Rate Distortion Performance of H.264/SVC in Full HD with Constant Frame Rate and High Granularity

Martin Slanina, Michal Ries
Brno University of Technology
Technicka 12
Brno, Czech Republic
e-mail: slaninam@feec.vutbr.cz, ries@feec.vutbr.cz

Janne Vehkaperä
VTT Research Centre of Finland
Kaitoväylä 1
Oulu, Finland
e-mail: janne.vehkapera@vtt.fi

***Abstract*—In this paper, we provide an analysis of performance of the scalable extension of the H.264/AVC video codec. We assume a fixed display scenario at full HD resolution with a constant frame rate. The encoded bit stream consists of one Coarse Grain Scalability (CGS) and one Medium Grain Scalability (MGS) layer with three sublayers, allowing for creating four quality layers from the complete bit stream. Hierarchical coding is employed, which results in dyadic decomposition of temporal layers. In order to increase the granularity of quality layers, the packets are selectively dropped from appropriate quality and temporal layers. We provide a performance analysis of such approach, compare the rate distortion performance to a mainstream H.264 encoder and analyze the composition of the bit stream at the considered operation points. Our findings show that quality enhancement of temporal layers has different effect on the overall performance depending on which temporal layer is enhanced.**

Keywords-High definition video; quality of service; scalability; video coding.

I. Introduction

It has been pronounced in the recent years that the video traffic forms a significant share in the data carried over the Internet. In 2011, the share of Internet video was 51 % of all consumer Internet traffic and is expected to raise to 55 % in 2016 [1]. These high numbers do not include video exchanged through peer to peer sharing. Counting them in, the share of video on the overall Internet traffic is expected to reach 86 % in 2016.

There is a rapid increase in the number of broadband connections to the Internet. Quite naturally, as the throughput increases, the quality demands of the users increase as well. Advanced Internet video (stereoscopic 3D and high definition) is getting an important share on the overall Internet video traffic and is expected to reach 46 % of the consumer Internet video traffic in 2014 [2]. These numbers clearly show the importance of efficient online delivery of high quality video.

The bit rate of the video carried over the data network is one of the most important aspects related to quality. When the bit rate is too low, the visual quality of the decoded video images may be degraded. On the other hand, when the bit rate is too high, the risk of network having issues delivering the high amount of data increases, which may lead to video freezes or longer waiting times (usually with TCP-based connections) or loss of data and visual impairments (for UDP-based connections). The optimum bit rate is thus limited from both ends and depends, to a large extent, on the individual parameters of the user's connection.

In order to cope with the different bit rate requirements of the delivered video, the most common approach is to offer several copies of the same content, encoded at different bit rates, and let the user or the video player select one of the bit rates according to the available throughput. Such fixed bit rate approach is, however, inefficient in case the available network throughput is changing over the playback. One possible solution is to allow for switching among the different bit rates offered by the provider, implemented in e.g. HTTP Live Streaming technique [3] or DASH - Dynamic Adaptive Streaming over HTTP [4]. A common feature of the mentioned adaptive streaming techniques is that it is the *client* who decides on the bit rate to receive. For such decision, the client needs to be informed on the actual available bit rate and take fast action when the available bit rate decreases. The advantage of the Scalable Video Coding (SVC) [5] extension of the H.264/AVC standard is that parts of the transmitted data stream have different importance. Through dropping certain packets in case of insufficient transmission capacity, SVC offers graceful degradation by decreasing the quality of the decoded received stream. A nice description of the concepts used in SVC can be found in [5].

The aim of this paper is to present a scheme for dropping SVC packets with high bit rate/quality granularity and to compare its performance to the non-scalable H.264/AVC encoder using multiple streams. We focus on the full HD resolution and full frame rate of 25 frames per seconds, as such format is very promising for the current and future video delivery.

