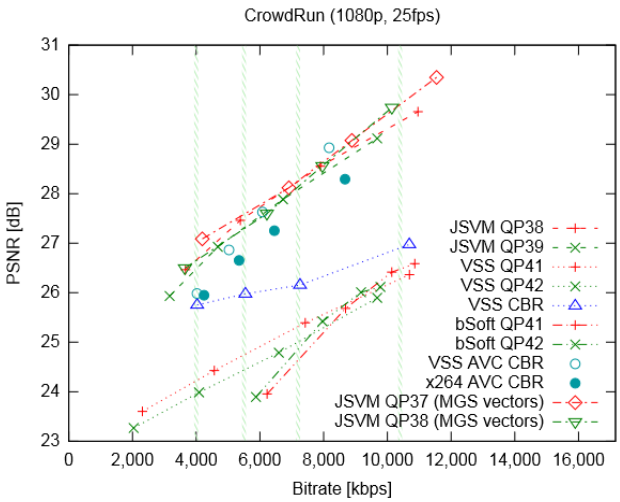


Figure 2. PSNR results of AVC and SVC encoders with 4 bitrates.

with a deltaQP=2. Additionally, the sequences were encoded with the VSS encoder in constant bitrate (CBR) mode as it was the only one of the tested SVC encoders to provide decent CBR support at all tested resolutions.

The VQM results for the tested sequences at 1080p resolution are shown in Fig. 1. Note that the y-axis of VQM results is an impairment scale from 1 (high distortion) to 0 (no distortion), indicating the expected quality of a video. For fixed QP mode, bitstreams with bitrates just below and just above the bitrate suggestions for the highest SVC layer (cf. Table I) are shown. The results for the MainConcept encoder and for AVC configurations in fixed QP mode are only shown for *PedestrianArea* for the sake of readability. The RD performance of the MainConcept encoder in relation to JSVM and VSS for the *PedestrianArea* sequence is representative for the other sequences. Results for the JSVM encoder in MGS vector mode are only shown for the *CrowdRun* sequence for the same reason.

As expected, AVC yields a higher RD performance than SVC with multiple MGS layers. However, at the lowest bitrate, the SVC bitstream from the VSS encoder in CBR mode (labeled *VSS CBR*) has only marginal overhead compared to the corresponding AVC bitstream from the same encoder (labeled *VSS AVC CBR*). Whether the x264 encoder outperforms VSS depends on the content.

The JSVM encoder with 4 MGS layers tends to reach the quality of the AVC encoders in several cases if we consider the expected SVC coding overhead discussed in Section III. Among SVC encoders, the JSVM yields the best RD performance, followed by MainConcept, VSS and bSoft. Note that the bSoft encoder distributes transform coefficients among MGS layers while other encoders perform manually configured requantization. This automatic distribution allocates only poor quality to the base layer. However, the bSoft encoder outperforms VSS for more complex sequences such as *CrowdRun*. For the JSVM encoder, the MGS vector mode yields slightly higher RD performance than requantization. Starting from virtually the same base layer quality as the requantization mode, the first two MGS slices

