



**HAL**  
open science

## Evaluation of MPEG4-SVC for QoE protection in the context of transmission errors

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

► **To cite this version:**

Yohann Pitrey, Marcus Barkowsky, Patrick Le Callet, Romuald Pépion. Evaluation of MPEG4-SVC for QoE protection in the context of transmission errors. SPIE Optical Engineering, Aug 2010, San Diego, United States. hal-00608337

**HAL Id: hal-00608337**

**<https://hal.science/hal-00608337>**

Submitted on 12 Jul 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluation of MPEG4-SVC for QoE protection in the context of transmission errors

Yohann Pitrey, Marcus Barkowsky, Patrick Le Callet, Romuald Pepion

IRCCyN Lab, Image & VideoCommunications Group.  
Ecole Polytechnique de l'Universite de NANTES, FRANCE.

## ABSTRACT

Scalable Video Coding (SVC) provides a way to encapsulate several video layers with increasing quality and resolution in a single bitstream. Thus it is particularly adapted to address heterogeneous networks and a wide variety of decoding devices. In this paper, we evaluate the interest of SVC in a different context, which is error concealment after transmission on networks subject to packet loss. The encoded scalable video streams contain two layers with different spatial and temporal resolutions designed for mobile video communications with medium size and average to low bitrates. The main idea is to use the base layer to conceal errors in the higher layers if they are corrupted or lost. The base layer is first upscaled either spatially or temporally to reach the same resolution as the layer to conceal. Two error-concealment techniques using the base layer are then proposed for the MPEG-4 SVC standard, involving frame-level concealment and pixel-level concealment. These techniques are compared to the upscaled base layer as well as to a classical single-layer MPEG-4 AVC/H.264 error-concealment technique. The comparison is carried out through a subjective experiment, in order to evaluate the Quality-of-Experience of the proposed techniques. We study several scenarios involving various bitrates and resolutions for the base layer of the SVC streams. The results show that SVC-based error concealment can provide significantly higher visual quality than single-layer-based techniques. Moreover, we demonstrate that the resolution and bitrate of the base layer have a strong impact on the perceived quality of the concealment.

## 1. INTRODUCTION

Communication networks are subject to data-loss and corruption, due to various factors such as congestion, routing errors or hardware failures. When transmitting video data on such networks, these phenomena create visual artifacts in the displayed video, which usually have a dramatic impact on the end-user's feeling of quality. Therefore these errors need to be detected and concealed using appropriate methods.

Due to the predictive structure of video coding techniques such as MPEG-4 AVC/H.264,<sup>1</sup> the errors are likely to propagate along the decoded data. Detecting and predicting the impact of loss on the visual quality are thus quite difficult tasks, as well as concealing the visual artifacts using reliable data. The errors are usually concealed using spatially or temporally close areas in the decoded domain.<sup>2,3</sup> Two types of error-concealment techniques can be distinguished. Spatial techniques rely on the surrounding areas to fill lost regions by using interpolation and smoothing operations.<sup>4</sup> Temporal techniques estimate the motion information of the lost parts and use the most plausible data as a basis for temporal prediction.<sup>5</sup> Also, the encoding structure can be modified so that consecutive data in the bitstream corresponds to scattered areas in the decoded video. Thus, the distortions caused by loss are likely to be scattered as well, reducing the visual impact and allowing for better error concealment.<sup>6</sup>

Other error-concealment techniques have been introduced, based either on statistical properties of the bitstream, or on more practical considerations. For instance a study was presented where different types of gap fillings for channel switching in mobile TV were considered.<sup>7</sup> Surprisingly, the observers preferred watching a commercial to seeing a black screen, a sender logo or a "please wait" screen. Thus, for longer outages, showing a prerecorded and prefetched commercial might be considered as an alternative to the application of error concealment strategies in particular when MPEG-4 AVC is used and no graceful degradation method exists.

Scalable Video Coding (SVC) has been introduced to provide a way to adapt video streams to variable decoding contexts. Several video layers are encoded in a single bitstream, with increasing spatio-temporal resolution and quality. In addition, inter-layer prediction is used to increase the global coding efficiency of the multi-layer stream. Thus, a layer can use the prediction and texture information from a lower layer called the *base layer* in order to get a better prediction. Scalable Video Coding therefore achieves higher compression rates than MPEG-4 AVC when encoding several layers

jointly. A scalable video coding standard called MPEG-4 SVC was finalized in 2007, based on the MPEG-4 AVC/H.264 standard.<sup>8</sup> It has been shown that it can allow for bitrate savings from 17% to 40% when compared to AVC, depending on the encoding configuration.<sup>9</sup>