Throughout this paper, the rate distortion performance will be measured using the Peak Signal to Noise Ratio (PSNR). Even though PSNR is not a good quality metric in most cases as the correlation of its outputs with results of subjective experiments is poor, it is suitable in this setup as a performance indicator: as it has traditionally been used for quantifying the performance of coding tools and algorithms, it is extremely wide spread and well understood in the video

 ISBN: 978-1-61208-262-2

coding community.

This paper is organized as follows: First, the related work is mentioned in Section II. The used packet dropping scheme and the resulting SVC performance are described in Section III. Finally, the paper concludes in Section IV.

II. Related Work

The performance of the SVC encoding algorithm has been studied even before the standard was released in 2007. In [6], the authors find that the overall rate-distortion performance of the SVC is approximately 10 % worse compared to AVC with identical settings. In other words, SVC needs approximately 10 % more bit rate to achieve the same quality in terms of PSNR. In [6], one base layer and one enhancement layer in spatial or quality domain is used. SVC performance analysis based on subjective tests has been done in [7], confirming the previous results. In [6] and [7], one base layer and one enhancement layer are used in terms of quality and spatial scalability.

A common approach to selecting the layers in the scalable bit stream is selecting a certain combination of D (dependency), T (temporal) and Q (quality) parameters. Then, all the inferior layers are included in all three domains. For fixed display conditions, i.e. fixed frame rate and fixed spatial resolutions, one only gets as many bit rate/quality levels as there are Q levels in the original stream. We utilize dropping in different T layers in order to increase the bit rate/quality granularity.

The authors of [10] provide an analysis of different approaches to dropping layers in the SVC bit stream in the mobile environment. Through a subjective study, different impact on Quality of Experience is introduced by different scaling approaches and for different contents. Generally, spatial scaling is regarded worse compared to temporal and quality scaling, which leads to a recommendation that the quality layers and some temporal layers should be dropped first. This is in complete agreement with our approach, where no spatial scalability is included.

In [11] and [12], the authors analyze the impact of unstable transmissions on the perceived quality when a scalable video bit stream is transmitted. In this context, the unstable transmission leads to varying number of layers received in the sequence duration. It has been shown that quality of unstable videos is subjectively perceived close to the quality of stable videos. Furthermore, it was found that temporal scalability introduces severe degradation of perceived quality while quality scalability leads to best results. These results are the motivation to keep the video frame rate constant through all operation points in our experiment.

III. SVC Performance

A. Packet Dropping Scheme

This subsection presents the packet dropping scheme employed for the increased bit rate granularity. Let us first describe the structure of the full quality bit stream as shown in Fig. 1. Using the hierarchical B frames, the full frame rate of 25 fps can be decomposed in a dyadic structure down to 3.125 fps in case four temporal layers are used. In the lowest frame rate (T = 0), only the 1st and 9th frame are encoded. For each of them, several quality layers can be defined (Q = 1 to Q = 3). The frames at the lowest frame rate serve as the basis for prediction of frames in the higher frame rate, i.e. T = 1. At T = 1, the 1st, 5th and 9th frame are available. Continuing such prediction leads to doubling the frame rate at each step, resulting in 25 fps at T = 3.

TABLE I. OPERATION POINTS.

T:	T = 0					T = 1					T = 2					T = 3					T = 4				
Q:	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
OP1	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠
OP2	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□
OP3	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□
OP4	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□
OP5	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□
OP6	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	⊠	□	□	□	□
OP7	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	⊠	□	□	□	□
OP8	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	⊠	□	□	□	□
OP9	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	□	□	□	□
OP10	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	⊠	⊠	⊠	⊠
OP11	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	⊠	⊠	⊠	□
OP12	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	⊠	⊠	□	□
OP13	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	⊠	□	□	□
OP14	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	⊠	□	□	□	□	⊠	□	□	□	□
OP15	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP16	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP17	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP18	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP19	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP20	⊠	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP21	⊠	⊠	⊠	⊠	⊠	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP22	⊠	⊠	⊠	⊠	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP23	⊠	⊠	⊠	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP24	⊠	⊠	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP25	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□
OP26	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	⊠	⊠	⊠	⊠
OP27	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	⊠	⊠	⊠	□
OP28	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	⊠	⊠	□	□
OP29	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	⊠	□	□	□
OP30	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□	⊠	□	□	□	□

The common approach to dropping SVC data is illustrated in Fig. 2. In order to keep the highest frame rate (T = 3), all the inferior frame rates are kept in full quality (Q = 3). Obviously, this technique can reach only 4 operation points at the full temporal resolution.

