Video Network Traffic and Quality Comparison of VP8 and H.264 SVC

Patrick Seeling
Dept. of Computing and New Media Technologies
University of Wisconsin-Stevens Point
Stevens Point, WI 54481
pseeling@ieee.org

Frank H.P. Fitzek
Dept. of Electronic Systems
University of Aalborg
Aalborg, Denmark DK-9220
ff@es.aau.dk

Gergö Ertli
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics
1111 Budapest, Goldmann György square 3, Hungary
egergo@aut.bme.hu

Akshay Pulipaka
School of Electrical, Computer, and Energy Engineering
Arizona State University
Tempe, AZ 85287-5706
akshay.pulipaka@asu.edu

Martin Reisslein
School of Electrical, Computer, and Energy Engineering
Arizona State University
Tempe, AZ 85287-5706
reisslein@asu.edu

ABSTRACT

Google has recently released the video compression format VP8 to the open source community. This new compression format competes against the existing H.264 video standard developed by the ITU-T Video Coding Experts Group (VCEG) in collaboration with the ISO/IEC Moving Picture Experts Group (MPEG). This paper compares these two video coding standards in terms of video bit rate-distortion (quality) performance and the video network traffic variability with different long video sequences. We find that VP8 presently does not fulfill its promise to achieve twice the quality at half the bandwidth compared to H.264. The rate-distortion (RD) performance of VP8 is rather slightly below the RD performance of H.264. On the positive side, in contrast to H.264, VP8 has no license fees.

Categories and Subject Descriptors

I.4.2 [**Computing Methodologies**]: Image Processing and Computer Vision—*Compression (Coding)*; C.2.m [**Computer Systems Organization**]: Computer-Communication Networks—*Miscellaneous*

General Terms

Performance

MoViD'10, October 25, 2010, Firenze, Italy.

Keywords

H.264, rate-distortion performance, traffic variability, VP8, video compression

1. INTRODUCTION

Video coding has been a very active research area over the last decades. The most prominent candidates came from the MPEG group (MPEG2 or MPEG4) or the ITU (H.261/H.263/H.263+). In 2003 a combined effort of the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG) led to the H.264/MPEG-4 AVC standard [5, 8] and subsequent scalability extensions in H.264 SVC [2, 6, 9]. Most of the current products such as Blu-ray, DVB, iTunes, and youTube are based on the H.264 standard.

But early in 2010, Google announced to open source the VP8 video codec they bought from On2 Technologies. The main motivation was to avoid license fees for H.264 based products that will begin in 2011 for Google's products, such as the Chrome browser or youTube. In order to make this video codec widespread and increase adoption possibilities of VP8 as the default HTML5 video standard, Google open-sourced the formerly closed-source developed video codec. Together with the introduction of VP8, claims were made that VP8 would give twice the video quality using half the network bandwidth over currently used video codecs.

Currently the license fees by MPEGLA (`www.mpegla.com`) are under discussion. Even the free usage of H.264 for the HTML5 has been proposed by MPEGLA until the end of 2016. Nevertheless, as given in Table 1 the license fees per unit per year are quite high (maximum yearly fee is limited to 5 million USD). In addition, there is a fee for video streaming service per streaming title and per subscriber. These high fees are a huge barrier for free application and services, such as mozilla, youTube, metacafe, and leave the door wide open for VP8. While there is currently an advantage of VP8 over H.264 with respect to license fees, the

Table 1: License fees for H.264.

range	price per unit
100.000 – 4.999.999 units per year	20 US cents
5.000.000 or more units per year	10 US cents

future of the developed video standard in this respect is not entirely clear. Early claims from different parties suggest that there might be a future patent pool formed by other parties, effectively removing the current licensing advantage of VP8 in the future, similar to the developments with Microsoft's video codec VC-1.

This paper investigates the video quality and network bandwidth usage of VP8 compared to H.264 for three different long feature video sequences. Even though the number of video sequence is small, the tendencies observed from all three video sequences are the same. Currently, the only comparisons of these two video coding standards are done by visual comparison of short video sequences or individual frames. As the visual quality of both standards is quite high and subjective to the observer, we utilize the average PSNR values for the complete video sequences as objective video quality metric.