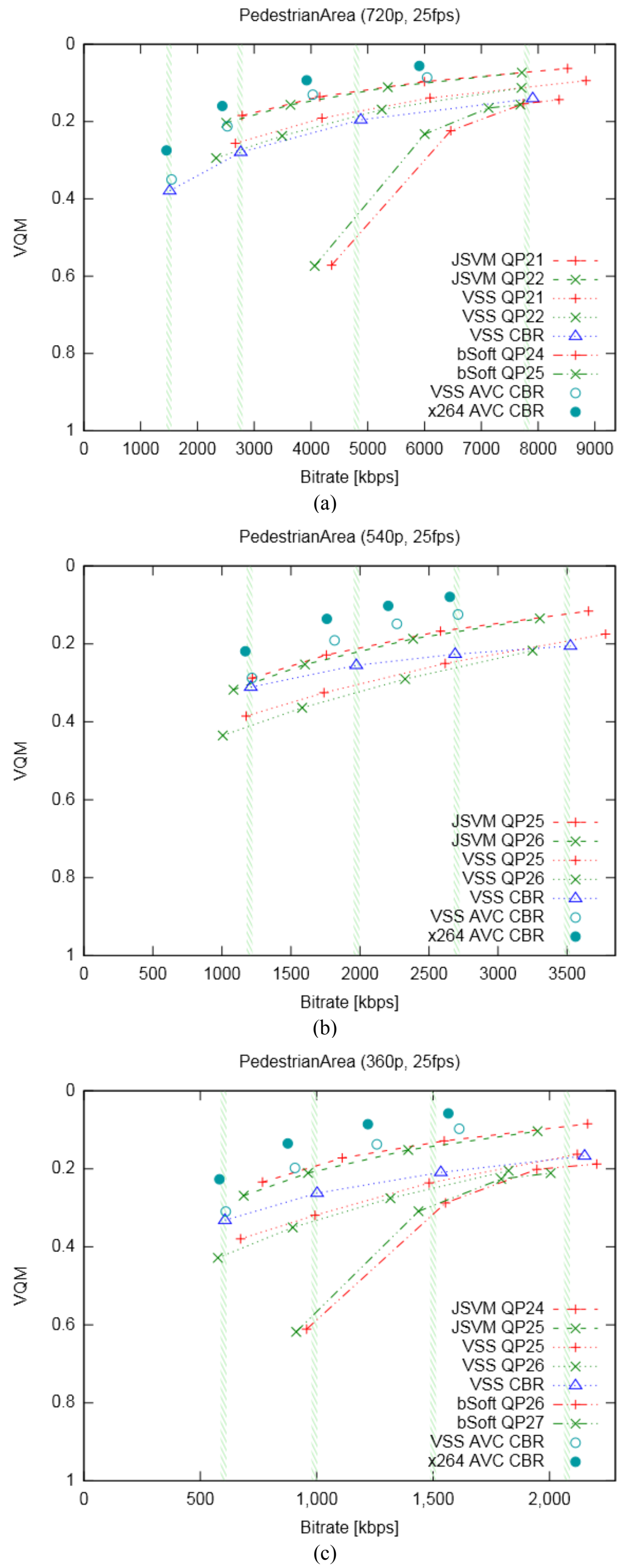


Figure 3. VQM results of AVC and SVC encoders with 4 bitrates at (a) 1280x720, (b) 960x540, and (c) 640x360 resolutions.

have higher RD performance than requantization mode.

```xml
<AdaptationSet>
  <Representation id="0" width="960" height="540"
        bandwidth="1200000">
    <SegmentList> <SegmentURL media="540p-BL-seg1.264"/>
    </SegmentList>
  </Representation>
  <Representation id="1" dependencyId="0" width="960"
        height="540" bandwidth="1975000">
    <SegmentList> <SegmentURL media="540p-EL1-seg1.264"/>
    </SegmentList>
  </Representation>  <!-- Further representations... -->
<Representation id="4" width="1920" height="1080"
        bandwidth="4000000">
    <SegmentList> <SegmentURL media="1080p-BL-seg1.264"/>
    </SegmentList>
  </Representation>  <!-- Further representations... -->
</AdaptationSet>
```

Listing 2. Simplified MPD for SVC streaming of multiple resolutions with one bitstream per resolution.

From the second MGS slice to the full MGS enhancement layer, VQM results almost stagnate. We find this behavior for other sequences as well.

