H.264/AVC Scalable Video Coding (SVC)

Composed by Shevach Riabtsev

slavah264@gmail.com

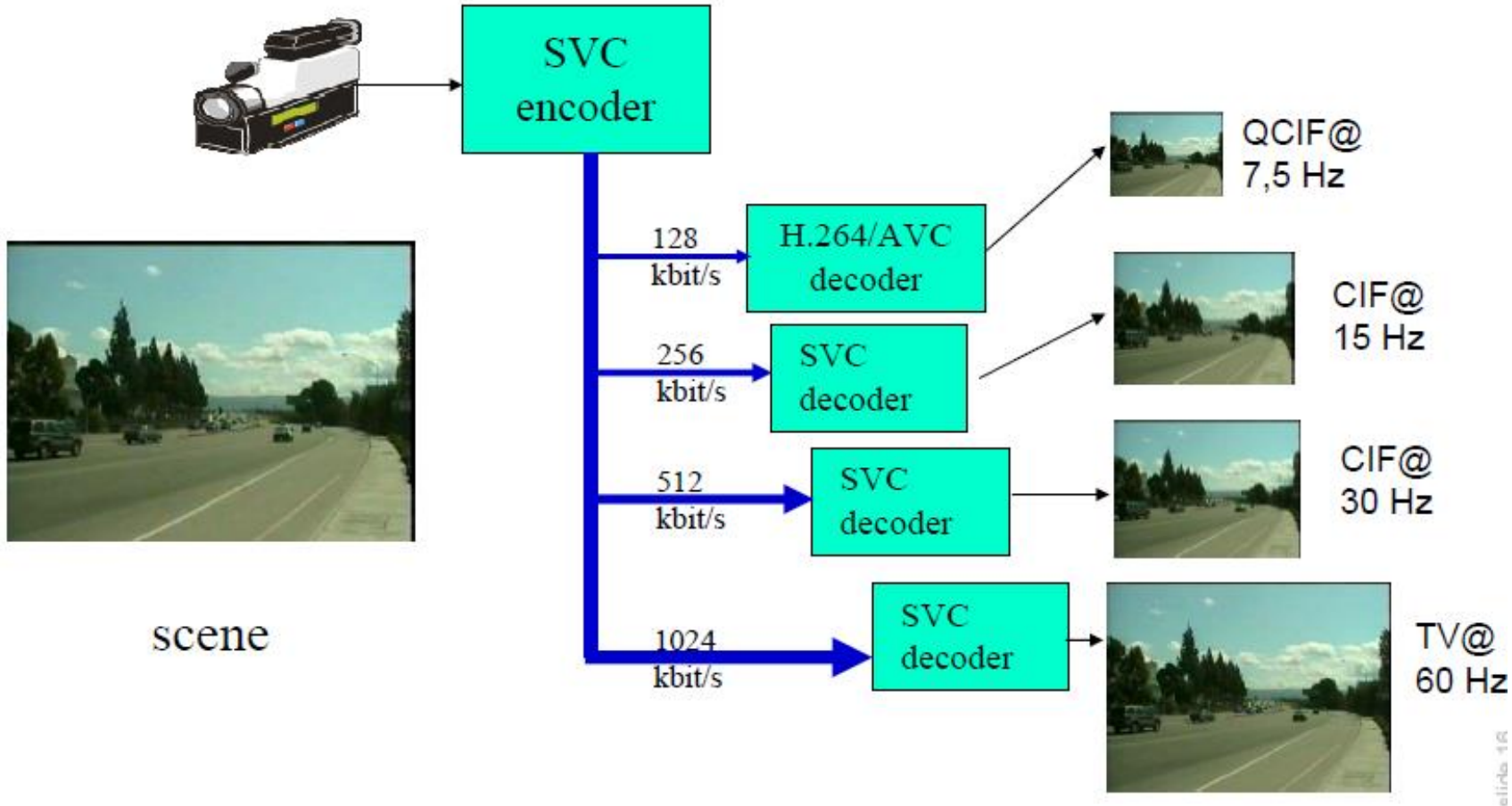
Definitions:

Video bit stream is called scalable if it's composed of several layers (from the base or the lowest layer to the highest layer) such that higher layers can be removed without compromising decoding of lower levels (i.e. the resulting stripped bit stream is still decodable). Generally speaking, each new enhanced layer is a 'delta' to the lower one. So, one can't decode k-th layer without decoding (k-1)-th lower layers.

Scalability Modes – three scalability modes are supported by AVC/H.264 : temporal, spatial (or resolution) and SNR (quality). Notice that there are other scalability modes which not explicitly supported by AVC/H.264, e.g. the bit-depth scalability (e.g. the base layer is SDR content and the enhanced layer is HDR).

Example: the spatial scalability means that the base layer is in the lowest resolution and each extension layer has a higher resolution than the lower layer.

Example: Spatial/Temporal scalability: higher layer has greater fps and resolution



SVC Basics

SVC is a straightforward extension to H.264/AVC with added complexity

Layered approach:

- One base layer
- One or more enhancement layers

The base layer is H.264/AVC compliant. In other words, SVC base sub-stream can be decoded by each H.264 decoder.

Enhancement layers enable Temporal, Spatial or Quality (SNR) scalability modes.

SVC Basics (cont.)

In Spatial and Temporal scalability, each subset of the bit-stream represents the source content with reduced picture size (Spatial Resolution) or frame rate (Temporal Resolution).

In case of quality scalability, also known as fidelity or SNR scalability, each subset of the bit-stream represents lower quality than the whole super-stream.

Multiple scalability modes can be combined to support various spatial-temporal resolutions and bit rates within a single super-bit-stream.

Simulcast vs. SVC

Simulcast mode: transmitting multiple bit-streams separately where each stream is independent of others, self-contained.

SVC mode: transmitting a single super-bitstream that contains several sub-streams (layers), ordered from the lowest layer (or the base layer) to the highest one. Each layer (excepting the base one) is dependent on the lower one.

Disadvantage of SVC is obvious: a decoder has to work hard, because to decode k-th layer Decoder needs to decode all (k-1) lower layers.

Advantage of SVC: inter-layer redundancy is exploited

Note: it's reported that SVC (in spatial scalability mode) provides a gain of 10% to 30% in bit rate reduction vs. Simulcast mode.

SVC Profiles

AVC/H.264 specifies three SVC profiles:

Scalable Baseline profile

Targeted for conversational and surveillance applications.
Support for Spatial Scalable coding is restricted to scaling ratios 1.5 and 2, between successive spatial layers.
Note: Interlaced video not supported.

Scalable High profile

Designed for broadcast, storage and streaming applications.
Spatial scalable coding with arbitrary resolution ratios supported.
Note: Interlaced video supported

Scalable High Intra profile

Designed for professional applications.
Contains only IDR pictures for all layers.
All other coding tools are same as Scalable High Profile.

SVC Spatial Scalability Inter-Layer Prediction Techniques

Prediction of a macroblock modes using modes of co-located macroblocks from a lower layer

Prediction of motion vectors using the up-sampled lower resolution motion vectors of co-located macroblock

Prediction of the residual signal using the up-sampled residual signal of the lower resolution layer