The remainder of this paper is structured as follows. In Section 2, we introduce the metrics used for the performance evaluation of VP8 and H.264. In Section 3, we present a performance evaluation of both video codecs with respect to rate-distortion performance and network traffic characteristics before we conclude in Section 4.

2. NETWORK TRAFFIC AND VIDEO QUALITY METRICS

Here we provide a brief overview of the most essential video traffic metrics used. For a video sequence consisting of N frames encoded with a given encoding layout, we let X_n $(n = 1, \ldots, N)$ denote the sizes (in byte) of the encoded video frames. We use a byte-wise approach, as the video data produced is typically byte-aligned. The mean frame size $\bar{X}$ [byte] of the encoded video sequence is then defined as

$$\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n, \tag{1}$$

while the variance S_X^2 of the frame sizes (whereby S_X denotes the standard deviation [byte]) is defined as

$$S_X^2 = \frac{1}{(N-1)} \sum_{n=1}^{N} (X_n - \bar{X})^2. \tag{2}$$

The coefficient of variation of frame sizes [unit free] is defined as

$$CoV_X = \frac{S_X}{\bar{X}} \tag{3}$$

and is widely employed as a measure of the variability of the frame sizes, i.e., the bit rate variability of the encoded video. Plotting the CoV as a function of the quantization scale (or equivalently, the PSNR video quality) gives the rate variability-distortion (VD) curve [7].

We define a *Group of Pictures* (GoP) of an encoded video stream as one I frame and all subsequent P frames before the next I frame in the stream, totalling M frames per GoP. The size of GoP m is denoted by Y_m $(m = 1, \ldots, \lfloor M/N \rfloor)$ [byte].

We use the Peak Signal-to-Noise Ratio (PSNR) as an objective measure of the quality of a reconstructed video frame $R(x, y)$ with respect to the uncompressed video frame $F(x, y)$. The larger the difference between $R(x, y)$ and $F(x, y)$, or equivalently, the lower the quality of $R(x, y)$, the lower the PSNR value. The PSNR is expressed in decibels [dB] to accommodate the logarithmic sensitivity of the human visual system. We note that the PSNR does not completely capture the many facets of video quality. However, analyzing a large number of videos subjectively becomes impractical. Moreover, recent studies have found that the PSNR is as good a measure of video quality as other, more sophisticated objective quality metrics [1]. We focus on the luminance PSNR values, since the human visual system is more sensitive to small changes in the luminance.

For the typical case of $N_x \times N_y$ frame consisting of 8-bit pixel values, it is computed as a function of the mean squared error (MSE) as

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{MSE}. \tag{4}$$

We denote the PSNR quality of a video frame n by Q_n and define the average PSNR quality $\bar{Q}$ of a video sequence as

$$\bar{Q} = \frac{1}{N} \sum_{n=1}^{N} Q_n. \tag{5}$$

3. PERFORMANCE COMPARISON

For the performance comparison, we utilized the *ffmpeg* version of the initial VP8 release and the H.264 SVC reference software encoder (ver. 9.19.9). For both encoders, we furthermore used only the basic encoding settings without additional optimizations. A GoP length of 16 frames was used with only I and P frames, resulting in a GoP pattern of *IPPPPPPPPPPPPPPP* for both encoders. We used the same quantization scale settings for I and P frames in both encoder settings, namely 10, 16, 22, 24, 28, 34, 38, 42, and 48. All encoded video sequences were in the CIF (352×288) resolution.

We illustrate the different produced video bitrates for VP8 and H.264 SVC with comparable average video sequence PSNR values as in Equation 5 (approx. difference of 0.3 dB) on a two-GoP level (32 frames) for the *Sony Demo* video sequence in Figure 1. We observe that overall, the two different encodings follow the same trends in terms of the resulting bitrate, but with H.264 SVC being significantly lower for most of the encoded video. Furthermore, we note that the traffic variability (in terms of the bitrate) seems to oscillate more for VP8.