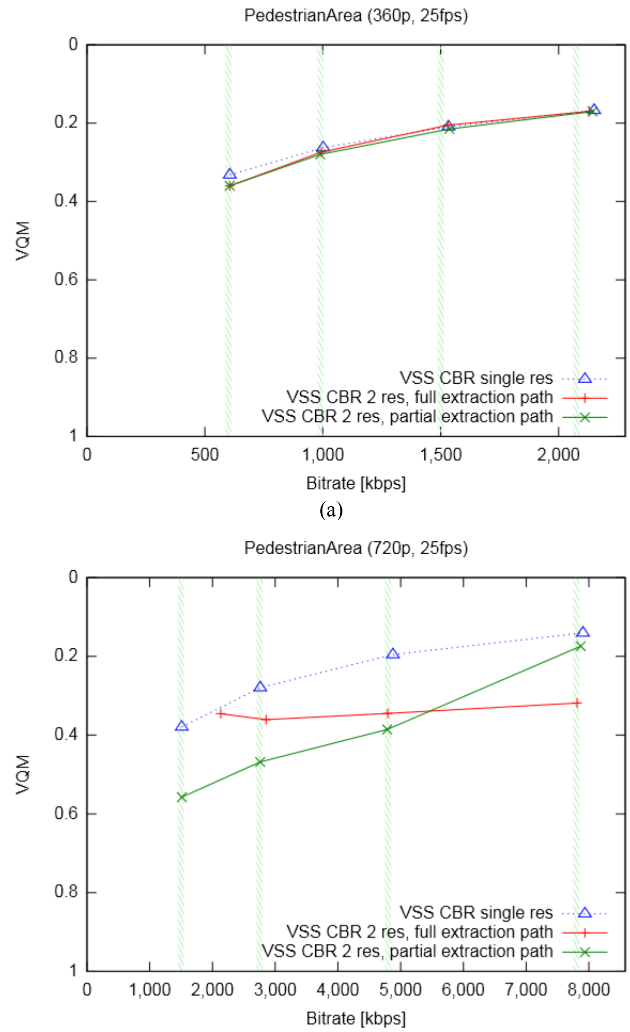
We note that the VSS encoder in CBR mode yields a surprisingly high quality at the base layer, but enhancement layers only bring little quality increase. On the one hand, such a low slope in RD performance makes bitrate switches in media streaming less perceivable. On the other hand, a higher bitrate that does not yield higher quality is basically a waste of bandwidth. Thus, we suggest that the bitrate recommendations for the base layer (i.e., 4,000 kbps) could be reduced, particularly with regard to mobile environments, even though low base layer quality impairs the quality of other SVC layers also [7].

For comparison, Fig. 2 shows the PSNR results for the *CrowdRun* sequence. It can be observed that several encoders yield better VQM performance than the PSNR results indicate. For example, the PSNR results for AVC bitstreams are below the RD performance of the JSVM, while the VQM results show the opposite. We find this behavior for encoding in CBR mode for various sequences and for the bSoft encoder at lower layers for all sequences. In contrast to the corresponding VQM results, the JSVM encoder with MGS vector mode only outperforms requantization mode at the full MGS enhancement layer in terms of RD performance.

Fig. 3 shows the VQM results at lower resolutions for the *PedestrianArea* sequence at the recommended bitrates. Again, results for the MainConcept encoder are omitted for the sake of readability.

As with 1080p, the JSVM encoder with 4 MGS layers tends to reach the quality of the AVC encoders in most cases considering the expected SVC coding overhead.

In terms of storage requirements (cf. Section II), SVC is more efficient than AVC with multiple representations for 4 MGS layers (if accepting small quality reductions in some cases) as shown in Table II. On average, an SVC bitstream requires 48% the disc space of the 4 AVC representations. Of course, the storage reduction comes at the cost of the discussed SVC coding overhead for every streaming session.



(b)

Figure 4. VQM results of spatial scalability for the VSS encoder. The lines labeled VSS CBR 2 res represent single bitstreams ranging over both resolutions (a) 640x360 and (b) 1280x720.

Fixed QP rate control mode is not designed to meet a specific bitrate. Thus, the bitrate may vary between frames. In particular, SVC has been shown to have high bitrate variability that poses challenges for media streaming [12]. For DASH, traffic variability only matters on a per-segment basis. Intuitively, the per-segment traffic variability is lower than that of individual frames. For SVC, DASH clients request the base layer prior to the enhancement layers of a segment. If an enhancement layer is not fully downloaded due to traffic variability, it will merely result in the playback of a lower quality. Additionally, typical DASH clients buffer three or more 2-second segments [3], further alleviating the impact of traffic variability. Due to the short duration of the test sequences (250 frames), we did not evaluate the bitrate variability in our tests.

