Transmission of Scalable Video in Computer Networks

Jânio M. Monteiro
University of Algarve and IST/INESC-ID, Portugal

Carlos T. Calafate
Technical University of Valencia, Spain

Mário S. Nunes
IST/INESC-ID, Portugal

INTRODUCTION

The Internet as a video distribution medium has seen a tremendous growth in recent years with the advent of new broadband access networks and the widespread adoption of media terminals supporting video reception and storage. This growth of Internet video transmission resulted from the advances in video encoding solutions and the increase in the bandwidth of terminals. However, it has also placed new challenges in the developments of video standards, due to the heterogeneous characteristics of current terminals and of the wired and wireless distribution networks.

As the terminal capabilities increase in terms of display definition, processing power, and bandwidth, users tend to expect higher qualities from the received video streams. Additionally, as different types of terminals will likely coexist in the same network, it will be necessary to adapt the content transmitted according to the receiving terminal. Instead of re-encoding (or transcoding) the bitstream, which requires a high computational power on intermediate nodes, rate adaptation would preferably be done by extracting parts of the original bitstream.

In terms of network scalability, encoding and transmitting the same video sequence in a large-scale live video distribution system is a challenge, which may only rely on point-to-multipoint transmissions like IP multicast or broadcast. The traditional solution for point-to-multipoint video transmission in heterogeneous networks and with terminals with very different capabilities relies on a process usually called simulcast or replicated streams transmission. In this process, a discrete number of independent video streams are encoded and distributed through the multicast or broadcast path. Terminals request and decode the video stream that better fits their characteristics and communication rate, switching between video streams according to bandwidth variations. However, the major drawback of *simulcast* is that much of the information carried in one stream is also carried in adjacent streams, and therefore the total rate required for a video transmission is much higher than the rate of a single stream. In these scenarios it would be preferable to encode different levels of quality—one base layer quality and one or more enhancement layers—which could be used to increase the quality of the base layer. Accordingly, terminals with lower bandwidth or computational power could request the reception of lower layers, and terminals with higher capabilities could request additional enhancement layers.

In this article we analyze scalable video transmission, from the perspective of video coding standards and the necessary developments in protocols that support media distribution in current and future network architectures. In the next section we start by describing the first contributions to this topic and following developments in related video coding standards. We then describe the structure of a scalable video bitstream, taking the novel H.264/SVC standard as reference, and we further proceed with an analysis of the protocols that can be used for the description, signaling, and transport of scalable video. We describe different network scenarios and examples where scalable video offers significant advantages, before moving on to some remarks on future trends in this area, discussing those mechanisms that must be associated with SVC techniques to achieve an efficient and robust transmission system, and concluding the article.

BACKGROUND

The use of layered video transmission in IP multicast was originally proposed by Deering (1993), who suggested the transport of different video layers in different multicast groups. With this solution the encoder would produce a set of interdependent layers (one base layer and one or more enhancement layers), and the receiver, starting with a base layer, could adapt his quality by joining and leaving multicast trees, each one carrying a different quality layer.

Deering's proposal was followed by several protocols like: receiver driven layered multicast (RLM) protocol (Mc-

Canne, Jacobson, & Vetterli, 1996), layered video multicast with retransmission (LVMR) protocol (Li, Paul, Pancha, & Ammar, 1997; Li, Paul, & Ammar, 1998), and ThinStreams (Wu, Sharma, & Smith, 1997).

The advantages of layered video multicast were confirmed by Kim and Ammar (2001) for scenarios where receivers are distributed in the same domain, with multiple streams sharing the same bottleneck link, as occurs in many video distribution scenarios.

