SUBJECTIVE QUALITY ASSESSMENT OF MPEG-4 SCALABLE VIDEO CODING IN A MOBILE SCENARIO

Yohann Pitrey, Marcus Barkowsky, Patrick Le Callet, Romuald Pépion

SUBJECTIVE QUALITY ASSESSMENT OF MPEG-4 SCALABLE VIDEO CODING IN A MOBILE SCENARIO

Yohann Pitrey, Marcus Barkowsky, Patrick Le Callet, Romuald Pepion

Institute for Research in Communications and Cybernetics of Nantes (IRCCyN)
Image and Video Communications (IVC) group
École Polytechnique de l'Université de Nantes – 44306 Nantes, FRANCE
{*yohann.pitrey, marcus.barkowsky, patrick.lecallet, romuald.pepion*}*@univ-nantes.fr*

ABSTRACT

Scalable Video Coding provides several levels of video encapsulated in a single video stream. In a transmission scenario such as broadcasting, this structure is quite advantageous as it can be used to address heterogeneous decoding targets with variable needs and requirements. However, this adaptability comes at a slight cost in coding efficiency when compared to single-layer coding. Based on subjective experiments, this cost is evaluated in this paper by comparing the new MPEG-4 Scalable Video Coding (SVC) standard with the now-established MPEG-4 AVC standard. Two scenarios are analyzed in the context of mobile transmission applications. The first scenario uses the same bitrate for SVC and AVC, leading to a slightly lower PSNR for SVC. The second scenario uses the same PSNR for SVC and AVC, leading to a slightly lower bitrate for AVC. The results of the subjective tests illustrate several interesting aspects of the relation between the performance of the two standards. First, we observe that the offset between AVC and SVC is not severe, though statistically significant in terms of user Mean Opinion Score (MOS) in such a context. Second, while adding another layer to SVC always leads to a performance loss, the impact of this loss decreases when the number of layers increases.

Index Terms— Scalable video coding, Subjective experiments, Quality evaluation.

1. INTRODUCTION

Video coding has become an essential area in today's visual communications. For nearly 30 years, research has been carried out to design technics to represent the video contents with as few data as possible. One of the latest accomplishments in the field of video coding is the H.264/MPEG-4 Advanced Video Coding (AVC) standard [1]. Using spatial and temporal prediction to remove redundancy, coupled with Discrete Cosine Transform (DCT), scalar quantization and entropy coding, AVC outperforms the coding performance of previous standards in a significant way. A Network Abstraction Layer (NAL) is additionally used to ease the transmission of the video on packet-based networks. Using this new standard, high quality can be reached at relatively low bitrates to produce video streams particularly adapted for video transmission.

Nevertheless, the need for flexibility is becoming crucial in current video applications. Typically, a video service provider has to address heterogeneous display devices such as mobile phones or residential televisions, using different transmission channels such as DSL links or wireless technologies such as WiFi or GSM. Despite its high performance when compared to previous video coding standards, AVC does not provide any convenient tools to adapt a stream to different targets. Consequently, several versions of the same video content have to be encoded for each type of target. In a broadcasting context, these different versions have to be transmitted all together in order to address all types of targets. This kind of scenario called *simulcast* obviously leads to a waste both in bandwidth and storage space.

MPEG-4 Scalable Video Coding (SVC) has been introduced by the Joint Video Team (JVT) as a response to provide flexibility for video coding and to minimize the waste of resources produced by simulcast [2]. It allows several resolutions of the same video content to be encoded as different *layers* of a single scalable stream. Three types of scalability are supported, to address a wide range of targets. *Spatial* scalability enables frame size tuning, to adapt the video contents for mobile phones as well as for higher spatial resolution display devices such as computers and high-definition televisions. *Quality* scalability (also referred to as *fidelity* scalability) provides increasing levels of quality of the stream (in terms of PSNR), essentially to adapt the bitrate needed to represent the video. *Temporal* scalability allows the number of frames per second to vary from one layer to another, in order to propose a balance between the motion smoothness and the amount of data to process. In addition to the three types of scalability, a new tool called *inter-layer prediction* is also provided in order to minimize the waste caused by the encoding of several versions of the video. Using inter-layer prediction, spatial and temporal prediction in an enhancement layer can

use the information from a lower layer called the *base layer* to get a better prediction. The redundancy between layers is therefore reduced, improving the coding efficiency of the whole video stream.