In contrast, the approach that we use throughout this paper allows dropping quality layers even from lower temporal resolutions. As long as the base quality layer is kept in each temporal layer, the full frame rate can easily be reconstructed. Such approach is shown in Fig. 3. Note the obvious inconsistency in layer T = 1: Although in the inferior layer (T = 0) no quality enhancement data was kept, higher quality layers are present at T = 1. As there are no references used to the *specific* Q at the lower layer, prediction can still be done and enhanced at the current layer. Moreover, we will show that in some cases this scenario can lead to higher

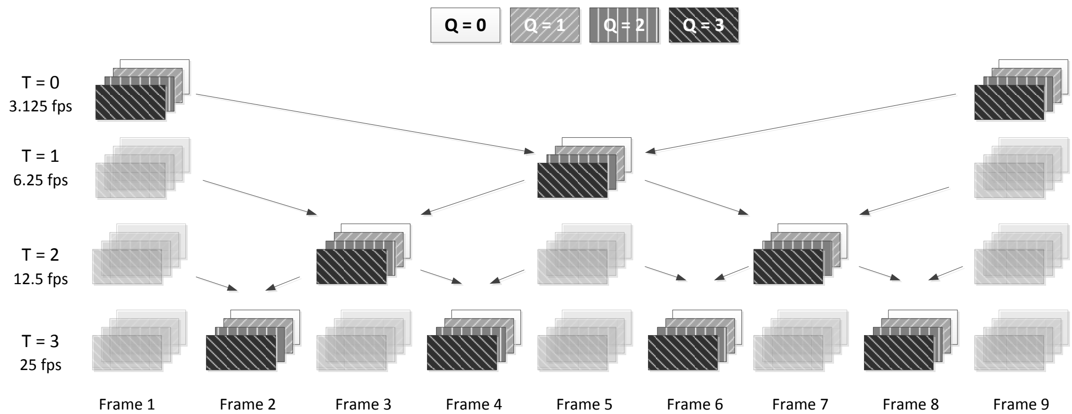


Figure 1. Layers in the SVC bit stream in full quality and full temporal resolution.

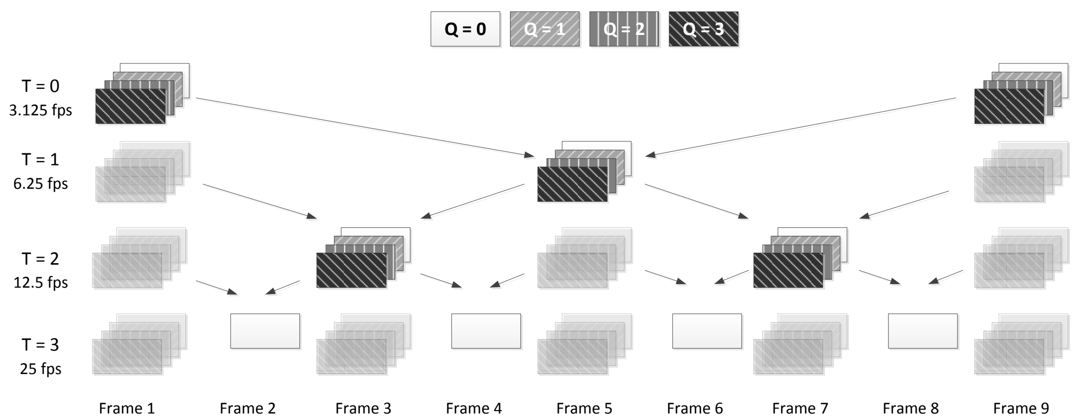


Figure 2. Layers in the SVC bit stream in case the highest temporal resolution is present with only the base quality layer.