In terms of video coding, layered video transmission requires a layered encoding of video, a process usually referred to as scalable video coding (SVC). Video coding standards like *ITU-T Recommendation H.263* from the International Telecommunication Union-Telecommunication (ITU-T, 2000) and *MPEG-2 Video* from the ITU-T and the International Organization for Standardization/International Electrotechnical Commission (ITU-T & ISO/IEC, 1994) included several tools that supported the most important scalability options. However, none of these scalable extensions was broadly implemented since they imply a loss in coding efficiency and also a significant increase in terms of decoder complexity.

In January 2005, the Joint Video Team (JVT) from ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) started developing a scalable video coding extension for the H.264 Advanced Video Coding standard (ITU-T & ISO/IEC, 2003), known as H.264 Scalable Video Coding (ITU-T & ISO/IEC, 2007). The H.264 SVC augments the original encoder's functionality to generate several layers of quality. Enhancement layers may enhance the content represented by lower layers in terms of temporal resolution (i.e., the frame rate), spatial resolution (i.e., image size), and the quality—specified as signal-to-noise ratio resolution (i.e., SNR).

By using the H.264 SVC, different levels of quality could be transmitted efficiently over both wired and wireless networks, allowing seamless adaptation to available bandwidth and to the characteristics of the terminal. However, the transport of SVC presents many challenges which must be considered in order to take full advantage of its potential.

CODING AND TRANSMISSION OF SCALABLE VIDEO

The most adequate technique for efficiently transmitting scalable video is highly dependent on the video encoding technology itself. Hence, for this article we have used the scalable video extensions to the H.264 standard as reference since these represent the most advanced technology currently available in this area.

The H.264 Advanced Video Coding (AVC) standard is currently emerging as the preferred solution for video services in third-generation (3G) mobile networks, which include packet-switched streaming services, messaging services, conversational services, and multimedia broadcast/multicast services (MBMS) (3GPP TS 26.346, 2005). It will also be used for mobile TV distribution to handheld devices (DVB-H) (ETSI TR 102 377, 2005).

SVC Bitstream Structure

The scalable extension of H.264/AVC includes several layers of quality. Relative to the base layer of an SVC bitstream, and for compatibility purposes, the JVT decided to make it compatible with the H.264/AVC profile.

The SVC bitstream may be composed of multiple spatial, temporal, and SNR layers of combined scalability. Temporal scalability is a technique that allows supporting multiple frame rates. In SVC, temporal scalability is usually implemented by using hierarchical B-pictures.

Quality (or SNR) scalability relies on both *coarse-grain quality scalable* (CGS) and *medium-grain quality scalable* (MGS) coding. While CGS encodes the transform coefficients in a nonscalable way, in MGS, which is a variation of CGS, fragments of transform coefficients are split into several network adaptation layer (NAL) units, enabling a more graceful degradation of quality when these units are discarded for rate adaptation purposes. The JVT also considered the possibility of including another form of scalability named *fine-grain scalability* (FGS), which was proposed in MPEG-4 Visual. FGS arranges the transform coefficients as an embedded bitstream, enabling truncation of these NAL units at any arbitrary point. However, in the first specification of SVC (Phase I) (ITU-T & ISO/IEC, 2007), FGS layers were not supported.

Spatial scalability provides support for several display resolutions (e.g., 4CIF, CIF, or QCIF) and is implemented by decomposing the original video into a spatial pyramid.

These spatial, temporal, and SNR layers are identified using a triple identification (ID), consisting of the *dependency ID* identifying the spatial definition, the *temporal ID,* and the *quality* (i.e., SNR) *ID,* which is referred as tuple (D,T,Q). For instance, a base layer NAL unit of the lowest temporal resolution and SNR scalability should be identified as (D,T,Q)=(0,0,0). Accordingly the network adaptation layer structure of H.264/AVC has been extended to include these three IDs. These three layers may be represented using a three-dimensional graph, such as the one in Figure 1.

SVC layers can be highly interdependent from each other, which means that the loss of an NAL unit of a certain layer may cause a severe reduction of quality or even prevent the correct decoding of other layers. This implies that lower layers should be protected from bit errors or packet losses.

Related Content