Coding a video stream using several layers provides intermediate quality and resolution levels. The stream can be adapted more easily to the different display devices and transmission conditions. Additional value is provided by the ability to adapt the stream according to the actual resources available at a given time. This is particularly useful in mobile environments as the bandwidth and receiving conditions can vary according to environmental factors. A typical use-case would concern a user receiving a video stream on a mobile phone through a given network such as WiMax. If the receiving conditions change due to a loss of the WiMax signal (e. g. in a vehicle in motion), SVC allows to adapt the resolution of the video to the new environment. The user gets a version of the video with lower size, temporal resolution or quality, instead of having to wait for the full-size video data to reach his mobile again.

The enhanced adaptability made possible by SVC logically comes with a price in terms of coding efficiency when compared to AVC. Indeed, coding a stream using several layers re-introduces redundancy, as inter-layer prediction is not capable to fully exploit the correlations between layers. Therefore, it is to expect that to reach a given quality, the bitrate needed to encode several SVC layers is slightly higher than the bitrate needed to encode a single AVC layer. In a similar way, getting the same bitrate for SVC and AVC means a slightly lower quality for SVC.

In this paper, we evaluate the loss introduced by SVC using subjective tests in a mobile transmission context. The test setup we propose involves all three types of scalability, and uses resolutions and bitrates well-suited for lightweight applications. We evaluate the video quality using a validated methodology and controlled viewing conditions, in order to get reliable data and to build a precise analysis of the test results. This paper is organized as follows. Section 2 presents the AVC and SVC standards. Section 3 presents the encoding scenarios we used for our tests, as well as the subjective quality evaluation methodology. Section 4 analyzes the test results, while section 5 concludes the paper.

2. MPEG-4 SVC PERFORMANCE EVALUATION

The performance of SVC has been studied both in terms of objective and subjective quality. Rate-distortion performance analysis using the PSNR as a quality metric, such as presented in [3, 4], confirm that the performance of SVC is not as good as the performance of AVC, but remains comparable if a small bitrate overhead is permitted. It is shown that with equal PSNR, SVC can save from 17% to 40% of bitrate when compared to an AVC simulcast scenario.

Objective quality metrics (and especially PSNR) do not always depict the actual perceived quality. Therefore some work has also been presented on subjective quality evaluation of the performance of SVC [5, 6, 7]. Although it involves expensive hardware and requires strict test conditions, this type of evaluation is closer to the final feeling of quality experienced by a viewer. Therefore the results of subjective tests usually allow a more detailed analysis of the impact of a particular effect on the videos.

For the verification test plan of SVC [5], subjective tests were performed on various SVC configurations, including videoconference, mobile and residential broadcasting and professional video production. This report shows that with a bitrate overhead of about 10% in favor of SVC, no significant visual quality difference could be noticed. Similar results were published by the European Broadcasting Union (EBU) [6], in which the performance of SVC is compared to AVC using a subjective metric in the context of residential broadcasting (typically from Standard Definition to High Definition television). It was stated that SVC can compete with AVC, particularly when using high bitrates and a good quality for the base layer. Most of the contributions for performance analysis for SVC focus on residential TV scenarios. The mobile transmission context has been studied in [5], but the test scenario does not fully exploit the possibilities of SVC.

In this paper we report the results of subjective quality assessment tests that were performed to compare AVC simulcast and SVC in a mobile transmission context. The next section describes the subjective quality methodology we used for these tests and the encoding scenarios that were presented to the viewers.

3. SUBJECTIVE EXPERIMENT SETUP

In our experiment, the purpose is to compare streams encoded using SVC to their equivalent in an AVC simulcast scenario. Four commonly-used video contents are processed : Harbour, Soccer, City and Crew. They reflect a wide variety of contents, ranging from sports to documentary. The SVC video streams contain four layers with frame size, temporal frequency (fps) and bitrate designed for typical mobile transmission applications. A summary is provided for reference in Table 1.

