Analysis of H.264/AVC Scalable Video Coding for Video Delivery to Heterogeneous Terminals

Thomas Rusert, Rickard Sjöberg, and Clinton Priddle

Ericsson Research, Multimedia Technologies
Stockholm, Sweden
{thomas.rusert,rickard.sjoberg,clinton.priddle}@ericsson.com

Abstract. In 2007, the ITU-T H.264 | ISO/IEC MPEG-4 AVC standard was extended to support temporal, spatial and fidelity (SNR) scalability in a framework that is referred to as Scalable Video Coding (SVC). Since the development of this SVC extension, its use has been proposed for several applications. It seems however there is not yet a broadly agreed understanding about the benefits of SVC compared to non-scalable coding. In this paper, we describe coding efficiency gain and cost measures for scalable video against non-scalable simulcast, and single layer non-scalable coding, respectively, in the context of video delivery to heterogeneous terminals. Our results show that the cost and gain from SVC are strongly dependent on the application and conditions. Specifically, it is shown that while SVC can theoretically provide promising gain in some applications, its cost is not negligible and in some cases this cost can outweigh the gain.

Keywords: SVC, H.264/AVC, standards, video, video coding, compression.

1 Introduction

A new scalable video coding framework, commonly referred to as SVC, has been standardized as an extension of the ITU-T H.264 | ISO/IEC MPEG-4 AVC standard [1] in 2007. While the use of SVC has been discussed in several applications [2], the benefits and limitations compared to non-scalable coding are not yet clear in practical operating points. In this paper, we provide more insight into the gain and cost that SVC induces compared to solutions based on non-scalable H.264/AVC profiles, which in the following we refer to as AVC. Our particular focus is on analyzing the use of SVC for video delivery to populations of terminals with unequal capabilities such as screen size, available bandwidth, and processing power, i.e. heterogeneous terminals.

Based on a layered coding principle, SVC enables coding of hierarchical video representations. The base layer represents the video sequence in the lowest quality, while one or more enhancement layers provide successive enhancement in terms of temporal resolution, spatial resolution or SNR quality. Scalability is provided along all of these dimensions, since the corresponding enhancement layer data may be omitted from the bit stream while the resulting bit stream is still decodable.

J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.): MOBIMEDIA 2010, LNICST 77, pp. 619–633, 2012.

Temporal scalability is not new with SVC – it can likewise be provided with AVC. On the other hand, simultaneous support of multiple spatial resolutions or SNR qualities with AVC requires transmission of multiple independent AVC streams, hereon referred to as AVC simulcast. Since AVC simulcast in contrast to SVC does not exploit redundancies between the representations, SVC can be expected to be more efficient. On the other hand, since provision of multiple video representations in an SVC bit stream imposes certain constraints compared to non-scalable coding, the coding efficiency for each individual video representation in the SVC bit stream can be expected to be lower when compared to dedicated (single layer) AVC coding.

SVC and AVC differ significantly with respect to encoders and decoders. In practical applications these differences will in turn imply differences in terms of coding efficiency, but also in other aspects. This includes video adaptation efficiency, error robust coding and transmission, error concealment and tune-in delay, as well as encoder and decoder complexity, and transmission complexity. In this paper, we focus on the differences in coding efficiency and their implications with respect to practical applications.

The paper is organized as follows. Firstly, a description of the delivery scenarios covered in this paper is given in Sect. 2. SVC efficiency measures that are used to evaluate the results presented later in the paper are formulated in Sect. 3. Sect. 4 provides experimental results for AVC and SVC coding, and in the case of AVC includes two different encoders. Sect. 5 is a discussion of the experimental results in the context of the delivery scenarios described in Sect. 2. Sect. 6 concludes the paper.

2 Video Delivery to Heterogeneous Terminals

Video delivery to heterogeneous terminals refers to a broad set of scenarios where video content is delivered to multiple terminals with different capabilities. The differences in capabilities can include different screen sizes, different connectivity, and different processing power. It is a challenge in such scenarios to deliver the video such as to maximize the video quality available on each terminal while minimizing the cost for delivering the video through the network.

Given that SVC can be used for coding video in several different representations simultaneously, it provides a potential solution to the above problem. It is important to note though that this is not a new functionality provided by SVC but that other, non-scalable, coding solutions can be likewise used. Compared to such alternative solutions, SVC does have potential to improve the efficiency of video delivery. The actual benefit SVC can provide, as will be shown later, depends a lot on the delivery mechanism and system settings in use.

In this paper, we focus on two major video delivery mechanisms, namely unicast video delivery (such as in on-demand TV or internet streaming), and broadcast/multicast video delivery (such as used in DVB-T, DVB-H, MBMS, or IPTV systems). Considering these mechanisms, we compare the benefit SVC provides over non-scalable coding. With respect to the evaluations performed in this paper, these quite broad scenarios cover many use cases on top of traditional TV.