SVC provides intermediate levels of video. Considering the bottom-up coding structure of SVC, the base layer can be independently decoded from the enhancement layers. Therefore, if the base layer is assumed to be unaffected by loss, the information it contains can be used in the enhancement layers in order to replace the erroneous areas. Postprocessing operations might be required as the layers are likely to have different spatial and temporal dimensions. Nonetheless, these postprocessings might produce spatial and temporal artifacts similar to classical error-concealment based on single-layer coding. Therefore, SVC represents an interesting candidate for video error-concealment. In addition, it can be easily combined with error-protection schemes. Especially, SVC is particularly adapted for unequal error-protection because of its hierarchical layer-based coding structure. The base layer is quite important for the decoding process, while it represents a limited amount of data when compared to the enhancement layers. A high-protection technique can then be applied on the base layer, while the enhancement layers can be protected in a more affordable way.<sup>10,11</sup>

Several studies have been proposed that use SVC as a way to perform error-concealment. For example, it has been shown that the erroneous areas in the enhancement layer may be ignored by using the motion and texture information from the base layer in order to conceal the errors.<sup>12</sup> This method is quite simple and shows interesting results in terms of concealment. A more complex method was published recently using the information from the upscaled base layer to replace the erroneous areas in the enhancement layer, based on a technique called frame-hallucination.<sup>13</sup> This technique involves a training process and requires a set of available images in order to choose the best image to use for concealment. It shows interesting results when compared to the simple method based on the reconstruction of the base layer, at the price of a relatively high computational complexity. Unfortunately, in both studies the results are presented in terms of average Peak-to-Signal Ratio (PSNR) improvement, which is not representative of the human-perceived quality.

In this paper, we propose two error-concealment techniques using the base layer of SVC to conceal errors in the enhancement layer. Specifically, we evaluate the influence of the spatio-temporal characteristics of the base layer as well as the bitrate used to encode it on the quality of the concealment. These two methods are compared to a classical error-concealment technique suited for single-layer AVC, in order to evaluate the interest of SVC in such a context. The comparison is carried out through a subjective experiment, therefore the presented results reflect the visual perception of human viewers.

This paper is organized as follows. Section 2 presents the process of creating the videos for the subjective experiment. Section 3 describes the setup for the subjective experiment. Section 4 presents the experimental results. Finally, section 5 concludes the paper.

## 2. DETAILS OF THE SUBJECTIVE EXPERIMENT

The process for generating the test videos is divided into four steps, from downscaling the reference sequences, encoding, transmission-error simulation to decoding and error-concealment. These four steps are described in detail in this section.

### 2.1 Input formats and video contents

The context of our experiment is mobile transmission. Thus we use videos in VGA format ( $640 \times 480$  pixels) at 30 frames per second (fps), which is a commonly used format for current smartphones and lightweight devices. Additionally, we use videos in QVGA format ( $320 \times 240$  pixels) at 15 and 30 fps in the SVC error-concealment scenarios as base layer videos. The input videos are generated from non-coded videos in full-HD ( $1920 \times 1080$  pixels) YUV 4:2:0 format at 60 fps using spatial and temporal downsampling. We use the tools provided by the JSVM reference software suite to perform these downsampling operations.<sup>14</sup> The source videos are first cropped to get a sequence with 4:3 aspect ratio from the 16:9 input sequence. Then spatial downsampling is performed using the Lanczos filter to get VGA ( $640 \times 480$ ) and QVGA ( $320 \times 240$ ) sequences. The cropping and downsampling steps are performed separately to ensure that the content does not appear horizontally or vertically stretched. Temporal downsampling is performed by discarding one frame out of two on each step. Thus one temporal down-scaling step is needed to get the streams at 30 fps, and two steps to get the streams at 15 fps.

We use twelve different video sequences containing a wide variety of contents, ranging from documentary to sports scenes and sequences with high spatial frequency details. Figure 1 displays snapshots of these video sequences illustrating the different types of contents.