In Figure 2, we illustrate the video bitrates on a two-GoP level for VP8 and H.264 SVC encodings of the *Terminator* video sequence with both encodings at a quantization scale setting of 28 and H.264 SVC having an approximately 51 kbps higher bitrate overall and a 1.66 dB higher average PSNR. We observe that both encodings exhibit a fairly similar behavior of the encoded video data bitrate over time. We note, however, that the H.264 SVC encoded video has overall slightly more pronounced 'peaks' of the encoded traffic, which attributes to the higher overall bitrate of the encoding, but also has periods where the traffic produced is slightly lower than for VP8. The differences in the encoded

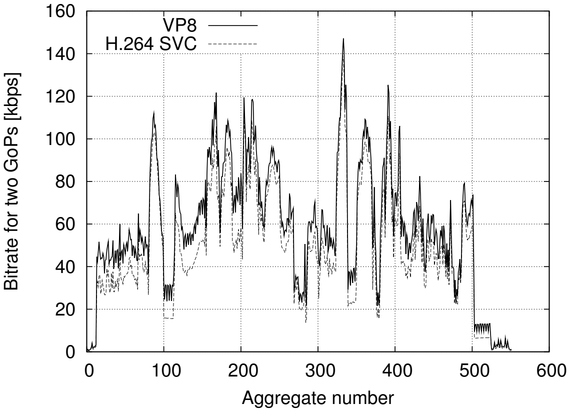

Figure 1: GoP (32 frames) video bitrate trace for comparable video qualities of the *Sony Demo* video sequence encoded in VP8 with quantization scale factor 10 and H.264 SVC encoded with quantization scale factor 22.

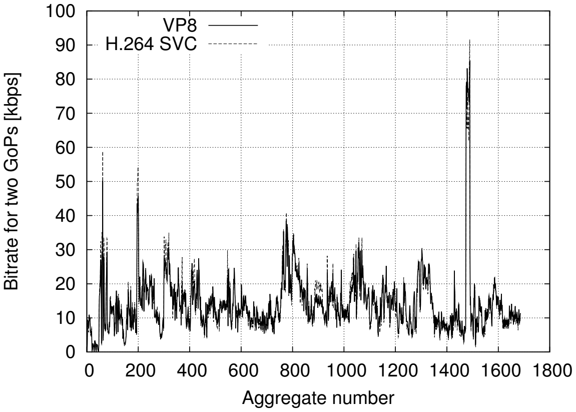

Figure 2: GoP (32 frames) video bitrate trace for comparable video qualities of the *Terminator* video sequence encoded in both VP8 and H.264 SVC with quantization scale factor 22.

bitrates we observe here in turn must be attributed to the individual encoder's efficiency to encode different types of content, e.g., high or low motion.

3.1 Rate-Distortion

Initial performance evaluation for video codecs typically investigates the rate-distortion performance, i.e., the video quality (distortion) as a function of the bitrate. We illustrate the rate-distortion (RD) curves for the three evaluated video sequences *Sony Demo*, *Die Hard*, and *Terminator* in Figures 3, 4, and 5, respectively. We observe for all encodings the typical logarithmic increase of the PSNR quality as the bitrate increases. For the *Sony Demo* video sequence, we observe the smallest difference in PSNR values for similar bitrate values, with the average PSNR value produced by H.264 SVC outperforming the one obtained for VP8 by approximately 1–2 dB. For the *Die Hard* and *Terminator* video sequences, this spread is larger at approximately 3–4 dB.

Overall, we note that the quality and bandwidth ranges

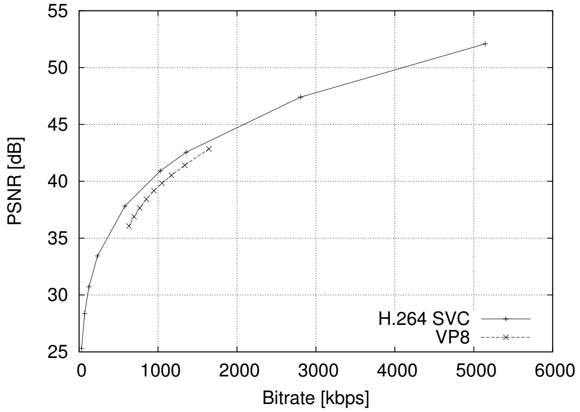

Figure 3: Rate-Distortion (RD) for the *Sony Demo* video sequence.

