Title: Scalable Video Coding in Networked Environment

Authors: Naeem Ramzan, Tomas Piatrik and Ebroul Izquierdo

Affiliation: Queen Mary University of London

Abstract:

State-of-the-art coders have been optimised over years according to the needs of the broadcasting industry. There are however key applications of coding technology (transmission, surveillance, etc.) whose challenges and requirements substantially differ from broadcasting. In this document an efficient approach jointly optimises the bit allocation between the wavelet-based scalable video coder and a forward error correction codes. Bit-rate optimization and adaptation is achieved by exploiting the scalability properties of the employed codec. In addition, temporal segments containing events relevant to surveillance application are encoded using high spatio-temporal resolution and quality while the portions irrelevant from the surveillance standpoint are encoded at low spatio-temporal resolution and/or quality.

1   INTRODUCTION

Advances in video coding technology along with the rapid development in network infrastructure, storage capacity, and computing power are enabling an increasing number of video applications. In advanced communication systems, users may access and interact with multimedia content from different terminals and via different networks such as: video transmission and access over the Internet and handheld devices like mobile telephones and Personnel Digital Assistants (PDAs); multimedia broadcasting; and video services over wireless channels. In the scenario depicted in Fig.1, the video server requires video contents of different fidelities, such as high quality material for storage and future editing and lower bit-rate content for distribution. In traditional video communications over heterogeneous channels, the video is usually processed offline. Compression and storage are tailored to the targeted application according to the available bandwidth and potential end-user receiver or display characteristics. However, this process requires either transcoding of compressed content or storage of several different versions of the encoded video.

None of these alternatives represent an efficient solution. Furthermore, video delivery over error-prone heterogeneous channels meets additional challenges such as bit errors, packet loss and error propagation in both spatial and temporal domains. This has a significant impact on the decoded video quality after transmission and in some cases renders useless the received content. Consequently, concepts such as scalability, robustness and error resilience need to be re-assessed to allow for both efficiency and adaptability according to individual transmission bandwidth, user preferences and terminals.

Scalable Video Coding (SVC) promises to partially solve this problem by "encoding once and decoding many". SVC enables content organization in a hierarchical manner to allow decoding and interactivity at several granularity levels. That is, scalable coded bitstreams can efficiently adapt to the application requirements. The scenario shown in Fig. 1 can truncate the SVC encoded bitstream at different points and decode it. The truncated bitstream can be further truncated to

some lower resolution, frame rate or quality. Thus, it is important to tackle the problems inherent to the diversity of bandwidth in heterogeneous networks and in order to provide an improved quality of services. Wavelet-based scalable video coding (W-SVC) provides a natural solution for error-prone transmissions with a truncateable bitstream.
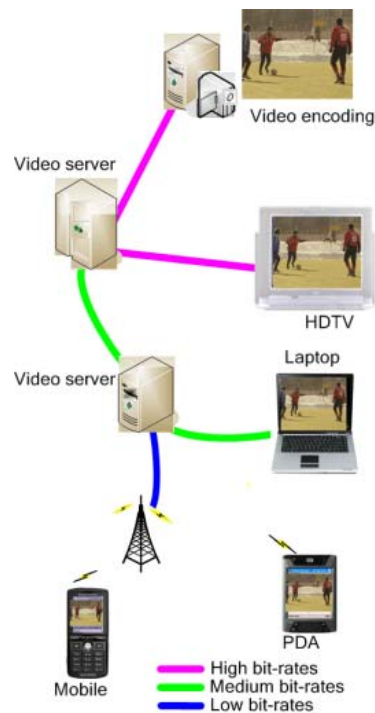

Fig. 1: Scalable video transmission: one video bitstream serves to different clients.

1.1 Main modules of Wavelet-based Scalable Video Coding

The W-SVC consists of three main modules:

Encoder: the input video is encoded by the W-SVC encoder, producing the bitstream of the maximum required quality which, if the application requires, can be up to quasi-lossless (resulting in imperceptible quality loss).