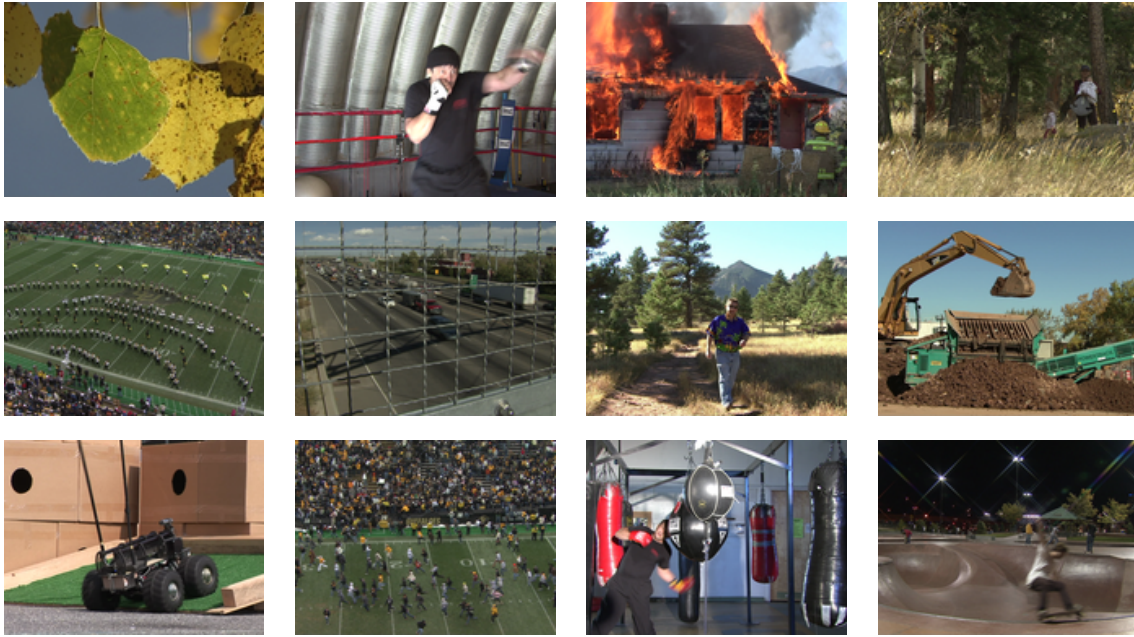


Figure 1. Video contents used for the experiment. Snapshots appear as they were cropped in VGA format.

## 2.2 AVC and SVC encoding

After downsampling, the videos are encoded using the MPEG-4 SVC reference encoder software. This software is capable of encoding both AVC and SVC streams, thus allowing us to compare the two standards using the same coding tools. The SVC streams contain two layers. The highest layer is always in VGA format at 30 fps, with a cumulated bitrate of 600 kilobits per second (*i.e.*: the total bitrate for the two layers is equal to 600 kbps). The base layer is always in QVGA format. We study four different scenarios involving changes in the temporal frequency and the bitrate of the base layer. In the first scenario, the base layer is at 15 fps and is encoded at 120 kilobits per second (kbps). In the second scenario, the base layer is at 30 fps with the same bitrate of 120 kbps. In the third scenario, the base layer is at 15 fps with a higher bitrate of 200 kbps. In the fourth scenario the base layer is at 30 fps and 200 kbps. The two scenarios at 15 fps allow us to evaluate the impact of temporal upscaling on the quality, while the two scenarios at 30 fps allow us to identify the impact of the bitrate of the base layer.

One AVC stream with the same spatio-temporal resolution and bitrate as the highest layer of the SVC streams is generated for each video content. It is thus in VGA format at 30 fps, encoded at 600 kbps. We use the same coding tools as for SVC, so that the only difference between the streams is the number of layers. The size of the groups-of-pictures for SVC is equal to 16. An IDR has been inserted after 64 frames. The bitrate control is ensured by the FixedQpEncoder utility delivered with the JSVM reference software. This tool repeats the encoding process using a logarithmic search scheme to reach the desired bitrate.

In the enhancement layer of the SVC stream as well as in the AVC stream, each frame is divided into four independent slices. Each slice is shaped as a horizontal rectangle and takes a quarter of the frame. Due to the format of the videos (*i.e.*: VGA), losing one slice in a frame represents a spatially-significant visual loss. Additionally, it is possible to choose whether the lost slice overlaps significantly with an area of interest or not.

## 2.3 Bitstream extraction and loss simulation

After the encoding process, we simulate packet loss in the bitstreams to generate visual distortions in the decoded videos. We first use the bitstream extractor provided by the JSVM software suite to build a packet trace of the bitstream. This packet trace is a text file that contains one line for each slice in the bitstream. This file is processed through a loss simulator that deletes the lines corresponding to a specific area in a specified set of consecutive frames. The distorted trace-file is then merged with the video bitstream using again the bitstream extractor. The slices corresponding to the deleted lines in the trace file are discarded in the output bitstream, creating a video bitstream that contains loss. One slice out of four is lost

in each distorted frame, which means that one quarter of the frame is missing during the loss interval. We manually chose to lose a slice that is part of a visually important area of the sequence. The length of the loss is set to one second out of ten seconds of video content for each sequence. The spatial and temporal locations at which the loss occurs are designed to destroy visually salient regions such as characters, high motion or fine details.