Note that we do not consider network-based transcoding [8] of video content here. Transcoding in the network may be too complex or not even possible (if the content is encrypted and the network not trusted) in practical video delivery scenarios.

2.1 Video Coding Options

We consider three different approaches for video coding and delivery, single layer coding, simulcast, and scalable coding.

Single layer coding is referred to video coding at a single dedicated spatial resolution, frame rate and bit rate. If the video is targeted at a single dedicated terminal or multiple terminals with the same capabilities, the coding can be done in a way that optimally suits the capabilities of the terminal(s). Thus the video quality for a given bit rate can be assumed to be optimal given the constraints of the video coding standard and encoder, or in other words, the maximum possible coding efficiency is achieved. If multiple terminals with different capabilities need to be addressed with the same single layer bit stream, then the bit stream needs to be encoded such that it can be decoded by all targeted terminals. After decoding, the video content needs to be adapted to the display capabilities of the respective terminal (e.g. scaling the incoming video to the actual display resolution of the terminal). We denote this option as terminal side adaption. While terminal side adaptation can be very efficient when terminal capabilities are similar, the provisioning of several different dedicated video representations becomes relevant when the spread of terminal capabilities is larger.

Simultaneous transmission of multiple representations of the same video content in independently coded single layer bit streams is denoted as simulcast. It can be used for transmission of different representations of the same content to terminals with different capabilities. The simulcast bit rate is the sum of bit rates of the independently coded video representations.

Having multiple video representations embedded in a single bit stream is referred to as scalable coding, for which SVC is an example. For simultaneous transmission of multiple video representations, scalable coding can be assumed to be more efficient than simulcast of single layer bit streams, since redundancies among representations can be exploited. The amount of savings over simulcast depends on the coding efficiency of the scalable coder compared the coding efficiency of the single layer coder used for simulcast coding. This is the focus of the rest of this paper.

2.2 Unicast Video Delivery

Consider a video on demand system where the video is stored at a server and on request transported to its clients via a dedicated point-to-point connection (unicast). The server may serve clients with different capabilities (e.g. maximum video resolution or downlink bit rate) by choosing from a set of pre-encoded bit streams or layers. The scenario is illustrated in Fig. 1.

In this scenario there are two main aspects to consider in terms of coding efficiency: The video storage space at the server, and the bit rates required for transmission in the network.

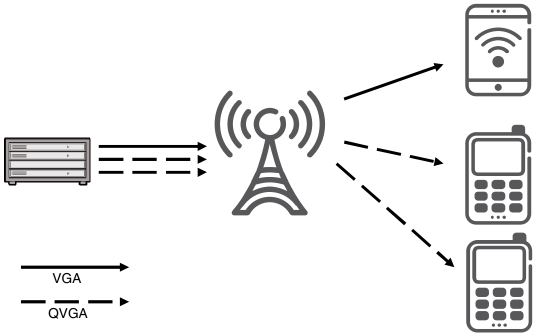

Fig. 1. Example of unicast delivery to heterogeneous terminals

If the capabilities of the receiving terminals are very similar, then single layer coding with terminal side adaptation may be the most viable solution. If that is not the case, and a dedicated point-to-point connection is used, it is possible to perform video adaptation with respect to terminal capabilities at the server side. One example of a server side adaptation mechanism is to let the terminal send a video resolution request to the server. The server then selects what version to transmit at session startup. In addition to this, the server may switch streams during transmission. In both cases, multiple content representations have to be stored at the server side. Thus, since scalable coding can provide bit rate savings by exploiting redundancies among representations, there are potential gains in terms of reduction of required storage at the server side. On the other hand, any inefficiency that scalable coding may exhibit compared to single layer coding will be disadvantageous in terms of bit rates required in the network.

2.3 Broadcast and Multicast Video Delivery

Consider a system similar to the unicast video delivery system where video is distributed to heterogeneous terminals (e.g. different maximum video resolution), but transported via a broadcast and/or multicast enabled network, i.e. in a tree-structured or one-to-many type of delivery scheme. An example of a multicast system is illustrated in Fig. 2.

In this scenario, the two main aspects to be considered with respect to video coding efficiency are the bit rates required for transmission in the backend network, and the bit rates required for transmission on the last network hop (e.g. a DSL link in an IPTV system, and the radio link in a mobile broadcast system).

