Video Content Adaptation Based on SVC and Associated RTP Packet Loss Detection and Signaling

Daniele Renzi\textsuperscript{1}, Peter Amon\textsuperscript{2} and Stefano Battista\textsuperscript{3}
\textsuperscript{1,3}bSoft ltd
\textsuperscript{2}Siemens Corporate Technology, Information and Communications
daniele@bsoft.info %p.amon@siemens.com bautz@bsoft.info

Abstract

The development of Scalable Video Coding (SVC) has been directed to serve a wide range of terminals over heterogeneous networks with the same encoded bit stream. SVC provides an elegant solution to adapt the video content, as it allows terminal and/or gateways accessing only a sub-part of the stream, without affecting the semantics of the source video signal. This paper describes adaptation mechanisms based on SVC, focusing on the implementation in the context of the IST project ENTHRONE. A particular emphasis is placed on adaptation techniques based on SVC NAL unit header extension, and on innovative RTP packet loss handling mechanisms, specifically conceived to handle losses in scenarios where SVC to SVC and SVC to H.264/AVC adaptation is performed and layered multicast cannot be used.

1 Introduction

The recent advances in video coding techniques led to the new standard H.264/MPEG-4 Scalable Video Coding (SVC), which has been conceived as an amendment of the widely adopted H.264/MPEG-4 AVC. SVC introduces innovative scalability features, enabling encoders to produce a single stream with multiple temporal, spatial and quality layers, while preserving high compression efficiency. SVC decoders or media aware gateways can easily access all or part of the entire compressed bit stream. That allows adapting the video at bit stream level. The source SVC stream can be cut in so-called operation points, i.e., sets of layers reconstructing a specific representation of the video content in temporal, spatial and quality dimensions. That is made possible by the hierarchical structure of the SVC stream.

In an End-to-End (E2E) framework based on Quality of Service (QoS), like the one devised and developed in the IST project ENTHRONE, such an adaptation can be automatically demanded by the terminal, by providing the adaptation modules with the information concerning the QoS metrics and the terminal resources.

The operation points can be transmitted using RTP as transport protocol in IP packet-based networks; e.g., in a layered multicast mode, where each scalable layer is transported in its own multicast group, or in a unicast mode, where a single RTP session carries multiple layers. The layered multicast is very attractive, especially in enterprise networks or research environments, because of its ease of implementation (even though it requires special mechanisms to synchronize the packets carried in different sessions). However, it might cause problems in scenarios where clients are protected by firewalls, as it requires opening different ports to multicast addresses. That is why we investigated methods to correctly handle RTP transmission modes different than layered multicast, and particularly to detect and signal packet losses.

2 Adaptation techniques for SVC

Various adaptation techniques to adapt an SVC stream have been investigated. E.g., adaptation techniques that use the NAL unit headers in SVC streams or media agnostic techniques like gBSD can be used [9].

2.1 NAL unit header

The high-level syntax of SVC obeys similar design criteria as those of H.264/AVC. Parameter sets, containing information for more than one picture, are normally transmitted out-of-band, using a reliable transmission protocol (e.g., TCP), but could also be repeated in-band, e.g., for broadcast applications.
The video data is transmitted in Network Abstraction Layer (NAL) units. The first byte of the SVC NAL unit header extension (see Figure 1) contains the syntax element priority_id and also indicates whether the NAL unit belongs to an IDR access unit (idr_flag). The second and third byte provide information on the scalability dimensions, by the fields dependency_id (Did), temporal_id (Tid) and quality_id (Qid), and also on the possibility to discard NAL units from the decoding of layers with higher Did (discardable_flag).

As stated in [4] and [5], the NAL unit header co-serves as payload header for an RTP packet in some cases: that arranges for the scalability information to be available at transport level. A further criterion is the backward compatibility to H.264/AVC: a legacy H.264/AVC decoder regards SVC NAL units as regular NAL units with unknown NAL unit types, and discards them while still being able to decode the base layer (or even the entire stream when only temporal scalability is used, as described in section 4.1).

2.2 Priority-based adaptation

Based on the information in the NAL unit header, a simple priority based adaptation can be performed. The syntax element priority_id gives information on the importance of a NAL unit. Since priority_id is a one-dimensional field, the three-dimensional vector containing the scalability dimensions (Did, Qid and Tid) have to be mapped into one dimension, as depicted in Figure 2.

![Figure 2. Mapping of scalability dimensions (DTQ) to layers (priority_id)](image)

This mapping already contains a predefined adaptation path, e.g., the decision, whether temporal or spatial resolution should be sacrificed first. Adaptation is done by setting a threshold for priority_id and skipping all NAL units having a higher value than the threshold.

2.3 DTQ-based adaptation

DTQ denotes the three scalability dimensions Did, Tid and Qid, which are set for every NAL unit. Adaptation based on DTQ allows for more flexibility, e.g., the adaptation path can still be determined at the time of adaptation, e.g., in case different terminals having different display size, receive the same content, but experience different adaptation decisions due to their capabilities. Adaptation is done using three threshold values for the respective syntax elements.