## 2.4 Decoding and error concealment

For the decoding process, we use a commercially available SVC decoder. Unfortunately, the latest JSVM reference decoder versions are not capable of decoding erroneous streams. Therefore it leads to unexpected crashes when missing or corrupted data occur. The decoder we use is capable of decoding both AVC and SVC streams and proposes a basic error-concealment scheme making it robust to loss. The error-concealment technique is based on a combination of a repetition of the decoder buffer and a motion vector based error concealment. Two examples of the visual artifacts produced by this technique are displayed in Figure 3.

In addition to the AVC decoded stream combined with the buffer repetition-based concealment technique, three scenarios are studied. In the first scenario, only the base layer of the SVC stream is decoded and up-scaled to get a video in VGA at 30 fps. This scenario is based on the assumption that the base layer is not affected by transmission errors. Thus, decoding and upscaling it to the full resolution provides a minimum, yet always available quality. For the two scenarios in which the base layer and the enhancement layer are both in 30 fps, the upscaling operation only consists in spatial up-sampling on each frame from QVGA to VGA. We use the Lanczos filter as an upscaler, which has shown a good efficiency while maintaining a reasonable computational complexity.<sup>15</sup> For the two scenarios in which the base layer is at 15 fps, temporal up-sampling is needed to get the same resolution as the enhancement layer. To get 30 fps, we simply duplicate each frame in the upscaled version of the base layer. This solution benefits from very low computational complexity, while avoiding temporal interpolation issues which are not in the scope of this work.

The two remaining scenarios exploit the structure of the SVC stream to perform error-concealment. The first method consists in switching to the upscaled version of a frame when distortions are detected due to transmission errors. It is referenced later on as the *switch* method, or *frame-level* concealment. In order to simulate the detection of a loss in a frame by the decoder, a frame-wise binary comparison is performed between the decoded video sequences with and without transmission errors. If a difference is detected, the decoder switches to the frame upscaled from the base layer, while it keeps the high resolution frame if no difference is detected.

The second concealment method is based on the same principle as the switch method, whereas the comparison between the frame before and after transmission is performed at the pixel level. Therefore a mask of the distorted areas of the frame is constructed and replaced by the pixels from the upscaled version. This method is referred to as the *patch* method, or *pixel-level* method. As in the previous method, the comparison is performed using a simple pixel-luminance difference. Some visual examples of the distortions produced by the different concealment methods are displayed in Figure 3.

The two concealment methods presented here are expected to have different behaviors, depending on the encoding characteristics of the layers. In the following we consider the upscaled version of the SVC base layer as a concealment method as well. This simplistic scenario might be interesting in the sense that the bitrate saved from not sending (or even encoding) the enhancement layer can be used to perform stronger error-protection on the base layer. Thus the video stream can resist more efficiently to attacks and losses, at the price of some up-scaling artifacts.

In the *switch* method, the frame is replaced at once by the upscaled frame. It represents an intermediate solution, as some high-quality areas might be replaced as well as distorted areas. Meanwhile, the frame keeps a globally constant (yet lower) quality, which might have an impact on the viewers opinion. The *patch* method takes advantage of the fact that the distortions might be localized in a particular area of the frame, which allows to keep the non-affected areas in high quality. The impact of the concealment might therefore be more limited, but discontinuities can be expected to appear at the border between the upscaled and the original areas.

In order to evaluate the performance of each of these SVC-based concealment techniques, we designed a subjective experiment to gather viewers opinions. The setup of this experiment is described in the next section.

### 3. SUBJECTIVE EXPERIMENT SETUP

The subjective experiment setup follows the recommendation ITU-R BT.500 concerning the room setup, the correct illumination and the calibration of the display. The viewing distance of 4H and other parameters which are specific for a setup using VGA resolution were taken from the VQEG Multimedia testplan.<sup>16</sup> In particular, all the sequences were displayed without upscaling on a 40 inch TV-Logic LMV401 reference LCD display operating at its native resolution of  $1920 \times 1080$  pixels at 60Hz. The VGA sequences were surrounded by a large gray border with a pixel value of  $Y=108$  in the YCbCr color space, corresponding to 0.25 of the maximum display brightness.

For the payout, the freely available AcrVQWin program of Acreo AB Sweden was used.<sup>17</sup> It implements an Absolute Category Rating with Hidden Reference (ACR-HR) protocol according to ITU-T P.910 with sequence randomization constraints as specified in the VQEG Multimedia testplan, e.g. it is forbidden to display two conditions with the same source content. The rating scale consists of five attributes ranging from “excellent” with a MOS score of 5 to “bad” with a MOS score of 1. In the interactive graphical user interface only the attributes are shown to the observers.