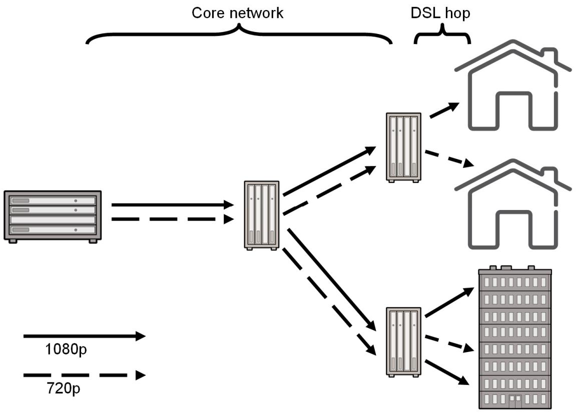

Fig. 2. Example of multicast delivery to heterogeneous terminals

As in the case of unicast delivery, terminal side adaptation is one option in this scenario. If terminal side adaptation is not reasonable, e.g. because the terminal capabilities are too different, the video needs to be transmitted in multiple representations. Then either simulcast or scalable coding can be used. The potential advantage with scalable coding is that bit rate can be saved in the backend network. On the other hand, in case a dedicated last hop exists in the network (e.g. the DSL link in a multicast IPTV system) over which only a single video representation needs to be provided to a single terminal (or a population of equivalent terminals), the advantage of simulcast is that the bandwidth available on the last hop can be optimally utilized, assuming that the simulcast includes a single layer coded video representation that optimally suits the requirements of the respective terminal.

3 SVC Coding Efficiency Measures

SVC supports three dimensions of scalability; temporal (frame rate), spatial (frame resolution) and SNR (pixel fidelity) scalability. Temporal scalability has been supported in AVC profiles already before SVC was introduced and can be provided without use of SVC. Thus it is not further considered here.

On top of what is available in the non-scalable AVC toolbox, several so-called inter-layer prediction tools were introduced in SVC to support SNR scalability and spatial scalability. The major new scalable coding tools added are inter-layer intra prediction, inter-layer motion prediction, and inter-layer residual prediction [2]. Using these tools when coding multiple different bit rates and/or spatial resolutions, higher SVC layers can re-use information from lower layers to reduce the bit rate for coding of the higher layers. The base layer is coded without the use of scalable coding tools, and thus can be decoded by a non-scalable AVC decoder. The use of scalable coding tools in higher layers leads to a bit rate reduction compared to simulcast coding, and we denote this bit rate reduction as the *SVC gain*.

On the other hand, although the inter-layer prediction tools in SVC help in exploiting redundancy among SVC layers, the bit rate savings provided by those tools typically cannot fully compensate for the bit rate overhead caused by the constraints imposed when multiple different video representations are coded together. In other words, providing a certain video quality using spatial or SNR scalable coding can be expected to require a higher bit rate than non-scalable AVC coding would require. We denote the additional bit rate required with SVC as *SVC cost*.

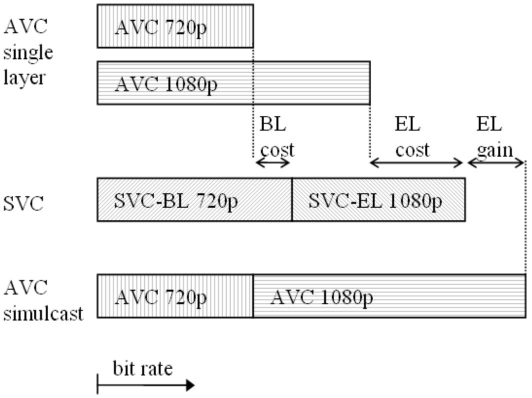

Fig. 3. SVC cost and gain compared to AVC for base layer (BL) and enhancement layer (EL)

Fig. 3 illustrates how SVC gain and cost measures are associated with different SVC layers. Note that the distribution of SVC cost and gain among layers depends on the encoding control used. A straightforward SVC encoding control would encode the base layer first, and then successively encode enhancement layers on top of that ("bottom-up" approach [6]). In this case, since all AVC coding tools are available, AVC single layer performance is achieved for the base layer, i.e. the base layer cost is zero. More sophisticated SVC encoding schemes such as [6] may perform joint encoding of multiple layers, trading the SVC costs among the layers more equally.

In the following, SVC gain and cost measures are formalized in order to allow for detailed comparison of coding options.

3.1 Relative SVC Cost and Gain

Assume a video sequence is delivered in N different qualities (different resolutions and/or bit rates). Using non-scalable AVC coding, N bit streams are required to represent the video sequence according to the different qualities. Let us assume that the representations are sorted in order of increasing quality, and the n^{th} bit stream ($1 \leq n \leq N$) is encoded at a bit rate $R_{\text{AVC},n}$.

