Scalable Encoding and Transcoding

ENTHRONE WORKSHOP
WP4

BSOFT ltd, Italy
Presentation: Daniele Renzi

Outline

- Content providers aim at best serving as many device groups as possible by generating scalable video streams with optimized decoding points. The **Scalable Video Coding** (**SVC**) answers this need in an optimum way.
- However, some new terminal devices do not adopt SVC immediately and other legacy terminals only implement **H.264/AVC** decoding.
- The SVC base layer is totally H.264/AVC compliant. Therefore, a set-top-box without SVC would only be able to decode the targeted quality of a mobile phone (i.e., the base layer quality). This is not satisfactory.
- Therefore, **adaptation** mechanisms are necessary.
- One possibility to address the problem is **transcoding** from the SVC to the H.264/AVC standard.

SVC in the ENTHRONE framework

- SVC gives the benefit of relocating the burden of adaptation from network modules, specifically conceived for such a task, to the content provider.
- ***Adaptation at the server level**:* The adaptation is performed on the initially stored video content in scalable format, before IP packetization and transmission.
- That simplifies the adaptation process and save some bandwidth compared to simulcasting single-layer streams.

Scalable Video Coding - **The SVC standard**

- **Scalability** has been a goal of video compression technologies for many years.
- For a long time the scalable video coding, e.g. MPEG-2 Scalable Extensions, has not had a big appeal in the market, mainly because of its high loss in terms of compression efficiency.
- The recent advances in video coding techniques led to the new standard **H.264/MPEG-4 Scalable Video Coding (SVC)** (Amendment 3 of ISO/IEC 14496-10, namely H.264/MPEG-4 AVC).

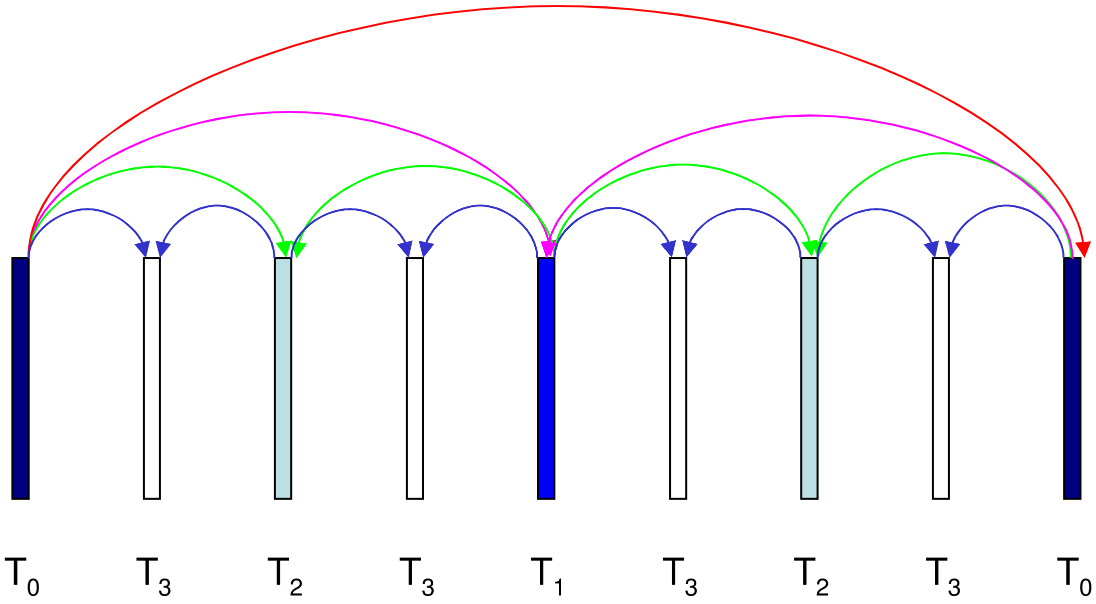
Scalable Video Coding - **Concepts**

- A video is called *scalable* when parts of it can be extracted as sub-streams which are still decodable to the decoder.
- Each sub-stream represents the source content in a reduced temporal, spatial and/or quality resolution compared to the original bit-stream.
- The source content is first encoded with low frame rate, low spatial resolution or low PSNR to form a ***base layer***. The residual information between the base layer and the original content is then encoded to form one or more ***enhancement layers***.