In total, 29 naïve observers participated in the test. They were tested for visual acuity using a Snellen chart and for color blindness using Ishihara plates. A test session lasted for about 40 minutes and was preceded by a training session. The analysis of their votes proofed that none of them needed to be rejected by the ITU-R BT.500 screening method or the VQEG Multimedia Test plan screening method.

### 4. EXPERIMENTAL RESULTS

Figure 2 displays the average Mean Opinion Scores (MOS) and their intervals of 95% confidence for each tested configuration. Additionally, we provide the results of a student t-test performed on the data in Table 2, in order to verify the statistical significance of the results we present.

The non-coded reference gets the highest score, above 4.5 which attests the high quality of the source videos. The non-damaged SVC videos get the second highest score, with only 0.5 less on the MOS scale than the reference. The damaged-AVC gets the worst score, with a value around 1.3 which is very low when compared to the other tested configurations. The error-concealment technique used on AVC therefore shows a strong impact on the perceived quality. The cause of this large quality drop might be the flickering effect produced by the concealment technique, which is quite disturbing for the viewer. Because of the buffer repetition-based error-concealment technique used by the decoder, about half a second of video gets repeatedly displayed with new content overlaid resulting from the extrapolation of the motion concealment. As it only concerns a part of the sequence, it creates both spatial and temporal discontinuities, which are judged to be very unpleasant.

Table 1. Video configurations used in the experiment and their associated labels.

Format number	Description / base layer format	Error concealment method
1	Reference	<i>none</i>
2	Non-damaged SVC	<i>none</i>
3	AVC	buffer repetition
4	15 fps / 120 kbps	upscale
5	15 fps / 120 kbps	patch
6	15 fps / 120 kbps	switch
7	30 fps / 120 kbps	upscale
8	30 fps / 120 kbps	patch
9	30 fps / 120 kbps	switch
10	15 fps / 200 kbps	upscale
11	15 fps / 200 kbps	patch
12	15 fps / 200 kbps	switch
13	30 fps / 200 kbps	upscale
14	30 fps / 200 kbps	patch
15	30 fps / 200 kbps	switch

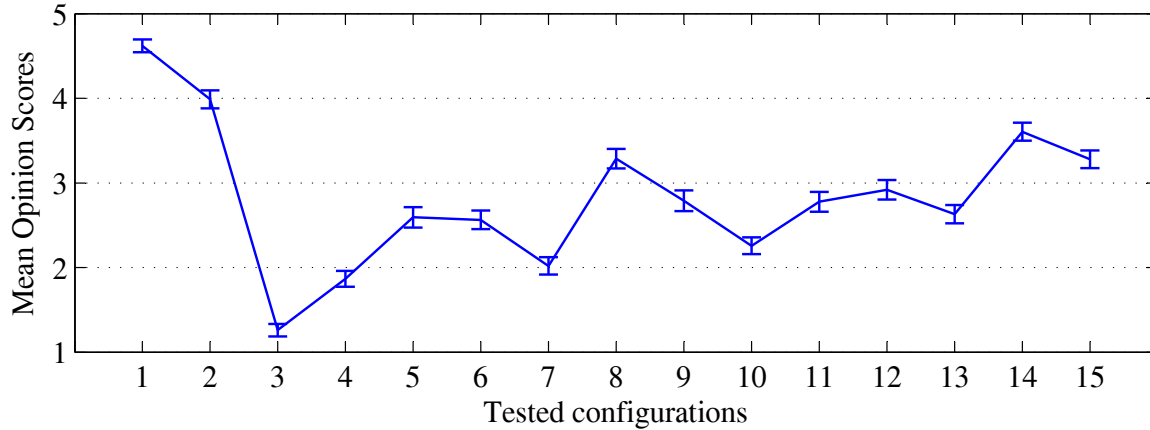


Figure 2. Average Mean Opinion Scores and intervals of confidence (95%) for the tested configurations. See Table 1 for the correspondence between numbers and configurations.

The four scenarios including the SVC-based concealment techniques show the same tendency, with MOS values globally increasing when the temporal resolution and bitrate of the base layer increase. The upscaled version obtains the lowest score in each scenario. It is to notice that with equal bitrate, the version with 30 fps is preferred to the version with 15 fps. The frames in the 15 fps version get a better individual quality as the bitrate is dispatched among half the number of frames from the 30 fps version. Still, the viewers seem to prefer a fluid motion to higher quality on individual images. This is an interesting result in itself illustrating that temporal artifacts are judged as more unpleasant than spatial artifacts. This will be confirmed in the rest of this analysis. The patched and the switched versions achieve significant improvement of the visual quality when compared to the upscaled base layer in each scenario. Still, the relative behavior of the two concealment methods differs depending on the temporal frequency of the base layer.