The bitrate recommendations from Table I for 4 MGS layers yield consistent qualities for all resolutions. The applied requantization with a deltaQP of 2 for fixed QP mode correlates with the bitrate suggestions of lower layers

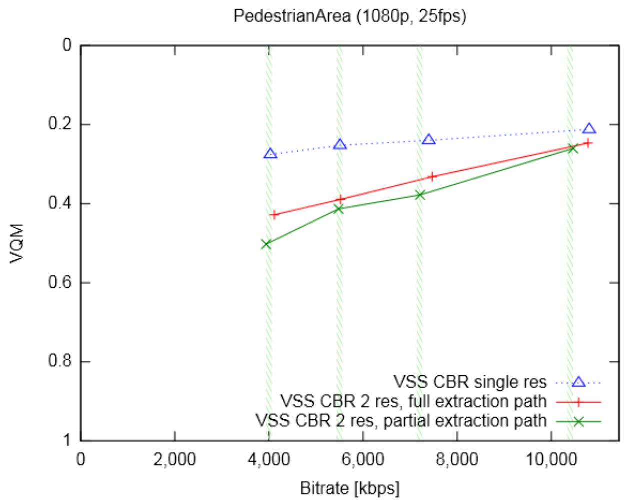


Figure 5. VQM results of spatial scalability for the VSS encoder. The lines labeled VSS CBR 2 res represent single bitstreams ranging over both resolutions 960x540 and the depicted 1920x1080.

to a reasonable extent for JSVM, MainConcept, and VSS encoders as further discussed in [7].

C. *Combination of Spatial Scalability and MGS*

SVC streaming of multiple resolutions can be achieved by either encoding one SVC bitstream that features spatial scalability or to encode several bitstreams, one per resolution. Example MPDs for the two approaches are depicted in Listing 1 and Listing 2 respectively.

In this section we evaluate the RD performance of SVC bitstreams with both spatial and quality scalability compared to hybrid SVC-DASH with just quality scalability.

An important aspect for a proper comparison is the way in which lower layers are obtained from an SVC stream that combines spatial and quality scalability. In SVC, each layer is identified by its dependency (i.e., resolution), quality, and temporal id, commonly denoted *DQT*. We consider two different extraction paths for achieving spatial scalability. An extraction path denotes the order in which SVC layers are removed from the stream. A quality layer q of an upper resolution d, e.g., DQT=(d,q,0), can either depend on the same quality layer of the previous resolution, i.e., DQT=(d-1,q,0), or on the highest layer Q of the previous resolution, i.e., DQT=(d-1,Q,0). The first extraction path (subsequently denoted *partial extraction path*), which is implemented in the JSVM reference software, yields a lower bitrate at the expense of discarded enhancement information from the lower resolution. The VSS encoder also supports the second extraction path (subsequently denoted *full extraction path*).

TABLE II. STORAGE REQUIREMENTS FOR SVC STREAMING PER RESOLUTION.

Resolution	AVC bitstreams	SVC bitstream	Reduction
1920x1080	23,000 kbps	10,400 kbps	54.8%
1280x720	14,000 kbps	7,800 kbps	44.3%
960x540	7,950 kbps	3,500 kbps	55.8%
640x360	4,350 kbps	2,080 kbps	52.2%

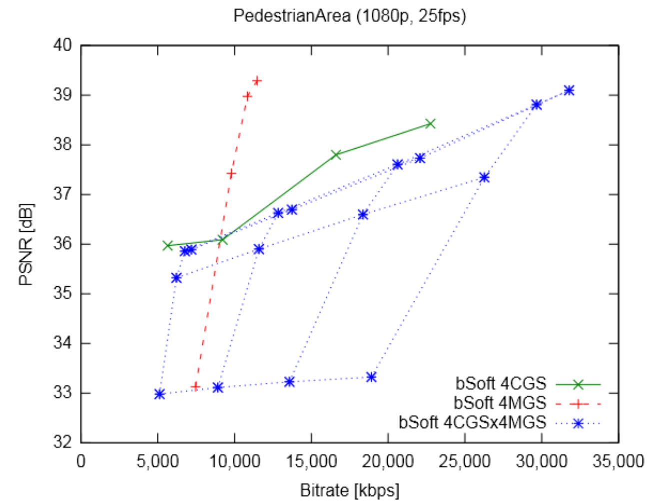


Figure 6. PSNR results for combination of CGS and MGS for the bSoft encoder.