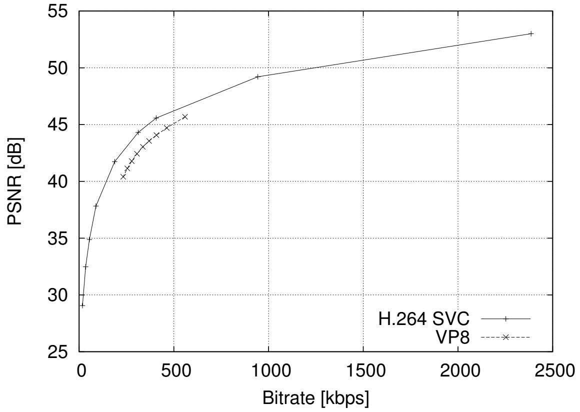

Figure 4: Rate-Distortion (RD) for the *Die Hard* video sequence.

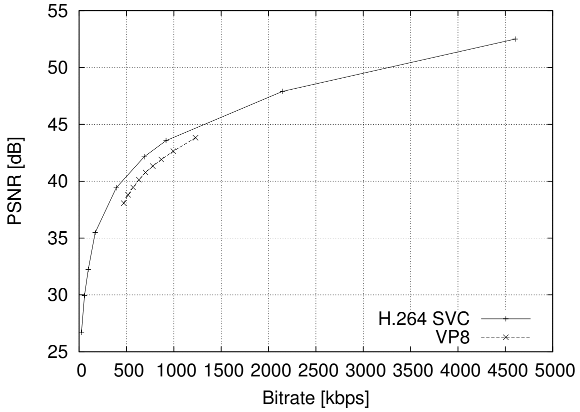

Figure 5: Rate-Distortion (RD) for the *Terminator* video sequence.

produced by VP8 are significantly smaller than those produced by H.264 SVC and content-dependent. Additionally, we note that the performance in terms of average PSNR values obtained for our evaluated long video sequences for a given bitrate are overall significantly smaller for the VP8 encoded videos.

3.2 Network Traffic Variability

For transmission over networks, the variability of the network traffic produced by video encoders is of high importance, as bandwidth provisioning needs to take not just the average video bitrate produced into account, but also the deviation from that bitrate [3, 4]. While oftentimes the peak-to-mean ratio of the produced traffic is considered, this metric is prone to individual outliers. The coefficient of variation, on the other hand, provides a more robust and comparable metric for comparison of the video traffic variability. In addition, the CoV is normalized with respect to the average bitrate (or video frame sizes). We illustrate the rate variability-distortion (VD) curve for the three evaluated video sequences *Sony Demo*, *Die Hard*, and *Terminator* in Figures 6, 7, and 8, respectively. We initially observe that overall, the variability for both video encoding standards decreases as the average sequence quality (measured in terms of the PSNR) increases. For all video sequences except *Sony Demo*, we observe a higher level of variability for VP8 compared to H.264 SVC. We also note that the behavior of the decrease in variability is mostly linear for VP8 while the H.264 SVC encoding exhibits a steady high level of variability for low quality levels and afterwards decreasing levels of variability as the quality increases.

Overall, the traffic variability of the encoded video produced by VP8 is significantly higher than the comparable H.264 SVC video traffic variability, except for the *Sony Demo* video sequence, which might be due to content dependency of the encoded data. For the streaming of encoded video, this overall favors H.264 SVC encoded video, as a lower traffic variability is more favorable from a network point of view.

3.3 Network Traffic Self-Similarity

The Hurst parameter H is commonly employed to measure the self-similarity of an underlying stochastic process. While mathematically well-defined, the parameter itself is typically *estimated* using different approaches and ranges from 0.5, indicating no presence of long-range dependence, to 1, indicating that the underlying process exhibits long-range dependence. We provide the Hurst parameter H, which we obtained from a least-squares fit of the R/S statistic pox plot for the individual video frame sizes X_n in Table 2. We initially note that for nearly all different encoders and settings, the Hurst parameter is larger than 0.8, indicating the presence of long-rage dependence of the video frame sizes over time. We observe furthermore that the Hurst parameter on a frame level typically starts closer to 1 with a low quantization scale (high quality) before dropping as the quantization scale is increased (lower quality), independent of the employed video encoder. This behavior can be explained with the loss in accuracy of the encoded content, which results in frame sizes that are less content-related and hence exhibit less similarity with preceding or subsequent frames.