The base layer (denoted as SVC1 in the following) is in QVGA format (320×240 pixels) at 15 frames per second. It is particularly suitable for devices such as mobile phones with small screen size and limited processing power. It is encoded at 100 kilobits per second (kbps), which makes it possible to transmit over UMTS-like channels that typically have a limited transmission capacity of 384 kbps. The second layer for the SVC coding is a temporal enhancement layer (denoted as SVC2), adding another 15 frames per second to the base layer. The decoded video is thus in QVGA format at 30 fps, which makes it affordable for mobile phones with small screen size but higher processing power. The total bi-

trate is equal to 250 kbps, which is suitable for HSDPA-like channels. The third layer is a dyadic spatial layer (SVC3), which doubles the frame size. This layer is thus in VGA format (640×480 pixels) at 30 fps. It is adapted for smartphone-like devices with a higher screen resolution and a medium processing power. The total bitrate is 750 kbps, which is designed for WiFi or WiMax networks. The fourth layer (SVC4) is a quality enhancement layer. Another 250 kbps is spent on enhancing the quality in terms of PSNR, leading to a total bitrate of 1000 kbps. This may be transmitted over WiMax if the network conditions are appropriate. The frame size and rate do not change from SVC3 to SVC4.

The input video sequences were generated from full-HD sequences (1920×1080 pixels at 60 frames per second), down-scaled using the DownConvert tool provided by the JSVM. The reference software for MPEG-4 SVC Version 8.6 was used to encode all streams [8]. This software is capable of encoding streams both in AVC and SVC. To encode an AVC stream, the number of layers is set to 1. In order to get a fixed bitrate, the FixedQPEncoder utility provided by the JSVM was used. This iterative tool executes the encoding with different parameters until a target bitrate value is reached. It should be stressed that the bitrates are expressed as total bitrates. This means that the sum of bitrate used to encode the full SVC stream containing four layers is equal to 1000 kbps.

From the bitstream used in condition SVC4, all the other conditions SVC3, SVC2, and SVC1 can be extracted without re-encoding. This is possible in SVC because each layer uses the lower layer as a base layer for inter-layer prediction. This is not a feature of AVC, thus two different scenarios for AVC are considered that can be useful to learn about the differences in coding efficiency between the two standards. In the first scenario, the same bitrate is allowed for AVC as for SVC. Four different encodings per video content are used, corresponding to the four conditions used for SVC. They are denoted as AVC-B1 to AVC-B4, where the letter B stands for "bitrate". Please note, that this is a different scenario for the network provider as it might no longer be possible to transmit all qualities over the same channel. The total bitrate for transmitting all four layers to the end user on the same network would be the sum of the individual bitrates, which is 2100kbit/s. In the second scenario, the same PSNR value is used for AVC as for SVC. For each decoded sequence of the four SVC layers, a PSNR value is calculated. These 16 PSNR values are then used to generate an equivalent AVC bitstream. Again, the FixedQPEncoder is used because it can handle a target PSNR value as well. These four conditions for each content are termed AVC-Q1 to AVC-Q4, where the letter Q stands for "quality".

To evaluate the visual quality of the different scenarios, the Subjective Assessment Methodology for VIdeo Quality (SAMVIQ) methodology is used [9]. This methodology uses multiple stimuli assessment, which means that a viewer is allowed to watch each video as many times as he wants. The viewer gives a score comprised between 0 and 100 on a continuous quality scale (0 being the score of the lowest quality and 100 being the score of the highest quality). The viewer can choose the order in which he wants to view the sequences, while every sequence has to be evaluated once. As the viewer has the opportunity to revise his judgment by viewing each sequence multiple times, the accuracy of the measure is increased when compared to other subjective experiment methods.

Table 1. Tested layer configurations for MPEG-4 SVC.

Layer Label	Frame Size	Frames p. sec.	Bitrate (kbps)	Scalability	Application
SVC1	QVGA	15	100	–	UMTS
SVC2	QVGA	30	250	TEMP	HSDPA
SVC3	VGA	30	750	SPAT	WiMax
SVC4	VGA	30	1000	QUAL	WiMax

Table 2. Test configurations for the AVC-B scenario.