Now using SVC, let $R_{\text{SVC},1..n}$ be the bit rate required to code the first n video representations in a single SVC bit stream. Here, let the n^{th} SVC layer represent the video in a quality equivalent to the quality of the corresponding n^{th} AVC video. Then the relative bit rate cost of the n^{th} representation using SVC is

$$C_n = \frac{R_{\text{SVC},1..n} - R_{\text{AVC},n}}{R_{\text{AVC},n}} = \frac{R_{\text{SVC},1..n}}{R_{\text{AVC},n}} - 1. \tag{1}$$

Similarly, the bit rate required for simulcast of the first n AVC representations is $R_{\text{AVC},1..n} = \sum_{i=1}^{n} R_{\text{AVC},i}$. Thus, comparing SVC against AVC simulcast, the relative SVC gain for the n^{th} video representations is

$$G_n = \frac{R_{\text{AVC},1..n} - R_{\text{SVC},1..n}}{R_{\text{AVC},1..n}} = 1 - \frac{R_{\text{SVC},1..n}}{R_{\text{AVC},1..n}}. \tag{2}$$

3.2 Achievable Gains

Eliminating the bit rate required for SVC coding, $R_{\text{SVC},1..n}$, by combing (1) and (2), the SVC gain can be written as a function of the AVC bit rates $R_{\text{AVC},n}$, which specify the operating points for AVC coding for the desired video representations, and the SVC cost C_n. This yields the following equation for the SVC gain.

$$\begin{gathered} G_n = G_n^{\max} - \frac{R_{\text{AVC},n}}{R_{\text{AVC},1..n}} C_n \\ \text{with } G_n^{\max} = 1 - \frac{R_{\text{AVC},n}}{R_{\text{AVC},1..n}} \end{gathered} \tag{3}$$

Given the fact that for practical applications, as of the constraints SVC coding imposes, the SVC bit rate $R_{\text{SVC},1..n}$ is never smaller than the AVC bit rate for coding the highest SVC operating point, $R_{\text{AVC},n}$, the SVC cost can be expected to be non-negative, i.e. $C_n \geq 0$. Thus, considering (3), $G_n^{\max}$ is the maximum achievable relative SVC gain compared to simulcast coding. This gain would be achieved in terms of relative bit rate saving compared to simulcast, if scalable coding of the n^{th} video representation required the same bit rate as non-scalable coding of that representation, i.e. if SVC coding came at zero cost.

The actual SVC gain G_n compared to simulcast decreases with increasing SVC cost C_n, i.e. with increasing SVC bit rate overhead compared to non-scalable AVC coding. This relationship can be nicely illustrated for the simple yet common example of two layers (e.g. spatial scalability 720p to 1080p). For the base layer, obviously,

$G_1 = -C_1$. For the enhancement layer, with $r = R_{\mathrm{AVC},2} / R_{\mathrm{AVC},1}$ the bit rate ratio between the two selected video representations, the following equation is obtained.

$$G_2 = G_2^{\max} - \frac{rC_2}{1+r}, \; G_2^{\max} = \frac{1}{1+r} \tag{4}$$

As can be seen from (4), the maximum achievable gain $G_2^{\max}$ is a function of the bit rate ratio r. The actual SVC gain, G_2, is determined by r and C_2 only. The relationship is illustrated in Fig. 4. Obviously the maximum gain is only achieved for vanishing SVC layer 2 cost $C_2 = 0$, and the actual gain decreases with increasing C_2 and r. Note that while G_2 is a well-defined function of r and C_2, the cost C_2 practically depends on several factors such as the video content, encoder control, and bit rates of base layer and enhancement layer (the latter two are reflected by the bit rate ratio r). This will be demonstrated in Sect. 4.

To conclude the discussion, comparing SVC coding over non-scalable AVC simulcast coding at comparable quality, the maximum theoretical relative SVC gain over AVC simulcast is determined by the following two factors that need to be taken into account when considering the use of SVC in practical deployments.

(a) The operating point as determined through the AVC bit rates $R_{\mathrm{AVC},n}$.

(b) The SVC cost C_n over single layer AVC coding (which depends on the operating point as well).

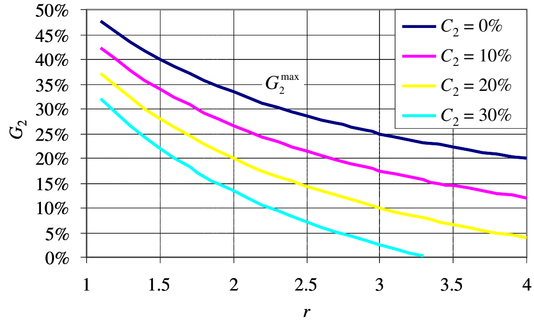

Fig. 4. Two layers SVC gain G_2 over bit rate ratio r

4 Experimental Analysis of SVC Coding Efficiency

Having defined the SVC cost and gain in Sect. 3 along with analysis of the theoretical gains, we now investigate actual cost and gain in practical coding experiments.

4.1 Practical Operating Points

As illustrated in Fig. 4, the maximum theoretical gain with SVC is highly dependent on the bit rate span between the base and enhancement layer bit, or more precisely, the ratio r between the bit rates required to code the base layer and enhancement layer representations using single layer AVC coding.