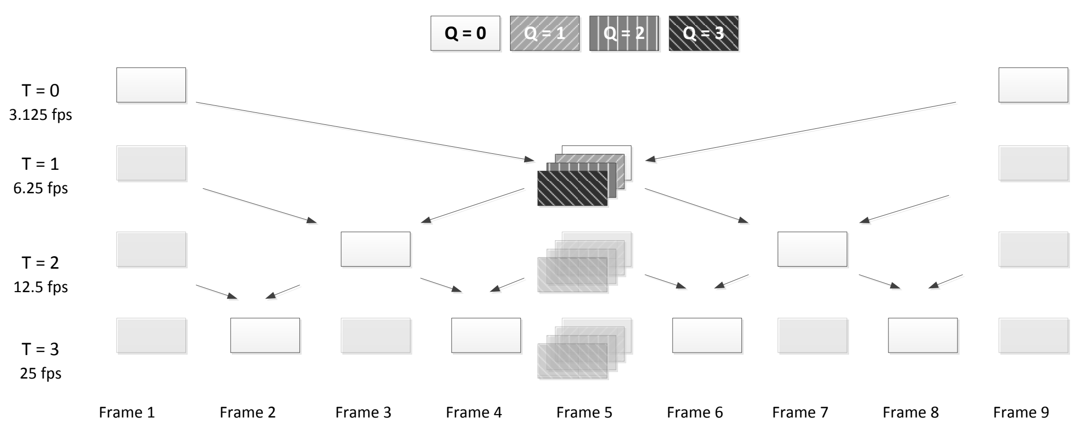


Figure 3. Layers in the SVC bit stream in case only the base quality layer is kept for several temporal layers.

 ISBN: 978-1-61208-262-2

quality at a given bit rate. All the operation points considered in this experiment are described by TABLE I. An operation point (OP) is defined by the subset of layers, employed in the respective sub-stream. Please note that OP25 and OP30 are made up from an identical set of sub-layers, which means one of them could be removed as redundant. However, we keep both of these operation points as they represent the lowest quality reached through different packet dropping strategies, which is useful in further descriptions and plots. In order to drop packets from an SVC bit stream, trace files were prepared and used with the bit stream extractor tool available with JSVM. The resulting streams were decoded with JSVM and compared to uncompressed originals in terms of mean PSNR of the luma component.

B. Test Setup

Two sequences were used in our experiment: The ”blue sky” sequence, which has midrange spatial and low temporal activity (a view of a sky through treetops with slow camera rotation) and the ”tractor” sequence, which has high spatial and temporal activity (a tractor working on the field with complex background and high motion). The sequences have been downloaded from [13].

The sequences were encoded with the JSVM encoder (Joint Scalable Video Model) version 9.19.13 [8]. As this version has no rate control mechanism implemented, we decided for a Quantization Parameter (QP)-based configuration and all the sequences were encoded with the same initial QPs. According to the complexity of each scene, the resulting bit rates and qualities differ.

The bit stream was formed by the component layers:

1) *The base layer,* identified by Q = 0. This layer was coded with the GOP size of 16 to allow efficient decomposition of frame rate. The IDR period was 32. The layer was coded using High profile, initial QP set to 48.
2) *CGS enhancement layer*, identified by Q = 1. The Coarse Grain Scalability (CGS) allows for enhancing the quality of a decoded picture by employing inter-layer prediction mechanisms, such as prediction of macroblock modes, motion parameters and residual prediction. [9]. The layer was encoded with initial QP set to 42.
3) *MGS enhancement layer*, identified by Q = 2 to 4. The Medium Grain Scalability (MGS) provides a finer granularity compared to CGS by splitting a given quality enhancement layer into several MGS layers. Basically, MGS divides the transform coefficients of each macroblock into multiple groups. In our configuration, we used 4 (most significant) coefficients for the lowest quality layer, 4 coefficients for the medium and 8 (least significant) coefficients for the highest quality MGS layer. The initial QP of the MGS enhancement layer was set to 30.