For the GoP level (aggregate of 16 frames constituting a Group of Pictures), we observe a less consistent trend,

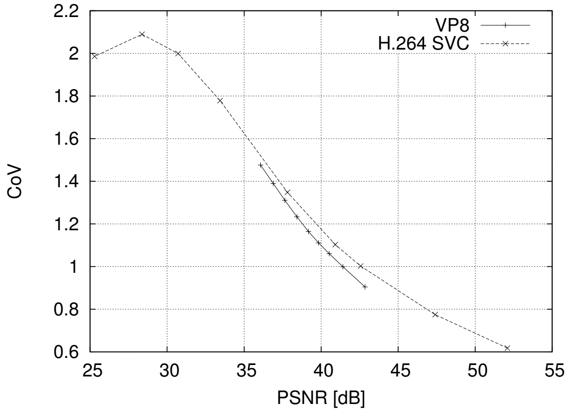

Figure 6: Rate Variability-Distortion (VD)for the *Sony Demo* video sequence.

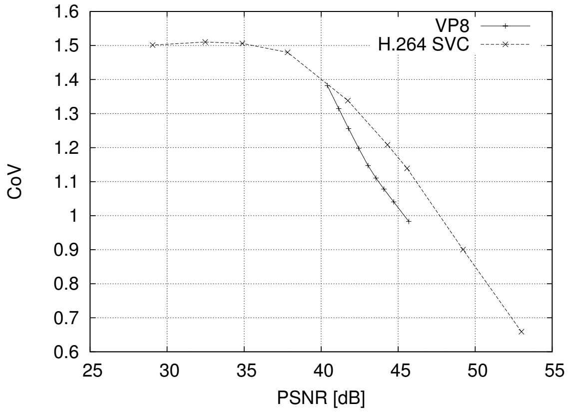

Figure 7: Rate Variability-Distortion (VD) for the *Die Hard* video sequence.

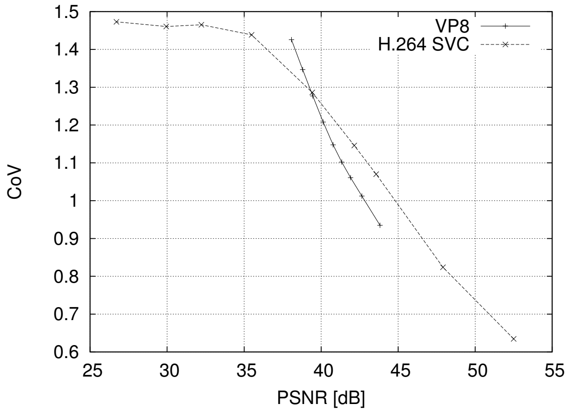

Figure 8: Rate Variability-Distortion (VD) for the *Terminator* video sequence.

Table 2: Overview of the Hurst parameter H estimated from a least-squares curve fitting of the R/S statistic pox plot of the video frame sizes for the video sequences *Sony Demo*, *Die Hard* and *Terminator*.