When the base and enhancement layers have a similar quality, prediction between the layers can be very efficient and thus large gains can be achieved if r is close to 1. On the other hand, in many practical cases, as the value of r decreases and the qualities between the layers become similar, the benefit of transmitting multiple different yet very similar representations decreases – it makes little sense to transmit two representations that are visually indistinguishable. For this reason, assuming that SVC coding will come at some non-zero cost, single layer coding with terminal side adaptation (e.g. scaling the video to the display resolution) will be the most viable solution at some point.

Towards the other end of bit rate ratios, as r increases and thus the difference in the quality of the representations increases, prediction between layers is getting less efficient, and the maximum gain reduces as the original cost of the lower layer becomes less significant to the total bit rate. This can be seen from Fig. 4.

Following the argumentation above, the main benefits of SVC coding can be expected within a certain range of bit rate ratios r not too far from 1.

In this paper we consider the following two practical applications in more detail; Mobile TV at QVGA and VGA resolution and TV at 720p and 1080p resolution. For the Mobile TV scenario, we consider SNR scalability at QVGA with two different bit rates. Here, we consider bit rate ratios between 1.5 and 2 relevant for observing clear differences in visual quality. Additionally, we consider spatio-temporal scalability from QVGA@12.5Hz to VGA@25Hz. Given that the amount of pixels increases by a factor of eight between these spatio-temporal resolutions, we consider bit rate ratios of more than 2 relevant. For the TV scenario, the number of pixels is more than doubled and we consider bit rate ratios of at least 1.5 relevant.

In summary, we consider three different scenarios that are in accordance with the scenarios considered in the SVC verification tests in MPEG [3],

(A) SNR scalability, QVGA@12.5Hz, using scalable baseline profile [1],

(B) spatio-temporal scalability, QVGA@12.5Hz to VGA@25Hz, using scalable baseline profile,

(C) spatial scalability, 720p50 to 1080p50, using scalable high profile [1].

4.2 SVC Cost and Gain in the JSVM Framework

In our first experiment, using the SVC reference software (JSVM 9.17) [4] and coding settings according to [3], we perform SVC encoding using four different base layer quantizer settings for each scenario and sequence, such as to roughly span a reasonable base layer quality range. For each of the base layer quantization parameters QP_{BL}, we use four different values $QP_{offset}=QP_{EL}-QP_{BL}$ (five values for scenario A), thus obtaining four (five) different enhancement layer quantization parameters QP_{EL} and a total of 16 (20) test points per sequence. SVC encoding is done with all inter-layer prediction tools enabled. As a reference, we encode each sequence and each representation with multiple quantization parameters using the non-scalable AVC compatible mode of the JSVM software. Based on the coding results, we calculate the cost and gain measures as well as the bit rate ratios between base layer and enhancement layer as described in Sect. 3 for each test. Then we take the average of the obtained enhancement layer cost, gain and bit rate ratio values over the test points with constant values of QP_{offset} for each sequence. Note that as of the "bottom-up" coding approach used in JSVM [6], the base layer representation is practically identical for the SVC and the AVC single layer case, thus the base layer cost is practically zero, $C_1 \approx 0$.

The results are illustrated in Figs. 5-7. Generally, with increasing QP_{offset}, the enhancement layer quality decreases towards the base layer quality and thus the bit rate ratio r between enhancement layer and base layer decreases. Consequently, as expected from the analysis in Sect. 3, increasing SVC gain can be observed as QP_{offset} increases. For most sequences, increasing SVC enhancement layer cost can be observed as well as QP_{offset} increases. The choice of QP_{offset} has an impact on the difference in pixel fidelity between the base layer and the enhancement layer, with $QP_{offset}>0$ indicating that the enhancement layer quantization is coarser (pixel fidelity lower) than the base layer quantization. As discussed in Sect. 4.1, the qualities of base layer and enhancement layer representations will become similar as the bit rate ratio r decreases (QP_{offset} increases). At some point there will be no justification to transmit two different representations. For SNR scalability (scenario A), a value of $QP_{offset}=-6$ corresponds to a reasonable bit rate ratio between 1.5 and 2. Here, average SVC costs C_2 between 20% and 33%, and average SVC gains G_2 between 15% and 23% are obtained. For QVGA-to-VGA scalability (scenario B), a reasonable value of $QP_{offset}=2$, corresponding to a bit rate ratio of around 2.5, leads to average SVC costs C_2 between 12% and 22% and average SVC gains G_2 between 10% and 21%. For 720p-to-1080p scalability (scenario C), the same value of $QP_{offset}=2$ leads to an average bit rate ratio of around 1.8, and average SVC costs C_2 between 11% and 31% and average SVC gains G_2 between 20% and 23% are observed.

In summary, the results indicate that for practical operating points, SVC gains and SVC costs in the enhancement layer are roughly in the same order of percentage.