Scalable Video Coding – **Temporal Scalability**

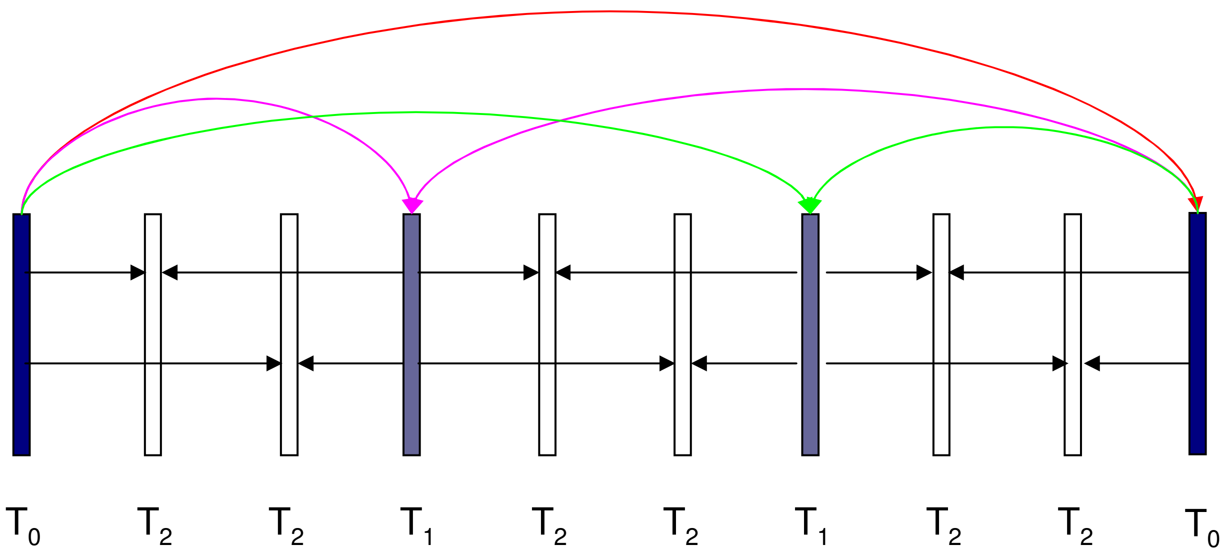
- Temporal scalability is generally enabled by restricting motion-compensated prediction to reference pictures with a temporal layer less than or equal to the temporal layer of the picture to be predicted.
- SVC usually employs **hierarchical B-pictures** to provide temporal scalability.
- SVC provides a considerably higher degree of **flexibility** on a picture and sequence level.

Scalable Video Coding – **Temporal Scalability – Dyadic prediction structure**



GOP (Group Of Pictures)

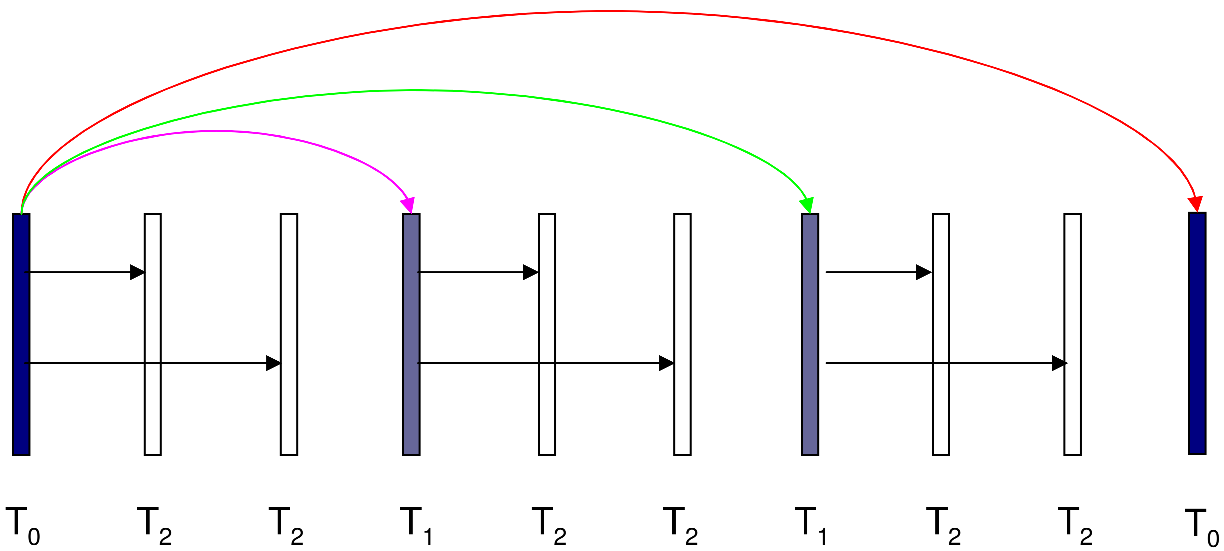
Scalable Video Coding – **Temporal Scalability – Non-dyadic prediction structure**