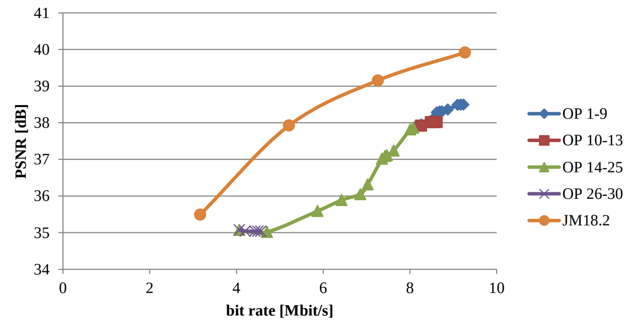


Figure 4. Bit rate vs. quality (PSNR) for the blue_sky sequence.

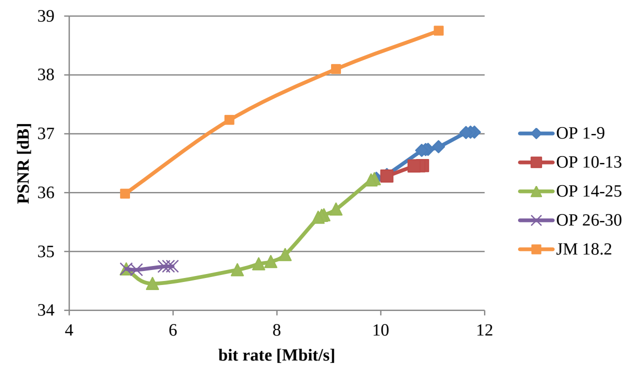


Figure 5. Bit rate vs. quality (PSNR) for the tractor sequence.

Furthermore, the sequences were encoded using the reference implementation of the H.264/AVC encoder [14] version JM 18.2. The motivation for using another encoder is to keep track of the overall performance of the scalable coding and the usability of the created bit streams. The AVC encoder was set to High profile, level 4.0, with hierarchical B-frames having the same structure as the B-frames in SVC.

C. Results

Figures 4 and 5 illustrate the average PSNR for 200 frames of each sequence along all the SVC operation points defined in TABLE I. Furthermore, a curve illustrating the AVC performance is shown. It can be observed that the coding performance of AVC is several dB better compared to SVC. On the other hand, the SVC offers flexibility in modifying a readily encoded bit stream such that the required bit rate can be easily altered with no need of transcoding.

An interesting phenomenon can be observed in the area of lower SVC bit rates for the ”tractor” sequence. Here, different qualities are achieved for the curve representing operation points 14-24, where the quality layers are dropped regularly from higher to lower temporal layers, and the curve representing operation points 26-29, where only the quality

 ISBN: 978-1-61208-262-2

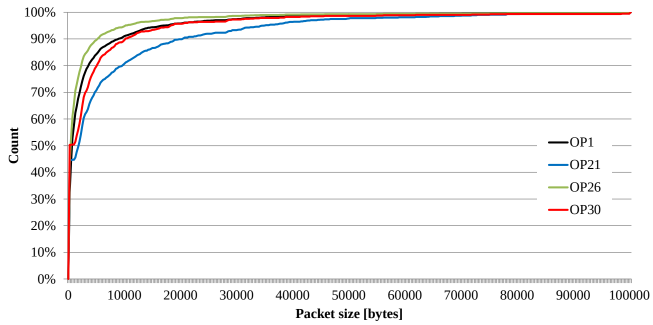


Figure 6. Cumulative distribution function for the blue_sky sequence.