4.3 Impact of Encoding Optimization

It is a commonly known fact that encoding optimization can have major impact on the performance of video compression systems, while also having a significant impact on the encoding complexity. Both need to be taken into account when evaluating SVC coding efficiency.

It was shown in [3][6] that the SVC cost can be traded between the base layer and the enhancement layer, and costs in the order of 10% over AVC encoding with JSVM can be achieved for both layers. This was done with an advanced SVC encoder that jointly optimizes SVC base layer and enhancement layer coding instead of applying the "bottom-up" approach used in the JSVM encoder. However, the approach in [6] can be considered to come at a significant increase in encoding complexity over the "bottom-up" approach, and a similar degree of optimization applied for AVC encoding may yield similar improvements. To illustrate the impact of advanced encoding on AVC coding efficiency, we show that significant gains over AVC encoding with JSVM can be achieved already without introduction of any new coding tools. This is done with pure encoding optimization using the same coding tools. To this end, we compare the results for AVC encoding using JSVM with advanced AVC encoding using the publicly available KTA software [5]. We use identical prediction structures and H.264/AVC encoding tools with both software packages, i.e. only H.264/AVC compliant coding tools are used in KTA. Using AVC high profile settings according to [3], we code the test sequences over a range of medium qualities and compute BD-rate differences according to [7]. As summarized in Tab. 1, coding efficiency improvements between 3% and 21% can be achieved, with average gains of 7.5%. This shows that SVC gains with optimized encoding shown over JSVM-based AVC encoding need to be assessed with a perspective that gains of similar orders may be achieved with pure AVC encoding optimization.

Considering the potential effects of encoding optimization on both coding efficiency and complexity, we conclude that for SVC coding efficiency comparisons with respect to practical applications, the encoding complexity needs to be comparable, which we believe is the case for our experiments described in Sect. 4.2.

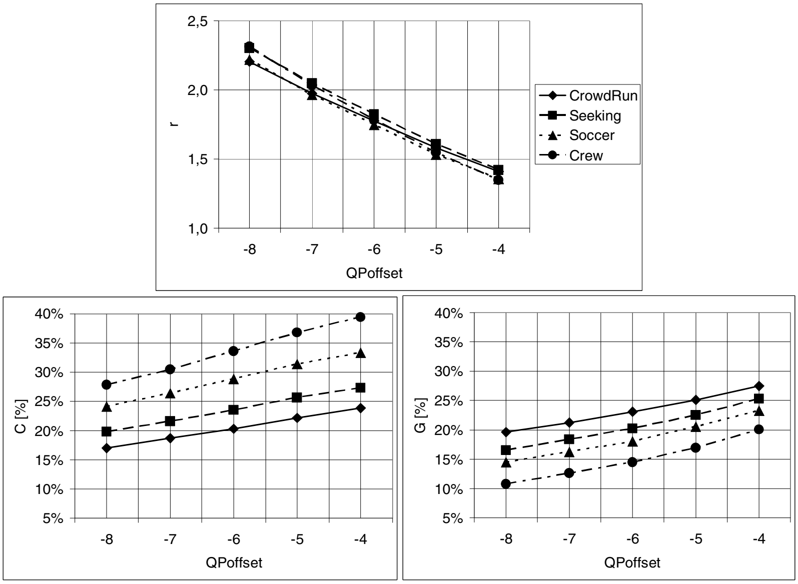

Fig. 5. Bit rate ratio and SVC enhancement layer cost and gain for scenario A (SNR scalability using QVGA@12.5Hz sequences with scalable baseline profile)

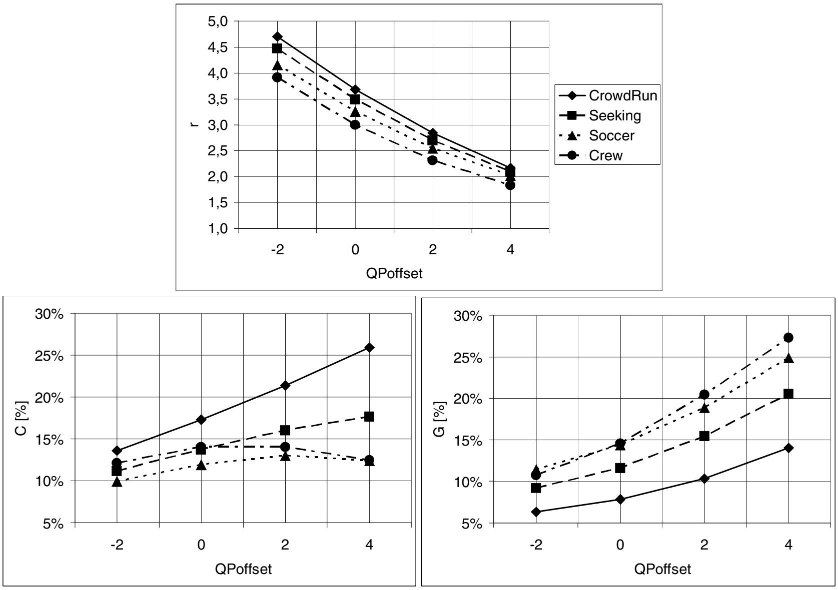

Fig. 6. Bit rate ratio and SVC enhancement layer cost and gain for scenario B (spatio-temporal scalability, QVGA@12.5Hz to VGA@25Hz, using scalable baseline profile)