GOP (Group Of Pictures)

Scalable Video Coding – **Temporal Scalability**

– Hierarchical prediction structure with delay of 0



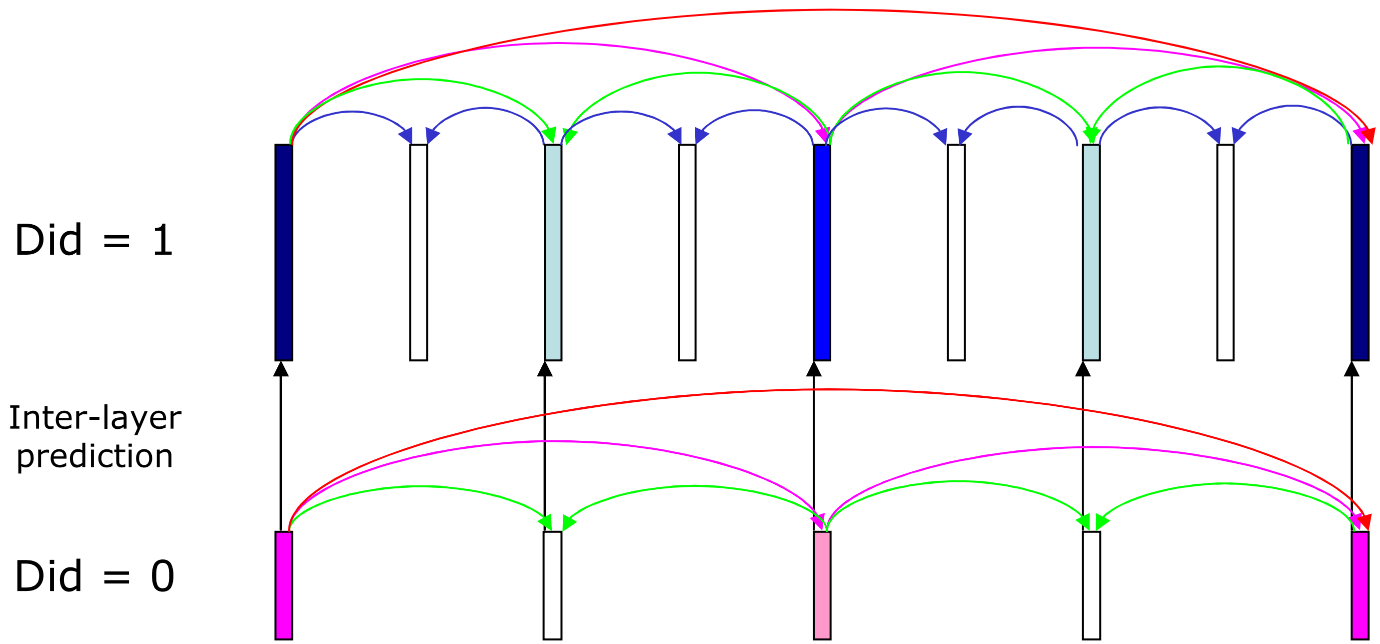
GOP (Group Of Pictures)

Scalable Video Coding – **Spatial Scalability**

- **Multiple-layer coding**.
- Each spatial layer corresponds to a supported spatial resolution and is uniquely labelled by a so-called ***dependency identifier (Did)***.
- Within each layer, motion-compensated prediction and intra coding are employed in the same way as in single-layer coding.
- In order to exploit the redundancy between spatial layers, additional **inter-layer prediction** mechanisms are incorporated.
- In order to limit the memory requirements and decoder complexity, SVC stipulates that all spatial layers should have an identical coding order.