When the temporal frequency of the base layer and the enhancement layer are identical, the patched version gets a significantly higher score than the switched version. The pixel-level concealment only affects the damaged areas in each frame, keeping the intact areas in high quality. The spatial discontinuities at the limit between the upscaled and the intact areas are preferred to the homogeneous but low-quality frame. This shows that the viewers tolerate spatial discontinuities when the global quality of the frames is higher. The impact of temporal discontinuities on the viewers' opinion is quite different. When the temporal frequency in the base layer is lower than in the enhancement layer, the patched and the switched methods get equivalent scores. This assessment is verified by the student t-test in Table 2. The quality increase with the switched version is actually quite constant from one scenario to another, so we can notice a loss in efficiency from the patched version when the temporal frequency is lower in the base layer. When upscaling the base layer from 15 to 30 fps, each frame is duplicated in order to double the number of fps. Therefore, using one frame from the base layer to conceal the errors in the enhancement layer causes spatio-temporal discontinuities in one frame out of two. This phenomenon is particularly visible when motion is high around the junction between the concealed and the non-concealed areas, such as demonstrated in Figure 3.

It is interesting to note that even the upscaled version of the lowest quality scenario (*i.e.*: 15 fps / 120 kbps) is preferred to the damaged AVC. This point shows that spatial and temporal consistency has a strong impact on the global perceived quality. Even a video sequence with low quality can reach a higher opinion score than a high quality video with badly-concealed errors. Furthermore, Table 2 shows that the version which was upscaled from the base layer at 30 fps / 200 kbps is statistically equivalent in terms of visual quality to the patched and switch versions in the scenario at 15 fps / 120 kbps. This confirms that upscaling a good quality base layer can reach the same quality level as versions concealed using a lower-quality base layer. This means that transmitting a QVGA base layer at 30 fps / 200 kbps and upscaling it to VGA produces the same quality as the VGA enhancement layer encoded at 600 kbps when transmission errors call for concealment. Moreover, it can be argued that the bitrate saved from using only 200 kbps could be used to add redundancy in the bitstream and perform error protection.

The patched and switched versions in the 30 fps / 120 kbps scenario reach better or equal quality when compared to the

15 fps / 200 kbps scenario. This is again in favor of the viewers preference for the spatial distortions rather than temporal distortions. In addition, the patched version in the 30 fps / 120 kbps scenario gets a similar score as the switched version in 30 fps / 200 kbps. The difference in bitrate between the two scenarios is indeed compensated by the superiority of the pixel-level concealment over the frame-level concealment. In this case, the spatial discontinuity produced at the border between the upscaled and the intact areas is preferred to a homogeneous quality on the whole frame. To explain this, one should consider that the quality change is more noticeable when it is applied on the whole frame than when only a limited area of the frame is affected. A temporal discontinuity is created in both cases, but its effects are more visible when the whole frame changes at once.

Table 2. Comparison of the experimental results from the tested configurations through a student t-test. If the cell at row  $i$  and column  $j$  contains a ‘.’ symbol, configurations  $i$  and  $j$  are not statistically different in terms of MOS. If the cell contains a ‘↑’ symbol, configuration  $i$  has a significantly higher score than configuration  $j$ . If the cell contains a ‘↓’ symbol, configuration  $i$  has a significantly lower score than configuration  $j$ .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
2	↓		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
3	↓	↑		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
4	↓	↑	↑		.	↑	↓	↓	↑	↓	↓	.	↓	↓	↓
5	↓	↑	↑	.		↑	↓	↓	↑	↓	↓	.	↓	↓	↓
6	↓	↑	↑	↓	↓		↓	↓	↓	↓	↓	↓	↓	↓	↓
7	↓	↑	↑	↑	↑	↑		↑	↑	↑	↑	↑	↓	.	↓
8	↓	↑	↑	↑	↑	↑	↓		↑	.	.	.	↓	↓	↓
9	↓	↑	↑	↓	↓	↑	↓	↓		↓	↓	↓	↓	↓	↓
10	↓	↑	↑	↑	↑	↑	↓	.	↑		.	.	↓	↓	↓
11	↓	↑	↑	↑	↑	↑	↓	.	↑	.		↑	↓	↓	↓
12	↓	↑	↑	.	.	↑	↓	.	↑	.	↓		↓	↓	↓
13	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		↑	↓
14	↓	↑	↑	↑	↑	↑	.	↑	↑	↑	↑	↑	↓		↓
15	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	

The presented subjective experiment answers several interesting questions. First, we showed that using SVC as an error concealment technique can produce significantly higher quality than using a classical error concealment technique on a single-layer AVC stream. It is well known that SVC introduces a loss in coding efficiency as encoding several layers of the same content produces redundancy. This loss has been evaluated in terms of visual quality by several studies and proved to produce a noticeable yet limited impact on the viewers’ opinion.<sup>9</sup> According to our tests, the differences between SVC and AVC are not as significant in a context such as lossy transmission, where the errors produced by data-loss have a stronger impact on the quality. Therefore, SVC can be advantageous when compared to AVC as it provides intermediate versions of the video that can be exploited for error concealment in the enhancement layers.