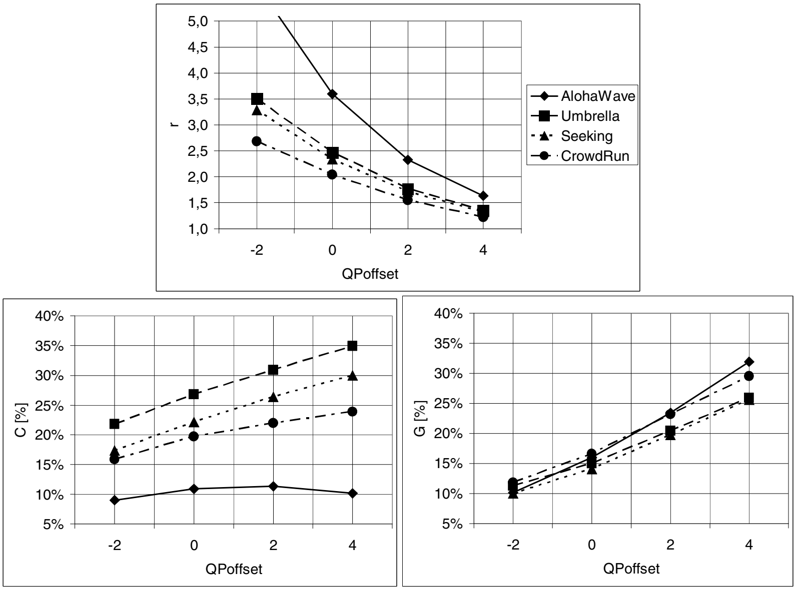

Fig. 7. Bit rate ratio and SVC enhancement layer cost and gain for scenario C (spatial scalability, 720p50 to 1080p50, using scalable high profile)

Table 1. BD-rate improvement for AVC encoding using KTA compared to JSVM

Sequence	**Resolution**	**Frame rate [fps]**	**BD-rate [%]**
CrowdRun	QVGA	12.5	-3,34
Seeking			-4,70
Crew		15	-5,55
Soccer			-3,48
CrowdRun	VGA	25	-3,88
Seeking			-4,13
Crew		30	-9,42
Soccer			-3,69
AlohaWave	720p	50	-13,25
CrowdRun			-5,54
Seeking			-7,74
Umbrella			-7,60
AlohaWave	1080p	50	-21,32
CrowdRun			-6,19
Seeking			-8,77
Umbrella			-11,24
Average			**-7,49**

5 Discussion on Using SVC for Video Delivery

Using the definitions of SVC gain and cost introduced in Sect. 3, and the experimental coding results in Sect. 4, we now discuss the benefits of SVC for the video delivery to heterogeneous terminals as outlined in Sect. 2.

5.1 SVC Cost and Gain in Unicast Delivery

As discussed in Sect. 2.2, if AVC is used to provide different video representations to heterogeneous terminals in a video on demand service, all representations need to be stored individually at the server, the total storage amount being the sum of the amounts required for the individual AVC-coded representations. With reference to Sect. 3, the total bit rate for storage is thus $R_{\text{AVC},1..N}$, while with SVC, the equivalent bit rate is $R_{\text{SVC},1..N}$. Consequently, there is an SVC gain G_N in terms of storage savings over storage of multiple simulcast coded representations. At the same time, the SVC cost C_N reflects the storage overhead induced by provisioning N different representations compared to provisioning only the highest quality representation coded with AVC.

On the transmission side, using AVC, a single non-scalable bit stream tailored to the terminal type is transmitted in each unicast session, thus providing AVC coding efficiency. Using SVC, some or all of the N video representations will require higher transmission bit rates as compared to the corresponding AVC representations, i.e. the SVC cost C_n applies for transmission of the n^{th} video representation.

In summary, the SVC gain (shown to be the order of 10-20% in Sect. 4) applies to the storage compared to an AVC simulcast storage solution. On the other hand the SVC cost (shown to be the order of 10-30% for the enhancement layer) applies to the network. Typically, network resources are more valuable than storage resources.