Scalable Video Coding – **Spatial Scalability –**

Multi-layer structure with inter-layer prediction



Scalable Video Coding – **SNR Scalability**

- One base layer with minimum quality/bitrate, one or more enhancement layers with higher quality/bitrate.
- Usually the **quantizer** is chosen as unique parameter for tuning the quality levels among the different layers.
- Depending on the applications, higher or lower **granularity** can be needed.
- ***Coarse-Grained Scalability*** (CGS) and ***Medium Grain Scalability*** (MGS).

Scalable Video Coding – **SNR Scalability – Coarse-Grained Scalability (CGS)**

- In principle, CGS is identical to spatially scalable coding with the only exception that all layers have an identical spatial resolution.
- Texture information is typically refined by **re-quantizing** the residual texture signal in the enhancement layer, with a smaller quantization step size compared to that used in the preceding CGS layer.

Scalable Video Coding – **SNR Scalability – Coarse-Grained Scalability (CGS) advantages**

- **Simplicity**.
- **Low complexity** compared to single-layer coding.

Scalable Video Coding – **SNR Scalability – Coarse-Grained Scalability (CGS) drawbacks**

- **Low granularity**: The CGS can only provide a very limited number of bit rate points.
- **Low efficiency:** The multi-layer concept of CGS scalability becomes less efficient when the relative rate difference between adjacent CGS layers gets relatively small.
- **Low flexibility:** CGS scalability is unable to provide sufficient flexibility for all the applications.

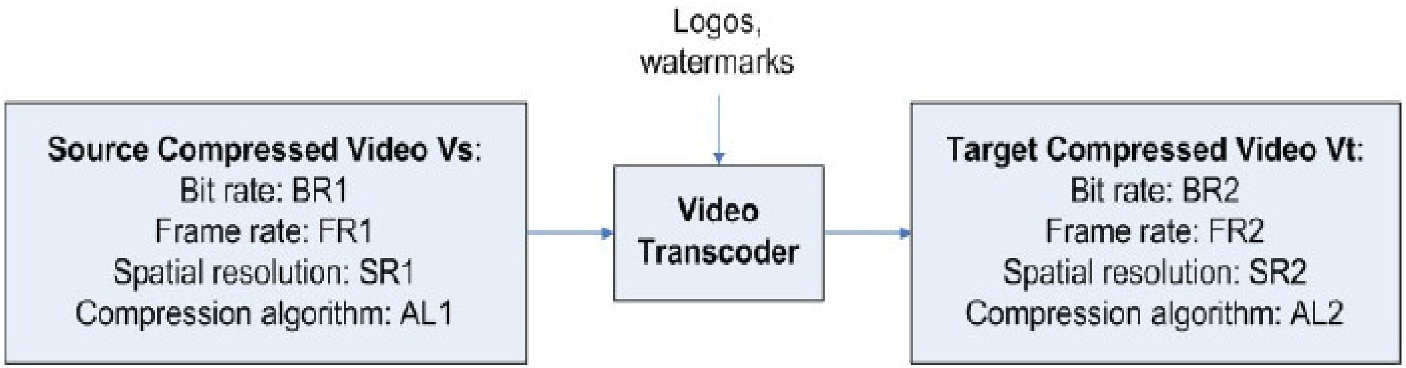
Scalable Video Coding – **SNR Scalability – Medium-Grained Scalability (MGS)**

- The MGS is advantageous over the CGS in that it contains a modified **high-level signaling**.
- That allows **bit rate switching** between different MGS layers in any access unit.
- **Graceful degradation**.
- With the MGS concept, any enhancement layer NAL unit can be discarded from an SNR scalable bit-stream, thus enabling **packet-based SNR scalable coding**.