Layer Label	Frame Size	Frames p. sec.	Bitrate (kbps)
AVC-B1	QVGA	15	100
AVC-B2	QVGA	30	250
AVC-B3	VGA	30	750
AVC-B4	VGA	30	1000

Among the sequences presented to the viewer during a test session, two sequences are used as references. The first sequence is explicitly labeled as the high quality reference for the time of the presentation. The second sequence is hidden and randomly presented amongst other coded sequences. In this experiment, the original non-coded video stream is used as the high-quality reference. As the SAMVIQ methodology recommends not to compare two different frame formats in the same test, the QVGA and the VGA formats are handled in two separate test sessions. As a result, the two conditions 1 and 2 are processed in a first test session, while the conditions 3 and 4 are processed in a second test session. As the viewers involved in the first session were different from the viewers from the second session, it is not possible to consider the results of the two sessions as parts of the same test.

Each test session involved 15 viewers, whose age ranged from 17 to 50, with an average of 25. The screen model was a Samsung SyncMaster 1100MB reference screen. According to the ITU recommendation, the distance between the viewer and the display was equal to 6 times the height of the displayed image for the videos in QVGA format, and 4 times for the videos in VGA format. Each viewer was asked to rate 28 video sequences, during a session of approximately 30 min-

Table 3. Results of the t-test and MOS comparison between the presented AVC and SVC scenarios.

QVGA	Ref	SVC1	SVC2	AVC-B1	AVC-B2	AVC-Q1	AVC-Q2
Ref		↑	↑	↑	↑	↑	↑
SVC1	↓		↓	·	↓	·	↓
SVC2	↓	↑		↑	↓	↑	·
AVC-B1	↓	·	↓		↓	·	↓
AVC-B2	↓	↑	↑	↑		↑	↑
AVC-Q1	↓	·	↓	·	↓		↓
AVC-Q2	↓	↑	·	↑	↓	↑	

VGA	Ref	SVC3	SVC4	AVC-B3	AVC-B4	AVC-Q3	AVC-Q4
Ref		↑	↑	↑	↑	↑	↑
SVC3	↓		↓	↓	↓	·	↓
SVC4	↓	↑		↓	↓	·	·
AVC-B3	↓	↑	↑		·	↑	·
AVC-B4	↓	↑	↑	·		↑	↑
AVC-Q3	↓	·	·	↓	↓		↓
AVC-Q4	↓	↑	·	·	↓	↑	

utes.

The experimental data from the presented tests is used in the next section to compare the coding performance of AVC and SVC in the presented scenarios.

4. EXPERIMENTAL RESULTS

Figure 1 compares the average Mean Opinion Score (MOS) of the AVC and the two SVC scenarios for the four video contents. We can observe that the scores of SVC and AVC-Q are very close for all tested conditions.

As the AVC-Q streams are designed to have the same PSNR as the corresponding SVC layers, we can state that in the context of our experiment, the PSNR is a sufficient approximation for visual quality when evaluating the same content. To reinforce this statement, Table 3 includes the results of the student t-test on the experimental data. The configurations presented to the viewer for each video content are compared in terms of statistical difference. Two configurations can be considered not statistically different when the corresponding cell in Table 3 contains a '·' symbol. On the opposite, if two configurations a and b are statistically different, it is possible to order them in terms of visual quality. If the MOS of configuration a is higher than the MOS of configuration b, the cell located at line a and column b in Table 3 contains a '↑' symbol. Otherwise (the MOS of configuration a is less than the MOS of configuration b), the cell contains a '↓' symbol. According to the t-test, we can confirm that there is no statistically significant difference between the SVC and AVC-Q scenarios.

For the first layer in QVGA at 15 frames per second, the t-test also confirms that there is no visible difference between SVC and the two AVC conditions. This result is expected as the base layer of SVC uses the same encoding algorithm as AVC.

Figure 2 displays the Mean Opinion Scores (MOS) obtained on the presented configurations for each video sequence. The results of the two test sessions (QVGA and VGA) are displayed together to ease the reading, while no relation between the two parts of the chart should be assumed.