## 5. CONCLUSION

In this paper we studied the performances of Scalable Video Coding to perform error-concealment in terms of visual quality. The scalable streams contain two layers, among which only the highest is subject to loss. The lower layer is then upscaled to replace the lost or corrupted data. Two techniques were proposed based on this principle, working at frame and pixel level. We compared them to a classical single-layer error-concealment technique based on MPEG-4 AVC/H.264 and to a scenario where only the base layer is transmitted and upscaled. The results of our subjective experiment show that the two SVC-based error-concealment techniques perform better in terms of visual quality than the AVC-based technique. Moreover, we identified the impact of the resolution and quality of the base layer on the visual quality of the concealment.



If the two layers are at the same temporal frequency, a pixel-level technique performs well, whereas a frame-level technique is better if the two layers do not have the same temporal frequency. In addition, the bitrate used to encode the base layer is of great importance. We show that a good quality base layer combined with upsampling can perform as good as two layers using the presented concealment techniques when the base layer is in low quality. Therefore the advantage of SVC concerning error-concealment techniques can be significant, as long as the base layer is encoded with sufficient quality.

## 6. ACKNOWLEDGEMENTS

This work was partly conducted for the SVC4QoE project funded by the French Direction Générale de la Compétitivité, de l'Industrie et des Services (DGCIS). This project is aimed at evaluating the performance of scalable video coding for the optimization of the quality of experience in a mobile video broadcasting context.

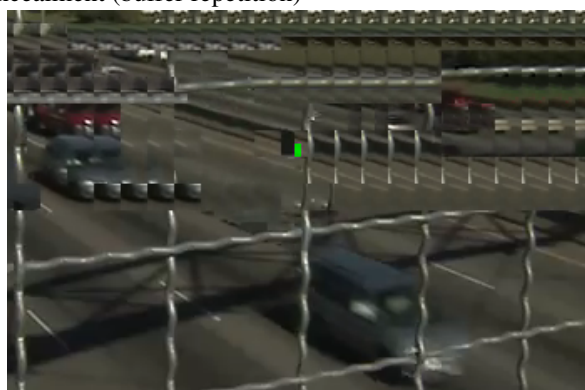
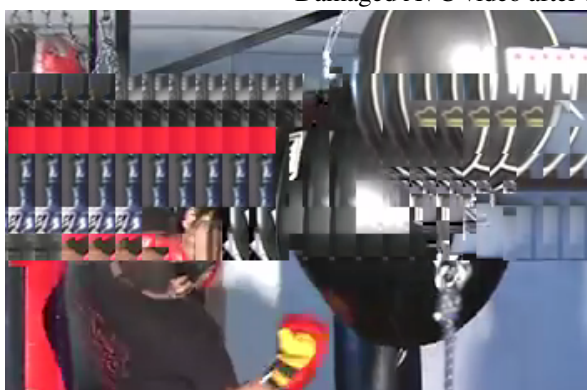
## REFERENCES