Transcoding – General Approaches

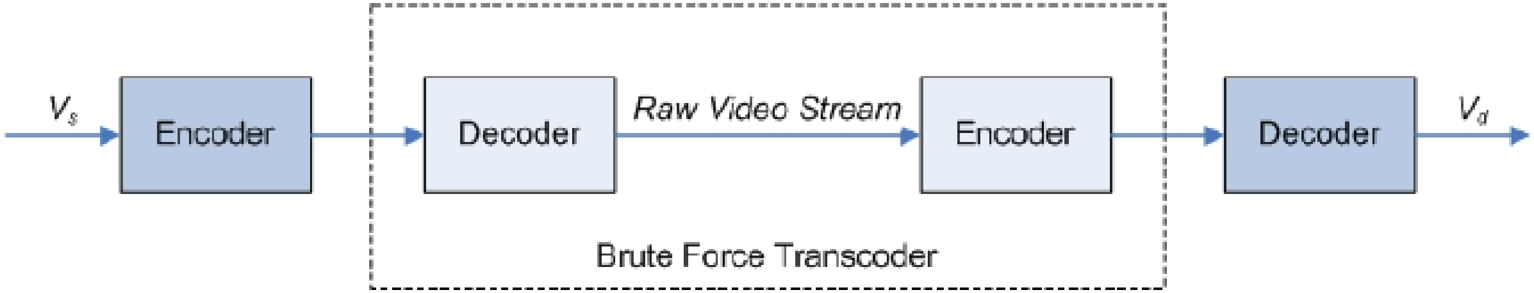
- **Video transcoding** can enable multimedia devices of different capabilities or formats to exchange video content.
- Generally a transcoder can have two major tasks: ***bit rate adjustment*** and ***format conversion***.
  - To suit available network bandwidth, a video transcoder can perform **dynamic bit rate adjustments** in the video stream without additional functional requirements in the decoder.
  - A video transcoder can provide **format conversion** to enable content exchange.
- For the time being, **several mainstream video compression standards** coexist in different multimedia applications.
- This makes transcoding necessary both within and across the standards to allow interaction between multimedia systems.

Transcoding – A Video Transcoder



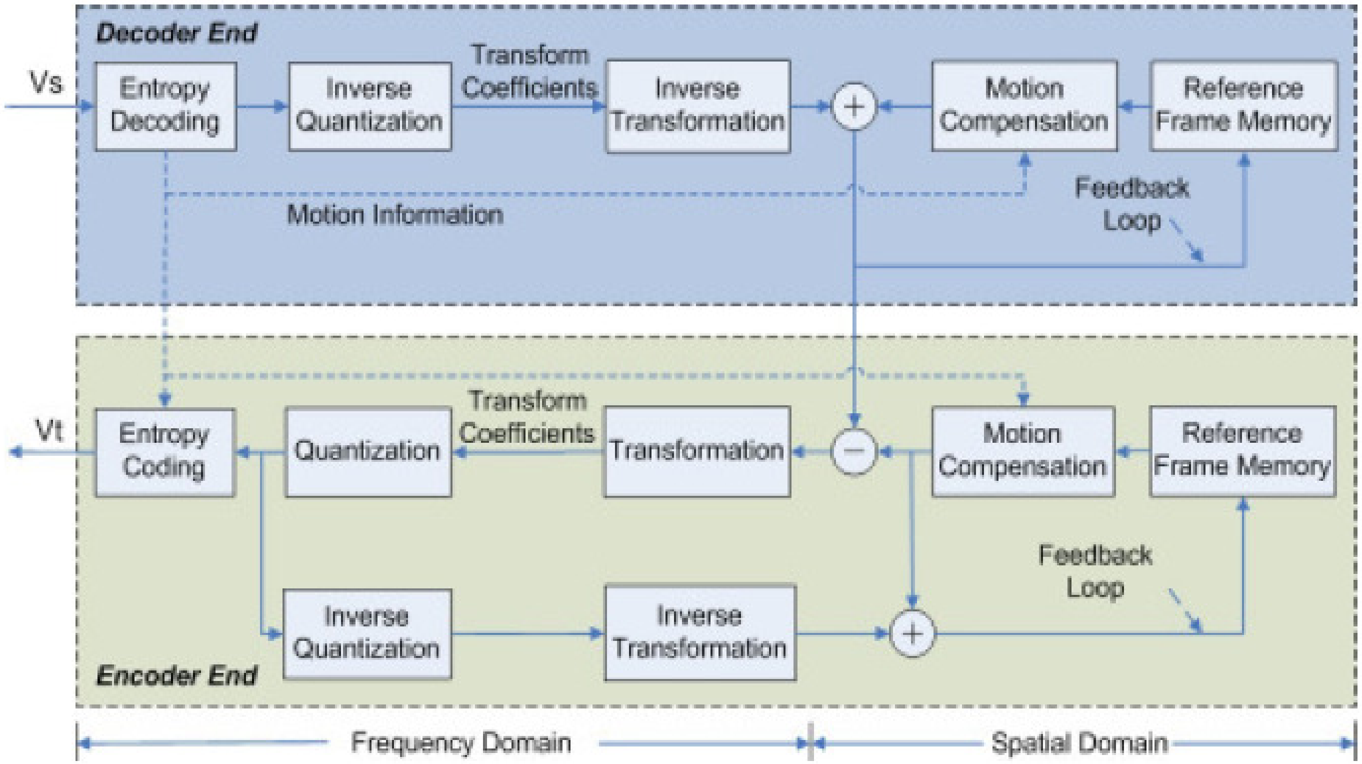
- ❑ Adjustment of coding parameters of the compressed video.
- ❑ Spatial and temporal resolution conversions.
- ❑ Insertion of new information such as digital watermarks or logos.
- ❑ Enhanced error resilience.