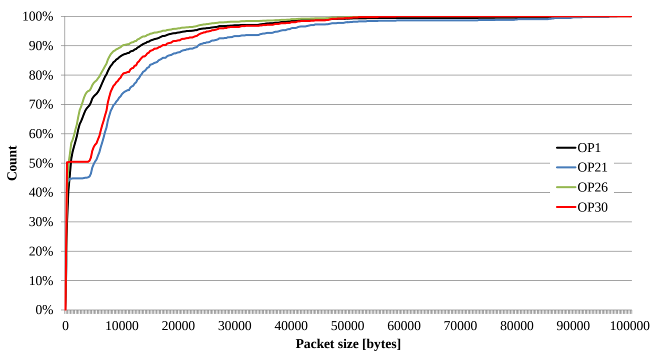


Figure 7. Cumulative distribution function for the tractor sequence.

enhancement data of the highest temporal layer are kept.

In OPs 26-29, the frames from the highest temporal layer are enhanced, which means quality enhancement information is present in every second frame of the sequence (see T = 3 in Fig. 1) - the number of frames in this temporal layer is higher than the number of frames in the inferior temporal layers. One would expect that high bit rate is required to present the frames of highest temporal layer at a higher quality (more quality layers). This might not be true in all cases since the frames at the highest temporal layers quite often do need lower bit rate than the frames at lower temporal layers (due to temporal prediction).

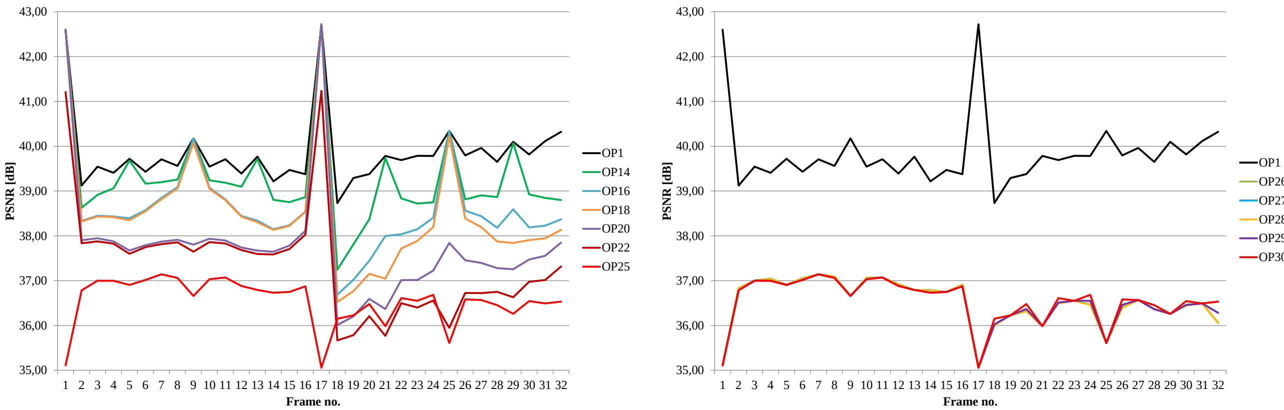
In order to understand the composition of the packets in the bit stream, we have plotted the cumulative distribution function of the packet sizes for both sequences in Fig. 6 and Fig. 7. In these CDFs, the operation point OP1 (full bit stream) is displayed using a black line, OP30 (lowest bit rate stream) is displayed with a red line, OP21 is drawn with a blue line and, finally, OP26 is drawn with a green line. The reason for selecting OP21 and 26 is that they follow different trends in Fig. 5 and use different parts of data for quality enhancement: While OP21 drops quality layers monotonously from highest to lowest temporal layers, OP26 uses quality enhancement in the highest temporal layer only. The cumulative distribution function plots show that the curve for OP 21 is generally below the other curves, which means that there is a low percentage of smaller packets in the bit stream. On the other hand, the curve for OP26 is clearly above all other curves, which means that the percentage of small packets in the bit stream is higher. An explanation for this is that in OP26, the quality enhancement information is present for a larger number of frames (higher temporal layer) using a lower amount of data in small packets.