Fig. 4 shows VQM results for both extraction paths at resolutions 640x360 and 1280x720 for the VSS encoder. Note that the bitstreams range over both resolutions. Single resolution SVC bitstreams are shown for comparison.

At the lower resolution, both extraction paths have roughly the same RD performance as the single resolution bitstream with only a slight overhead at the base layer. Note however that the PSNR results for both extraction paths are at the base layer 0.2 dB lower and at the highest layer around 0.7 dB lower than for the single resolution bitstream.

At the higher resolution, the *full extraction path* starts at a quality that is on par with the single resolution bitstream RD performance. Since the bitstream for *full extraction path* depends on the highest layer of the lower resolution, it starts at a bitrate of 2,134 kbps. Subsequent enhancement layers do not increase the quality of the bitstream; rather the first enhancement layer even reduces the quality.

On the other hand, the *partial extraction path* starts at a low quality but increases almost to the quality of the single resolution bitstream for the highest layer. Even with the low starting quality, we argue that the *partial extraction path* as far better suited for multi-resolution SVC streaming.

The VQM results for both extraction paths at resolutions 960x540 and 1920x1080 are shown in Fig. 5. As there is only negligible loss at 960x540 (similar to Fig. 4 (a)), only the higher resolution results are shown. Since the lower resolution has a target bitrate of 3,500 kbps at the highest layer and the higher resolution starts at 4,000 kbps, the *full extraction path* is able to meet that target bitrate and the quality increases with enhancement layers at the higher resolution. Still, we consider the *partial extraction path* to be better suited for spatial scalability in SVC streaming.

TABLE III. PSNR LOSS FOR SPATIAL SCALABILITY OF *PARTIAL EXTRACTION PATH* COMPARED TO *HYBRID SVC-DASH*.

Resolution	Layer 0	Layer 1	Layer 2	Layer 3
Resolution 1	0.13 dB	0.25 dB	0.31 dB	0.47 dB
Resolution 2	2.54 dB	2.64 dB	3.00 dB	0.77 dB

In terms of PSNR, the average quality loss due to coding overhead across all sequences and both resolution pairs are shown in Table III for the *partial extraction path*. As spatial scalability only yields around 23% reduction of storage requirements, we argue that one SVC bitstream per resolution is better suited for the given use case.

D. *Combination of CGS and MGS*

In the following test, we evaluate the RD performance for combining CGS and MGS modes in one bitstream. The encoding configuration comprises 4 CGS layers and 4 MGS layers, resulting in 16 quality layers per stream.

Fig. 6 shows the PSNR results for the combination of CGS and MGS for the *PedestrianArea* sequence encoded with the bSoft encoder with the QP at the highest layer set to 28. For comparison, PSNR results of the bitstream with 4 CGS layers and the bitstream with 4 MGS layers are also shown. The combination of CGS and MGS is depicted with lines that show the possible extraction paths for each quality layer. For example, starting at the base layer, we can either add one MGS layer, resulting in the layer with DQT=(0,1,0), or add one CGS layer in order to obtain the layer with DQT=(1,0,0). From either of these two layers, the layer with DQT=(1,1,0) can be reached.