Transcoding – Brute-Force Transcoding



- Methodology: **fully decode the incoming source video stream and then re-encode the decoded source video** into the target bit rate and/or format.
- The full decoding and re-encoding is **complex** and consumes tremendous processing time and possibly requires extra equipment.
- While still maintaining acceptable quality, significant savings in complexity can be achieved by reusing as much as possible the information contained in the original incoming bit-stream.

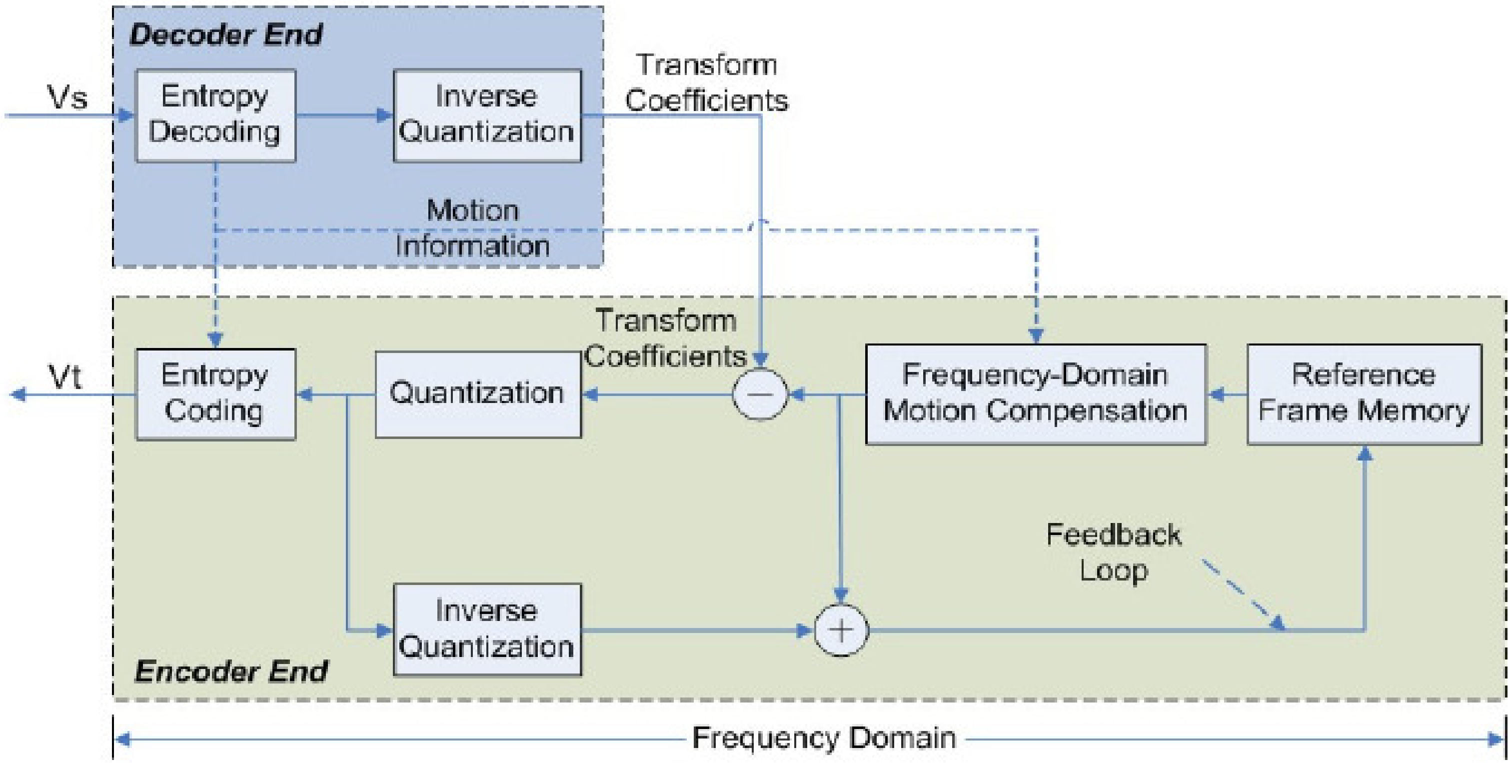
Transcoding – Spatial Domain Transcoding Architecture (SDTA) - 1