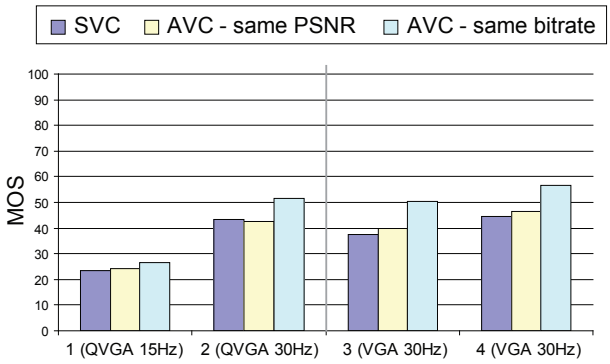

Fig. 1. Average Mean Opinion Score for the four processed sequences.

The scores obtained with the first condition are very low, typically located between 10 and 30, which corresponds to *bad* or *low* quality. The difference between AVC and SVC for this layer is very small, because the base layer of SVC is coded as a single AVC layer. Additional header information is included in the SVC version, which explains the slight difference in favor of AVC. On the second condition the scores are higher, because of the higher number of frames per second (*cf.* Table 1) which gives a smoother motion information. For this condition the difference between SVC and AVC is more noticeable in terms of MOS, but remains below 10 MOS points for three sequences out of four which can be considered equal. The difference is higher for CREW which is known to be a complex sequence to encode. It contains high motion and camera pan with camera light flashes causing abrupt luminance changes. The temporal enhancement between condition 1 and 2 may not allow efficient inter-layer prediction, as these light flashes are typically located on several isolated frames. So the additional frames in condition 2 may not be

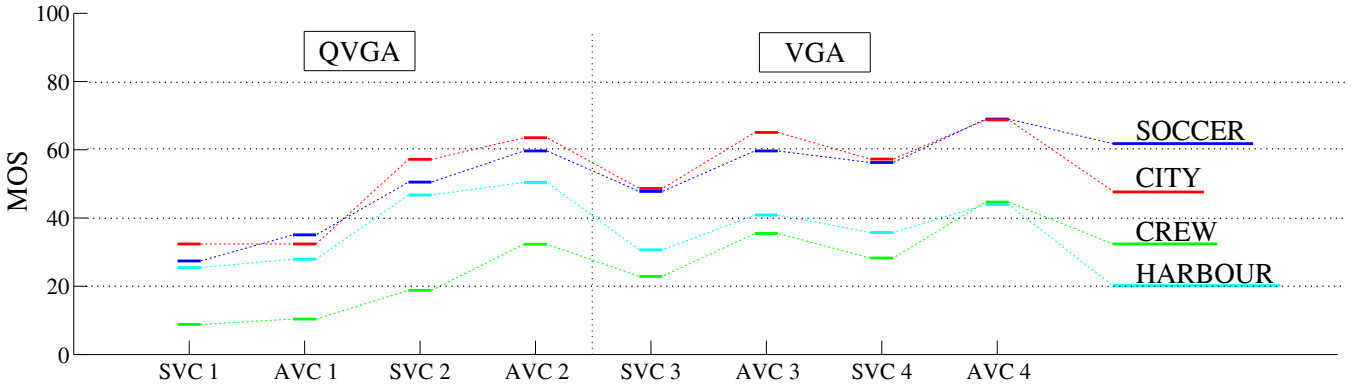

Fig. 2. Mean Opinion Scores (MOS) obtained for AVC and SVC using the SAMVIQ methodology.

close in terms of content to the frames in the base layer, leading to poor inter-layer redundancy.

On the third condition (spatial enhancement from QVGA to VGA), the difference in MOS between SVC and AVC is quite constant for the four sequences, supporting the hypothesis that the lower efficiency previously observed on CREW is related to temporal scalability. The difference in terms of MOS varies around 10 points between AVC and SVC, which is a significant difference according to the t-test.

For SVC, the scores of condition 4 (VGA at 30 fps + quality enhancement) are significantly higher than the scores of condition 3. This layer adds 250 kbps of quality refinement to condition 3 (*cf.* Table 1), which represents 25% of the total bitrate. On the contrary for AVC, using the same bitrate constraint (750 kbps and 1000 kbps) the sequences are perceived as equivalent in a statistical sense.