Due to the sharp decrease of PSNR for lower layers for MGS mode as observed in Section IV.B, also the combination of CGS and MGS suffers from this behavior along MGS layers. Thus, the depiction of PSNR results resembles a grid, where MGS layers form the vertical lines. This also means that the bitstream contains many extraction points that just have a high bitrate but very low PSNR. In particular, adapting to the layers with DQT values of (1,0,0), (2,0,0), (3,0,0), (1,1,0), (2,1,0), or (3,1,0) would be a waste of bandwidth. We conclude that out of the entire 16 SVC layers only the 10 layers forming the outer curve of *bSoft 4CGSx4MGS* are useful for adaptation in terms of RD trade-off, but at poor overall RD performance of the SVC bitstream.

The discussed configuration of 4 CGS layers and 4 MGS layers was not supported by the tested version of the VSS encoder. The configurations of the JSVM and MainConcept encoders do not allow for combination of CGS and MGS.

V. CONCLUSIONS

In this paper, we have investigated deployment options of SVC for Dynamic Adaptive Streaming over HTTP (DASH) with a special focus on scalability options. We performed several performance evaluations of major encoder implementations with HD content. Our tests have validated the bitrate recommendations deduced from industry solutions. We argue that the target bitrate of the base layer could be further reduced, depending on the scenario. Our findings suggest that a hybrid SVC-DASH approach with one SVC bitstream featuring quality scalability per resolution provides a good trade-off between the advantages of SVC and its coding overhead. Furthermore, we tested the combination of CGS and MGS modes in one bitstream. The results show that 10 out of 16 SVC layers are useful for adaptation but at poor overall RD performance.

Our future work will focus on integrating our findings with an adaptive end-to-end media delivery system. We will also investigate other codecs such as High Efficiency Video Coding (HEVC) and its scalable extensions.

ACKNOWLEDGMENT

This work was supported in part by the EC in the context of the ALICANTE project (FP7-ICT-248652). The authors thank MainConcept GmbH, Vanguard Software Solutions Inc., and bSoft Ltd. for providing their encoders.

REFERENCES

[1] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming over the Internet," *IEEE Multimedia*, vol. 18, no. 4, 2011, pp. 62-67.

[2] Y. Sánchez et al., "iDASH: improved dynamic adaptive streaming over HTTP using scalable video coding," *ACM MMSys 2011*, Chapel Hill, NC, USA, 2011, pp. 257–264.

[3] C. Müller et al., "Using Scalable Video Coding for Dynamic Adaptive Streaming Over HTTP in Mobile Environments," *EUSIPCO 2012*, Bucharest, Romania, 2012.

[4] A Polycom Whitepaper, "More Scale at Lower Cost with Scalable Video Coding," Oct. 2011. http://docs.polycom.com/global/documents/whitepapers/more-scale-lower-cost-scalable-video-coding-wp.pdf (last access: Apr. 2013)

[5] Y. Sánchez et al., "Efficient HTTP-based streaming using Scalable Video Coding," Signal Processing: Image Communication, vol. 27, no. 4, pp. 329–342, Apr. 2012.

[6] P. Ni et al., "Frequent layer switching for perceived quality improvements of coarse-grained scalable video", *Multimedia Systems*, vol. 16, no. 3, 2010, pp. 171-182.

[7] M. Grafl, C. Timmerer, H. Hellwagner, W. Cherif, and D. Negru, "Scalable Video Coding Guidelines and Performance Evaluations for Adaptive Media Delivery of High Definition Content," *ISCC 2013 Submission*, unpublished. Available http://itec.aau.at/~mgrafl/am.pdf.

[8] M. Wien, H. Schwarz, and T. Oelbaum, "Performance Analysis of SVC," *IEEE Trans. on CSVT*, vol. 17, no. 9, 2007, pp. 1194-1203.

[9] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," 2008.

[10] H. Kirchhoffer, H. Schwarz, and T. Wiegand, "CE1: Simplified FGS", Joint Video Team (JVT), Doc. JVT-W090, Apr. 2007. http://wftp3.itu.int/av-arch/jvt-site/2007_04_SanJose/JVT-W090.zip (last access: Apr. 2013)

[11] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011..

[12] G. Van der Auwera, P. David, and M. Reisslein, "Traffic and Quality Characterization of Single-Layer Video Streams Encoded with the H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 698–718, 2008.