Still, the explanation for the increased PSNR in higher operation points in Fig. 5 is missing. For a better understanding, we have plotted the PSNR values in time for two GOPs in Fig. 8 and Fig. 9. In all plots, OP1 is drawn with black line, OP30 (equal to OP25) is drawn with red line. First, let us focus on Fig. 8 showing the ”blue_sky” sequence: Fig. 8a shows the behaviour for OPs 14-25. With the increasing operation point index, there is a monotonous decrease of quality for all frames. Sharp PSNR peaks can be observed at frames number 1 and 17, which are in fact the boundaries of a GOP (I frames). These peaks mean that the the PSNR of the frames in the lowest temporal layer is high in case all the quality enhancement information is available. When the enhancement information is discarded from the lowest temporal layer, the PSNR of I frames drops drastically. As a result, what we can observe for both extreme cases, being the highest and the lowest overall mean PSNR for OP1 and OP25, is that there is a significant quality fluctuation present within a group of pictures. This unwanted effect should be eliminated by the rate control mechanism of the encoder. Unfortunately, no rate control has been implemented in the JSVM reference encoder for scalable video coding so far. Fig. 8b shows the PSNR of frames within two GOPs of OPs 26-30. In this case, enhancing the higher temporal layers only brings no improvement over OP30.

An interesting thing can be observed in the plots for the ”tractor” sequence in Fig. 9. In contrast to high PSNR fluctuations appearing for OPs 14-25 (Fig. 9a), the enhancement of higher temporal layers only results in a smoother PSNR curve as shown in (Fig. 9b) It can be expected that the PSNR fluctuations are annoying for the user and their elimination is desirable. When properly configuring the encoder, enhancement of the higher temporal layers only should lead to a more equal improvement of the quality of all frames throughout the sequence.

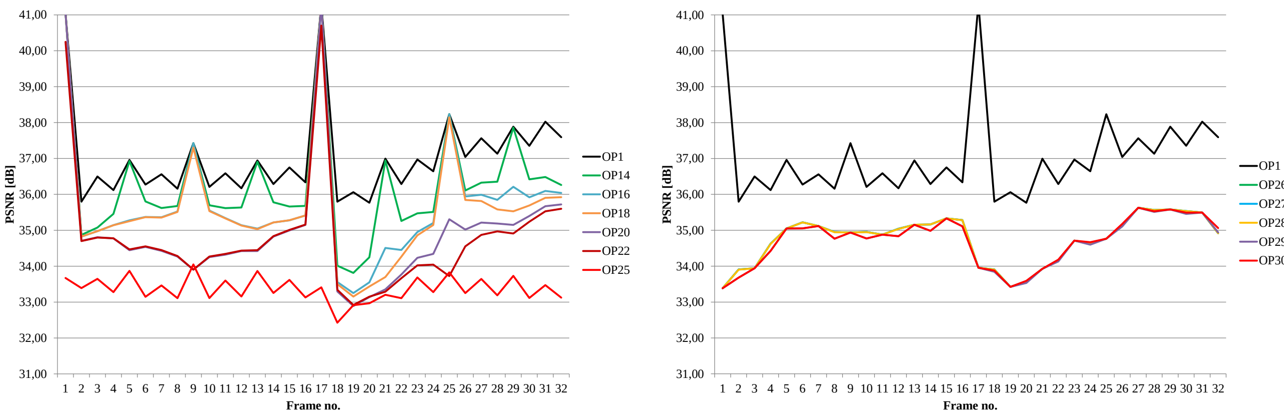
IV. Conclusion

The usability of H.264/SVC for efficient adaptive video transmission depends, to a larger extent, on the desired usage scenario. In our experiment, the usage scenario was high definition video, whose application is very likely to be found in IP-based television or wireless multimedia services. An approach for increasing the quality granularity of a scalable bit stream was presented, dropping the quality enhancement packets from different temporal layers according to a defined scheme.



a) b)

Figure 8. The blue_sky sequence: PSNR per frame for operation points 1 and a) 14 to 25; b) 26 to 30.



a) b)

Figure 9. The tractor sequence: PSNR per frame for operation points 1 and a) 14 to 25; b) 26 to 30.