When comparing AVC to SVC for the VGA format, it can be observed that AVC at 750 kbps outperforms SVC even at 1000 kbps. It has to be kept in mind though that the SVC bitstream for the VGA format already contains three layers, which allow decoding the video at intermediate quality levels.

Table 4 reports the average difference in MOS between SVC and AVC for each layer. It can be observed that between conditions 1 and 2 this difference varies more than between conditions 3 and 4. This tends to show that the loss introduced by SVC is higher when the number of layers is low. Going from one single layer to two layers introduces a higher loss than going from 3 layers to 4 layers. An important parameter of this effect might be the position of the added layer in the scalability hierarchy. Adding a new layer at the bottom of the hierarchy has an impact on the whole structure of the stream, whereas adding it at the top of the hierarchy has a more limited impact. This result could help a service provider to decide whether he should add a new layer in a scalable stream configuration. In such a situation, our tests show that besides the importance of this new layer for the new target to address, its impact on the overall coding efficiency should be estimated.

Layer 1 (QVGA@15Hz)	**Layer 2** (QVGA@30Hz)	**Layer 3** (VGA@30Hz)	**Layer 4** (VGA@30Hz+)
2.9	8.2	12.7	12.2

Table 4. Average MOS difference between SVC and the corresponding AVC-B stream.

5. CONCLUSION

In this paper we presented the results of subjective tests to compare the performance of MPEG-4 SVC and AVC in a mobile communication context. Two scenarios were studied for AVC, first using the same bitrate as SVC, second using the same PSNR as SVC. Using a detailed analysis of the results, we were able to confirm that the loss introduced by SVC is not severe though statistical differences are observed for all enhancement layers. The difference compared to AVC in terms of MOS is usually below 10%, which can be considered as sufficiently small. Secondly, we showed that this loss tends to stabilize when the number of layers increases. The impact of a given layer on the global coding efficiency is also related to its position in the scalable hierarchy, meaning that the more layers rely on it for prediction, the higher the impact. Finally, it was noticed that the scenario encoded with equal PSNR leads to equal visual quality, making the PSNR a good approximation of the viewer's opinion when used on each content individually. As a conclusion, our tests confirm the results previously published, while providing a precise analysis of the performance of SVC in a mobile context that fully exploits the capabilities of SVC. They show that SVC is of significant interest in mobile broadcasting, as it can address various targets with a limited cost in visual quality.

6. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard,"

Circuits and Systems for Video Tech., IEEE Trans. on, vol. 13, no. 7, pp. 560–576, July 2003.

[2] J. Reichel and H. Schwarz and M. Wien , " Scalable Video Coding - Joint Draft 4 ," Tech. Rep., Joint Video Team, doc. JVT-Q201 , year = 2005 ,.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable H.264/MPEG4-AVC Extension," *International Conference on Image Processing (ICIP)*, 2006.

[4] M. Wien, H. Schwarz and T. Oelbaum, "Performance Analysis of SVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2007.

[5] ISO/IEC JTC 1/SC 29/WG 11 N9577 , " SVC Verification Test Report ," Tech. Rep., Joint Video Team , 2007.

[6] A. Kouadio, M. Clare, L. Noblet and V. Bottreau , " SVC – A highly-scalable version of H.264/AVC ," *EBU Technical Review – 2008 Q2* , 2008 .

[7] N. Staelens, S. Moens, W. Van den Broeck, I. Marien, B. Vermeulen, P. Lambert, R. Van de Walle and P. Demeester, "Assessing the perceptual influence of H.264/SVC Signal-to-Noise Ratio and temporal scalability on full length movies," *IEEE International Workshop on Quality of Multimedia Experience (QoMEx)*, 2009.

[8] http://ip.hhi.de/imagecom_G1/savce/downloads/ , " JSVM Reference Software ," Version 8.6 .

[9] Question ITU-R 81/6 , " SAMVIQ Subjective Assessment Methodology for video quality ," *ITU-T Recommendation 6Q/23E* , 2003 .