Transcoding – Spatial Domain Transcoding Architecture (SDTA) - 2

- SDTA is **flexible** in that its decoder-loop and encoder-loop can be independent of each other.
- The SDTA shown in previous figure **reuses the incoming motion vectors**. This process is indicated by dotted lines in the figure: the incoming motion information is transmitted to the Motion Compensation (MC) module in the encoding end and reused there.

Transcoding – Frequency Domain Transcoding Architecture (FDTA) - 1



Transcoding – Frequency Domain Transcoding Architecture (FDTA) - 2

- Only **entropy decoding** and **inverse quantization** is performed in the decoder end to get the transform coefficients of each macroblock.
- At the encoder end the motion compensated residual errors are encoded through **re-quantization** and **entropy coding**.
- After inverse quantization, the reference frame memory stores the DCT values, which are then fed to the frequency-domain motion compensation module to reduce the drift error.
- FDTA may require less computation but may suffer from the **drift** problem due to non-linear operations.
- FDTA also lacks flexibility and is mostly **appropriate to bit-rate transcoding**.

Transcoding – Transcoding of SNR Layers to H.264/AVC Single Layer - 1

- For the goal of converting CGS layers to H.264/AVC, a fast implementation technique known as the **CGS bit-stream rewriting** was proposed within the JVT (*JVT-U043*).
- In this approach, the syntax and semantics of the CGS layer are changed to enable a **fast rewriting of a CGS bit-stream into an H.264/AVC-formatted bit-stream**.
- No drift, no need for reconstructing the intensity values of the original sequence, but still able to derive the identical output that would otherwise be generated from an SVC decoder.
- **Merging multiple CGS layers** together.
- Basic idea: the additional overhead carried by SVC bit-streams is redundant for AVC bit-streams.
- It is then beneficial to **remove the SVC overhead** from the bit-stream, to reduce the bit rate required to deliver the same reconstruction quality.

Transcoding – Transcoding of SNR Layers to H.264/AVC Single Layer - 2

- Changes to the **Inter-coded macroblocks** of the CGS layer to enable the direct SVC-to-AVC mapping:
  - macroblocks that are inferred from base layer macroblocks must use the same **transformation size** as the base layer macroblock;
  - the mapping of an enhancement layer macroblock from a base layer macroblock shall occur in the **transform level domain**.
- **Intra-coded macroblocks** impose additional difficulties to the SVC-to-AVC rewriting: an intra-coded macroblock cannot be reconstructed by adding a signaled residual to a spatial prediction from its neighbors, as in IntraBL mode.

  Hence, changes to the Intra-coded macroblocks of the CGS layer to enable the **direct SVC-to-AVC mapping**:
  - A **modified decoding process** that maps the intra prediction mode from the base layer to the enhancement layer. Intra prediction is then performed at the enhancement layer.
  - **Transform type for IntraBL macroblocks** must be the same as the co-located base layer macroblock.
  - Macroblocks coded by the **16x16 transform** in the base layer are also coded by 16x16 transform in the enhancement layer.