It has been shown that the overall performance of H.264/SVC in this scenario is considerably worse compared to the very well performing implementation of H.264/AVC. However, one has to keep in mind that the SVC reference encoder lacks efficient rate control, which degrades its coding efficiency. It can be expected that with other encoder implementations, the rate-distortion performance of the SVC encoder approaches AVC more closely.

Analyzing the cumulative distribution function of the packet sizes for the considered operation points has revealed the fact that quality enhancement data for higher temporal layers are carried in a higher number of smaller packets, which can be a useful information from the network point of view. It is also in a perfect agreement with the fact that with high temporal layer enhancement, a higher number of frames is enhanced by a smaller PSNR difference.

Finally, it has been shown that when the enhancement of the lower temporal layers only is kept, the quality of video frames in time tends to exhibit fluctuations in terms of PSNR. This effect does not appear when the enhancement of higher temporal layers only is employed. In order to exploit the impact of the quality fluctuations on the quality perceived by the consumer, the authors plan to conduct a subjective testing session. In case it is proved that removing these fluctuations would bring an overall improvement of the video quality, an optimization of the packet dropping strategy based on high temporal layer enhancement will be considered.

ACKNOWLEDGEMENT

The paper was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. The support of the project CZ.1.07/2.3.00/20.0007 WICOMT, financed from the operational program Education for competitiveness, is gratefully acknowledged. This work was supported by the Czech Ministry of Education under project number LD11081.

REFERENCES

[1] Cisco Systems, Inc. Cisco Visual Networking Index: Forecast and Methodology, 2011-2016, San Jose, CA, USA: Cisco Systems, 2012.

[2] Cisco Systems, Inc. Cisco Visual Networking Index: Forecast and Methodology, 2009-2014, San Jose, CA, USA: Cisco Systems, 2010.

[3] D. Van Deursen, W. Van Lancker, and R. Van de Walle, "On media delivery protocols in the Web," In 2010 IEEE International Conference on Multimedia and Expo (ICME), 2010, pp. 1028 - 1033.

[4] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over internet," IEEE Multimedia, vol. 18, no. 4, 2011, pp. 62-67.

[5] 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, Special Issue on Video Coding, vol. 17, no. 9, 2007, pp. 1103-1120.

[6] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, 2007, pp. 1194-1203.

[7] T. Oelbaum, H. Schwarz, M. Wien, and T. Wiegand, "Subjective performance evaluation of the SVC extension of H.264/AVC," In 15th IEEE International Conference on Image Processing, ICIP 2008, San Diego (CA, USA), 2008, pp. 2772-2775.

[8] SVC Reference Software (JSVM) [Online, retrieved: January, 2012] Available at: http://ip.hhi.de/imagecom_G1/savce/downloads/SVCReference-Software.htm

[9] R. Gupta, A. Pulipaka, P. Seeling, L. J. Karam, and M. Reisslein, "H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation," IEEE Transactions on Broadcasting, vol. 58, no. 3, 2012, pp. 428 - 439.

[10] A. Eichhorn and P. Ni, "Pick your Layers wisely - A Quality Assessment of H.264 Scalable Video Coding for Mobile Devices," In IEEE International Conference on Communications, ICC, Dresden (Germany), 2009, pp. 1 - 6.

[11] L. C. Daronco, V. Roesler, and J. V. de Lima, "Subjective Video Quality Assessment Applied to Scalable Video Coding and Transmission Instability," In Proceedings of the 2010 ACM Symposium on Applied Computing, SAC'10, Sierre (Switzerland), 2010, pp. 1898 - 1904.

[12] L. C. Daronco, V. Roesler, J. V. de Lima, and R. Balbinot, "Quality analysis of scalable video coding on unstable transmissions," Springer Multimedia Tools and Applications, 2011, DOI: 10.1007/s11042-011-0760-y.

[13] Xiph.org Video Test Media [Online, retrieved: June, 2012], Available at: http://media.xiph.org/video/derf/ .

[14] K. Suehring, The H.264/AVC Reference Software, version 18.2. [Online, retrieved: April, 2012]. Available at: http://iphome.hhi.de/suehring/tml/download/ .