Extractor: the main aim of the W-SVC extractor is to truncate the scalable bitstream according to the scaling requirements and to generate the adapted bitstream and its description. The adapted bitstream is also scalable and can be fed back into the extractor for another stage of adaptation. This scenario corresponds to the situation of multiple-point adaptation where the adapted bitstream is sent to the next network node and is adapted by another extractor.

Decoder: W-SVC decoder is capable of decoding any adapted bitstream by W-SVC extractor or encoded by W-SVC encoder.

2 APPLICATIONS

There are number of multimedia applications where SVC helps to reduce the complexity and provides natural solution of the problem.

2.1 Event-based Scalable Coding of Surveillance Video

The basic principle behind the event-based scalable coding is to use different encoding settings for time segments representing different events in a surveillance video. For this purpose we classify temporal segments of the surveillance video into two types:

- temporal segments representing an essentially static scene (e.g. only random environmental motion is present – swaying trees, flags moving on the wind, etc.)

- temporal segments containing non-randomised motion activity (e.g. a vehicle is moving in a forbidden area).

To enable this classification, background subtraction and tracking module is used as Video Content Analysis (VCA). The output of this module defines parameters of compressed video. For actual encoding the W-SVC is employed.

VCA: Video background subtraction module based on Gaussian mixture model is used as VCA. This module is able to deal robustly with light changes, bimodal background like swaying trees and introduction or removal of objects from the scene. Value of each pixel is matched against weighted Gaussians of mixture. Pixels whose value is not within 2.5 standard deviations of the Gaussians representing background are declared as foreground.
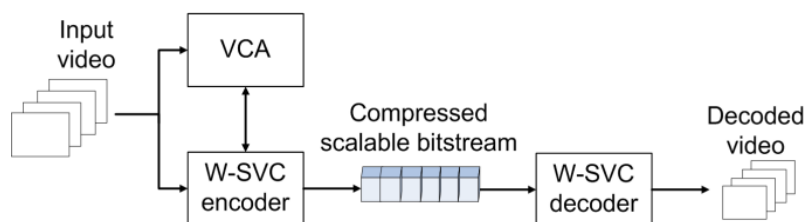

Fig. 2: Event-based scalable video encoding framework.

At each time instance the W-SVC encoder communicates with the VCA module (background subtraction and tracking). When the input video is essentially static the output of the background subtraction does not contain foreground regions. This can be used to signal to the W-SVC encoder to adapt captured video at low spatio-temporal resolution and quality, as shown in Fig. 2.

This allows, for instance, storing and/or transmitting the portions of the video containing long, boring, static scenes using low quality frame-rate and spatial resolution. On the other hand, when some activity in the captured video is detected, the VCA module notifies the W-SVC encoder to automatically switch its output to a desired much higher spatio-temporal resolution and quality video. Therefore, decoding and use of the video at different spatio-temporal resolutions and qualities corresponding to different events is achieved from a single bitstream, without multicasting or complex transcoding. Moreover, additional optional adaptation to lower bit-rate is also possible without re-encoding the video. This is, for instance, very useful in cases where video has to be delivered to a device with a low display capability. Using this approach, the bit-rate of video portions that are of low interest is kept low while the bit-rate of important parts is kept high. Since in many realistic applications it can be expected that large portions of the captured video have no events of interest, the proposed model leads to significant reduction of resources without jeopardizing the quality of any off-line event detection module that may be present at the decoder.

2.2 Joint Source Channel Coding for Scalable Video

The progressive nature of scalable bitstream facilitates to apply the Unequal Error Protection (UEP) in the JSCC. Normally, the JSCC consists of two main modules: scalable video encoding and channel encoding. At the sender side, the input video is coded using the W-SVC coder. The resulting bitstream is adapted according to channel capacities. The adaptation can also be driven by terminal or user requirements when this information is available. The adapted video stream is then passed to the channel encoding module where it is protected against channel errors. The channel coding module performs paketization, addition of CRC bits, and the channel error correction coding using a rate-distortion (R-D) optimization. After modulation, the video signal is transmitted over a lossy channel. At the receiver side, the inverse process is carried out.