3 RTP packet loss handling

3.1 Typical use cases

As described in [7], the most common network distribution models pertaining to the scalable video are:

- multicast/broadcast to terminals with heterogeneous connectivity;
- multicast/broadcast, with a Media Aware Network Element (MANE) to aggregate different sessions;
- unicast, where the server aggregates in one RTP session possibly more than one layer.

In the first scenario, commonly called layered multicast (Figure 3), each layer is transported in its own IP multicast group, and terminals subscribe to layers utilizing the IP multicast mechanisms. This mode has the advantage of simplicity for, e.g., detecting packet losses, thanks to the one-to-one correspondence between layers and RTP streams.

![Figure 3. First network distribution scenario: layered multicast](image)

However, in this scenario the client devices need to open different ports to multicast addresses to get the best possible quality, which might cause problems with firewalls. Another practical hurdle is the sparse application of multicast in today’s networks.

For the previous reasons, other network distribution models need to be used, like the ones in the second and third scenario, even though they rely on more complex
adaptation methods. In such scenarios a MANE performs either trimming or aggregation of scalable layers, in order to tailor the terminal needs with a single RTP session. E.g., in the third scenario, depicted in Figure 4, the server aggregates multiple SVC layers in the same RTP stream; a network adapter forwards only the packets carrying the required layers; if there is a packet loss, the adapter should be somehow able to map the lost packet with an SVC layer, and to signal the loss to the terminal, by inserting a gap in the RTP sequence number of the adapted stream. But, if along with the packet, also the DTQ information has got lost, then a sequence number gap cannot be mapped to a scalable layer.

The second solution we propose is based on the use of Single Time Aggregation Packets (STAPs) and sub-sequence information SEI (SSEI) messages. The NAL unit 1 in Figure 7 of [4] is a sub-sequence information SEI message (section D.1.11 of [1]), where the sub_seq_layer_num field indicates the layer, while the sub_seq_frame_num field has the same semantics of the field frame_num in the slice header, but with reference to a sub-sequence rather than to the entire sequence. We propose to check gaps on sub_seq_frame_num to detect frame losses, specifically for each layer. In order to provide a higher granularity of loss detection than at frame level, e.g., to handle losses at slice level, we propose to enable the adapter access the slice header information, in order to parse the slice_num field, and use it in addition to the sub_seq_frame_num to uniquely identify a slice in a sub-sequence (layer).

4 Implementation of concepts

Two examples of adaptation scenarios have been tested to check the effectiveness of the devised SVC adaptation (section 2) and packet loss detection and signaling mechanisms (section 3). Four modules have been implemented: a real-time SVC Encoder, an SVC2AVC Network Adapter, an SVC Network Adapter, and a Media Player with AVC/SVC decoding capabilities. All tests have been performed in standard IEEE 802.3 and 802.11 LANs, where packet losses have been simulated forcing random losses in the SVC Encoder module.

4.1 SVC to H.264/AVC Adaptation with only temporal scalability

According to the SVC standard, it is possible to generate a temporal scalable SVC stream based on H.264/AVC compliant coded slice NAL units, preceded by prefix NAL units, describing the SVC layer this NAL unit belongs to. The SVC stream can be decoded by both SVC clients and H.264/AVC clients; the latter would drop the prefix NAL units.

The SVC2AVC Network Adapter makes use of the information in the prefix NAL units for making adaptation decisions. This process is shown in Figure
6: if the SVC2AVC Network Adapter has been configured to forward the lowest two layers, corresponding to the frame rates of 7.5 fps and 15 fps, the packets related to the third temporal layer (30 fps) will be dropped, along with all prefix SVC NAL units. The result is an H.264/AVC stream with 15fps.

If the source SVC stream is conveyed over a single RTP session, one of the methods specified in section 3 of this paper can be used to signal packet losses to the Media Player. E.g., when the first method is used, the SVC Adapter processes the information contained in the RTP header extension to infer which SVC layer each lost packet was related to, and inserts a sequence number gap in the related out-going stream.

4.2 SVC to SVC adaptation

The SVC Network Adapter provides for a solution to extract multiple stream versions from a single SVC stream, to modify the description of the stream both in-band and out-of-band, to duplicate the streams when necessary, to provide a valid RTP stream after adaptation, and to detect and signal packet losses in each SVC layer. By accessing the DTQ information, the SVC Network Adapter figures out, whether the packet needs to be forwarded. In general, the sequence number field of the RTP packet, as well as the other data fields described in section 3, need to be conveniently translated, to hide any packet dropping in the adapter from the client. Other RTP fields, such as the packet timestamps, are preserved.

5 Conclusions

In this paper we described some advanced video content adaptation mechanisms based on Scalable Video Coding. The focus is on the end-to-end QoS based adaptation framework, devised in the context of the IST project ENTHRONE. A particular emphasis is placed on adaptation techniques based on the SVC NAL unit header extension, and on innovative packet loss handling mechanisms, for detecting packet losses in scenarios where SVC stream adaptation is performed, and the layered multicast network model cannot be used. That is why we have investigated methods to handle network distribution modes, where single RTP sessions are used for carrying multiple SVC layers, and especially to detect and signal packet losses, using standards-compliant solutions.

Acknowledgement

This work has been co-funded by the IST European project ENTHRONE under contract IST-2006-038463.

References