Quant.	Sony Demo				Die Hard				Terminator			
Scale	Frame		GoP		Frame		GoP		Frame		GoP	
	VP8	H.264	VP8	H.264	VP8	H.264	VP8	H.264	VP8	H.264	VP8	H.264
10	0.94	0.99	0.94	0.86	0.9	0.90	0.83	0.83	0.93	0.91	0.89	0.87
16	0.91	0.98	0.95	0.89	0.88	0.9	0.84	0.82	0.92	0.93	0.89	0.87
22	0.90	0.95	0.95	0.91	0.88	0.88	0.84	0.83	0.91	0.92	0.89	0.88
24	0.89	0.93	0.94	0.9	0.88	0.88	0.84	0.83	0.91	0.91	0.89	0.88
28	0.88	0.89	0.94	0.9	0.87	0.87	0.84	0.84	0.90	0.89	0.9	0.88
34	0.86	0.81	0.95	0.9	0.86	0.85	0.84	0.84	0.89	0.87	0.9	0.87
38	0.83	0.79	0.95	0.9	0.85	0.84	0.85	0.84	0.88	0.86	0.9	0.87
42	0.81	0.77	0.95	0.91	0.85	0.84	0.85	0.84	0.87	0.85	0.9	0.87
48	0.8	0.79	0.96	0.9	0.85	0.83	0.86	0.84	0.86	0.84	0.9	0.87

whereby the overall Hurst parameters determined for both encoders remains fairly level over the range of the evaluated quantization scales. As aggregation takes place for the GoP level, the overall motion and content similarities for a GoP of the underlying video sequences are more responsible for the sizes on a GoP level. Comparing the VP8 results for the frame level with those obtained for H.264 SVC, we observe that the Hurst parameters of VP8 encodings seem slightly lower than the ones obtained for H.264 encodings for low quantization scale settings. As the quantization scales increase, however, the Hurst parameters estimated for the H.264 encodings become slightly lower than their VP8 counterparts, i.e., they drop faster. To explain this behavior, we need to take into account that the obtained qualities for VP8 encodings are in a smaller range than those obtained for the H.264 encodings, as illustrated in the rate-distortion plots in Figures 3, 4 and 5.

If we compare the bitrates for the *Sony Demo* sequence and the resulting Hurst parameters, we observe that, e.g., quantization scale factor 16 for H.264 is closest to quantization scale factor 42 for VP8. In this comparison, the self-similarity for H.264 encoded video at a similar bitrate is significantly higher than for VP8. If we compare for similar video qualities, we again note that a quantization scale factor of 28 for H.264 is close to 38 for VP8 and results in a higher self-similarity of the video frame sizes.

4. CONCLUSION AND OUTLOOK

In this paper we compared the newly introduced video codec VP8 with H.264 in terms of bandwidth and video quality. While VP8 has the advantage of currently being license free, H.264 seems to offer lower bandwidth usage and better video quality. In stark contrast to the announcement that VP8 would use half the bandwidth offering twice the video quality compared to H.264, our findings show that H.264 outperforms VP8 in terms of rate-distortion performance by 1–4 dB for a given bit rate.

Currently, we investigate additional video sequences and will make results publicly available on our main video trace web page while taking future developments into account. Another focus will be to investigate further parameters, e.g., CPU usage and energy consumption, of those two video standards on a mobile device. Besides the bandwidth usage and the video quality, the energy consumption on a mobile platform is very important to become a wide spread standard.

5. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation through Grant No. CRI-0750927.

This work was partially financed by the CONE project granted by Danish Ministry of Science, Technology and Innovation.

6. REFERENCES

[1] Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II, Aug. 2003. available from `http://www.vqeg.org`.

[2] 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*, 54(3):698–718, Sept. 2008.

[3] G. Van der Auwera and M. Reisslein. Implications of smoothing on statistical multiplexing of H.264/AVC and SVC video streams. *IEEE Transactions on Broadcasting*, 55(3):541–558, Sept. 2009.

[4] T. V. Lakshman, A. Ortega, and A. R. Reibman. VBR video: Tradeoffs and potentials. *Proceedings of the IEEE*, 86(5):952–973, May 1998.

[5] D. Marpe, T. Wiegand, and G. Sullivan. The H.264/MPEG–4 advanced video coding standard and its applications. *IEEE Communications Magazine*, 44(8):134–143, Aug. 2006.

[6] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, Sept. 2007.

[7] P. Seeling and M. Reisslein. The rate variability-distortion (VD) curve of encoded video and its impact on statistical multiplexing. *IEEE Transactions on Broadcasting*, 51(4):473–492, Dec. 2005.

[8] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576, July 2003.

[9] M. Wien, H. Schwarz, and T. Oelbaum. Performance analysis of SVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1194–1203, Sept. 2007.