5.2 SVC Cost and Gain in Broadcast/Multicast Delivery

Following Sect. 2.3, for delivery of different video representations to heterogeneous terminals over broadcast and multicast, assuming AVC is used, if terminal side adaptation is not an option, the different video representations must be transmitted as simulcast in the entire network, except possibly the last hop(s) for multicast delivery, see Fig. 2. Considering Sect. 3 and assuming N different types of terminals, the video bit rate in the core network is thus $R_{\mathrm{AVC},1..N}$ for AVC. With SVC, the bit rate in the core network is $R_{\mathrm{SVC},1..N}$. Thus, there is an SVC gain G_N in the core network. At the same time, the SVC cost C_N indicates the transmission overhead induced by provisioning N different representations compared to transmitting only the highest quality representation with AVC coding.

In the case of broadcast, the SVC gain mentioned above applies in the entire network. In multicast however, this is not the case. On the final hop(s), the network transmits only those parts of the coded video that are actually requested by each particular terminal. In the AVC case, only one single non-scalable bit stream is transmitted on the last hop, requiring a bit rate of $R_{\mathrm{AVC},n}$ for video representation n. Using SVC, the corresponding bit rate is $R_{\mathrm{SVC},1..n}$. Thus, on the final multicast hop(s), the SVC cost C_n applies for transmission of the n^{th} video representation.

In summary, the SVC gain (shown to be the order of 10-20% in Sect. 4) applies to the bandwidth usage in the entire network for broadcast, but only for the core network for multicast. The SVC cost (shown to be in the order of 10-30% for the enhancement layer) represents the bandwidth overhead induced by provisioning multiple different representations instead of only a single representation. In particular, this cost applies to the last hop(s) in multicast systems (e.g. the DSL hop). For video broadcast where multiple representations need to be transmitted, the SVC gain seems promising. In multicast scenarios, the last link may be a valuable and/or very limited resource as is the case for IPTV over DSL and multicast Mobile TV (e.g. over MBMS).

6 Conclusion

SVC enables coding of multiple video representations in a single bit stream, with exploitation of redundancies between the representations. As such, it can provide coding efficiency gain over AVC simulcast. As shown in this paper, the achievable gain depends on the range of video qualities provided in the SVC bit stream, showing a decrease with increasing distance between layers. The actual gain additionally depends on the SVC cost over single layer coding, which depends on video content and encoder optimization.

For the case of two layers, while it has been shown that the cost can be around 10% per layer when advanced encoding is used [6], our results with straightforward bottom-up encoding using JSVM show significantly higher costs in the enhancement layer. The SVC gains obtained in our experiments are in the same order of percentage as the costs. When considering the use of SVC in practical applications, both the SVC gain and cost in terms of coding efficiency need to be taken into account. With that respect, among the applications that we have considered in this paper, it appears that video broadcast to terminals that require different video representations (typically different resolutions), is the most viable SVC application. In any case, provision of multiple different video representations with SVC will come at a significant bit rate cost compared to provision of a single representation.

Further aspects that we believe need to be evaluated in detail when considering the use of SVC include possibilities for video adaptation and transcoding, error robust coding and transmission, error concealment and tune-in delay, as well as complexity aspects such as encoder control and decoder complexity, and transmission complexity. When evaluating SVC spatial scalability, non-scalable coding with advanced image resampling techniques should be considered as a benchmark.

As we have shown in this paper, SVC can theoretically provide promising gains in certain applications. The SVC cost was, however, shown to not be negligible and in certain practical scenarios this cost may outweigh the gains provided by SVC. Careful consideration of both the costs and gains, and where these apply, needs to be made when considering the introduction of SVC in such systems.

References

1. ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services (2007)
2. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the scalable video coding extension of the H. 264/AVC standard. IEEE Trans. Circ. Syst. Video Techn. 17(9), 1099–1102 (2007)
3. SVC verification test report. Doc. MPEG-N9577, ISO/IEC JTC1/SC29/WG11. Antalya, Turkey (2008)
4. Joint Scalable Video Model (JSVM) software, :pserver:jvtuser:jvt.Amd.2@garcon.ient.rwth-aachen.de:/cvs/jvt (module "jsvm")
5. Key Technical Area (KTA) software, `http://iphome.hhi.de/suehring/tml/download/KTA`
6. Schwarz, H., Wiegand, T.: R-D optimized multi-layer encoder control for SVC. In: Proc. IEEE Int. Conf. Image Proc. ICIP 2007, vol. 2, pp. 281–284 (2007)
7. Bjøntegaard, G.: Calculation of average PSNR differences between RD-Curves, ITU-T SG16 Q.6 Document, VCEG-M33. Austin, Texas, USA (2001)
8. Moiron, S., Ghanbari, M., Assuncao, P., Faria, S.: Video Transcoding Techniques. In: Grgic, M., Delac, K., Ghanbari, M. (eds.) Recent Advances in Multimedia Signal Processing and Communications. Springer, Berlin (2009)