- [1] Iain E. Richardson, [*H.264 and MPEG-4 Video Compression*], John Wiley and Sons (2003).
- [2] Y. Wang and Q.-F. Zhu, "Error Control and Concealment for Video Communication: A Review," in [*Proceedings of the IEEE*], **86**(5), pp. 974–997 (1998).
- [3] KUNG Wei-Ying and KIM Chang-Su and KUO C.-C. Jay, "Spatial and temporal error concealment techniques for video transmission over noisy channels," *IEEE transactions on circuits and systems for video technology* **16**(7), 789–802 (2006).
- [4] Olivia Nemethova and Ameen Al-Moghrabi and Markus Rupp, "Flexible Error Concealment for H.264 Based on Directional Interpolation," *WirelessCom Conference on Wireless Networks, Communications and Mobile Computing* **17**(2), 425–450 (2006).
- [5] Donghyung Kim and Jongho Kim and Jechang Jeong, [*Temporal Error Concealment Based on Optical Flow in the H.264/AVC Standard in Advanced Concepts for Intelligent Vision Systems*], vol. 4179 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg (2006).
- [6] D. Agrafiotis and D.R. Bull and T-K Chiew and P. Ferre and A.R. Nix, "Enhanced error concealment for video transmission over WLANs," *Int'l Workshop on Image Analysis for Multimedia Interactive Services* **1** (2005).
- [7] Werner Robitza and Shelley Buchinger and Patrik Hummelbrunner and Helmut Hlavacs, "Acceptance of Mobile TV Channel Switching Delays," in [*Quality of Multimedia Experience Workshop (QoMEX)*], (2010).
- [8] Julien Reichel and Heiko Schwarz and Mathias Wien, "Joint Scalable Video Model JSVM-11, doc. JVT-X202," tech. rep., Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (2007).
- [9] ISO/IEC JTC1/SC29/WG11 MPEG2007/N9189, "SVC Verification Test Plan, Version 1," tech. rep., Joint Video Team (2007).
- [10] Amir Naghdinezhad and Mahmoud Reza Hashemi and Omid Fatemi, [*Unequal Error Protection for the Scalable Extension of H.264/AVC Using Genetic Algorithm*], vol. 6 of *Communications in Computer and Information Science*, Springer Berlin / Heidelberg (2008).
- [11] Bakker Dirk and Cromboom Dennis and Dams Tim and Munteanu Adrian and Barbarien Joeri, "Priority-based Error Protection for the Scalable Extension of H.264/SVC," *SPIE Optical and digital image processing* **7000**, 70000H–70000H–11 (2008).
- [12] Jong-Tzy Wang and Pao-Chi Chang, "Error prevention and concealment for scalable video coding with dual-priority transmission," *Journal of Visual Communications and Image Retrieval* **14** (2003).
- [13] Qirong Ma and Feng Wu and Jian Lou and Ming-Ting Sun, "Frame loss error concealment for spatial scalability using hallucination," *IEEE Packet Video Workshop* **1** (2009).
- [14] Joint Video Team, "JSVM Reference Software, Version 9.18." [http://ip.hhi.de/imagecom\\_G1/savce/downloads/](http://ip.hhi.de/imagecom_G1/savce/downloads/) (2009).
- [15] Yohann Pitrey and Marcus Barkowsky and Patrick Le Callet and Romuald Pepion, "Subjective Quality Evaluation of H.264 High-Definition Video Coding versus Spatial Up-Scaling and Interlacing," in [*Workshop on Quality of Experience for Multimedia Content Sharing (QoEMCS), EuroITV Conference (ACM)*], (2010).
- [16] David Hands and Kjell Brunnström, [*Multimedia Group Test Plan Draft Version 1.21*], Video Quality Experts Group (VQEG) (2008).
- [17] Kjell Brunnström, "AcrVQWin runs subjective experiments for video quality in Windows." [http://www.acreo.se/templates/Page\\_\\_\\_\\_7172.aspx](http://www.acreo.se/templates/Page____7172.aspx) (2005).

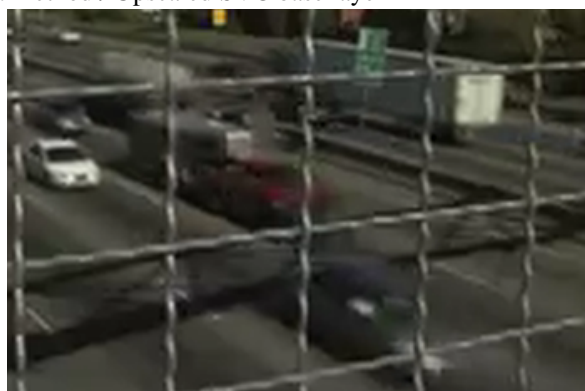
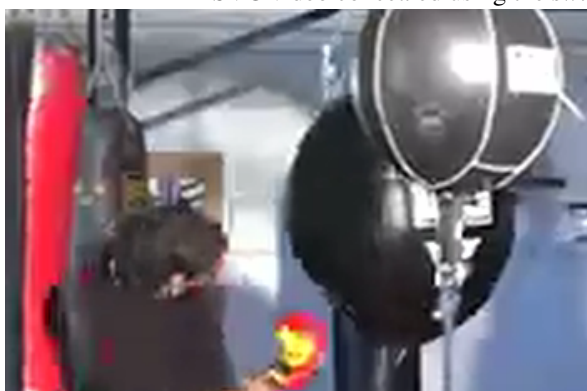
Non-coded reference



Damaged AVC video after concealment (buffer repetition)



SVC video concealed using the *switch* method / Upscaled SVC base layer



SVC video concealed using the *patch* method



Figure 3. Examples of the distortions produced by the different error-concealment techniques.