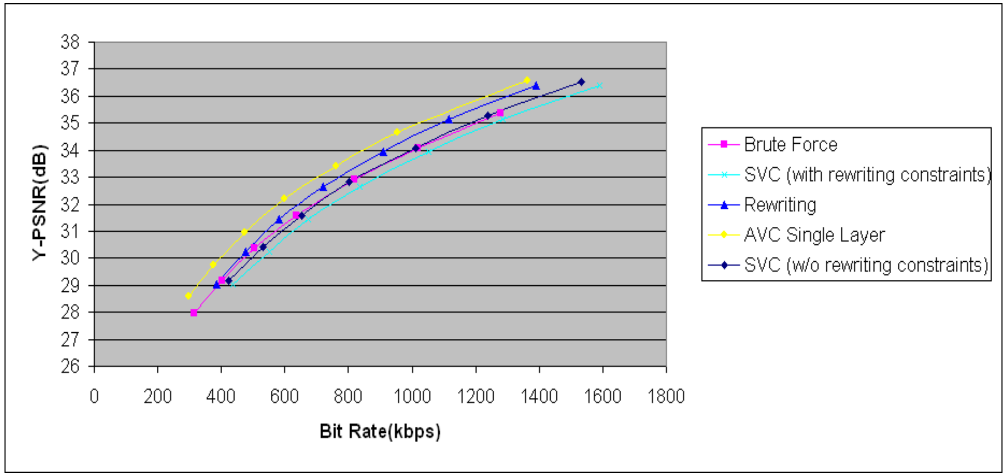
Transcoding – Transcoding of SNR Layers to H.264/AVC Single Layer – Test Results - 1

- Analysis was performed using the **JSVM** reference software.
- At first, CGS rewriting is compared with H.264/AVC single-layer coding.
- Second, CGS rewriting is compared with two-layer CGS coding to examine whether the CGS rewriting together with the subsequent single-layer decoding can outperform the normal SVC decoding for CGS bit-streams.
- Finally, CGS rewriting is compared with the brute force transcoding.

Transcoding – Transcoding of SNR Layers to H.264/AVC Single Layer – Test Results - 2

QP for BL	QP for EL	AVC Single Layer		Brute Force		Rewriting	
		Bit rate [kbps]	Y-PSNR [dB]	Bit rate [kbps]	Y-PSNR [dB]	Bit rate [kbps]	Y-PSNR [dB]
33	27	1363.97	36.57	1280.48	35.37	1389.12	36.41
35	29	955.38	34.65	1017.98	34.08	1116.95	35.14
37	31	762.69	33.41	820.27	32.89	909.24	33.93
39	33	598.14	32.19	635.17	31.57	719.47	32.66
41	35	474.94	30.94	504.75	30.35	580.91	31.45
43	37	376.35	29.75	402.89	29.15	477.81	30.25
45	39	297.94	28.59	313.37	27.97	385.23	29.03

QP for BL	QP for EL	SVC *with* rewriting constraints		SVC *without* rewriting constraints	
		Bit rate [kbps]	Y-PSNR [dB]	Bit rate [kbps]	Y-PSNR [dB]
33	27	1589.38	36.41	1533.96	36.54
35	29	1285.48	35.14	1238.01	35.27
37	31	1053.21	33.93	1012.47	34.06
39	33	835.10	32.66	802.11	32.82
41	35	674.74	31.45	652.91	31.59
43	37	550.35	30.25	531.07	30.41
45	39	436.56	29.03	424.08	29.16



Transcoding – Transcoding of SNR Layers to H.264/AVC Single Layer – Test Results - 3

Tests conclusions: at the same source sequence, format and QP settings, the PSNR of the decoding result from the rewritten bit-stream is identical to that of SVC coding with rewriting constraints. This meets the property that the CGS rewriting and SVC decoding should generate identical outputs.

Conclusions – Scalable Video Coding

- The recent advances in video coding techniques led to the new standard **H.264/MPEG-4 Scalable Video Coding (SVC)**.
- The SVC standard guarantees higher efficiency than older video coding standards supporting scalability, such as MPEG-2 Scalable Extension.
- That allows using SVC to perform easy bit stream **adaptation**.
- SVC gives the benefit of <u>relocating the burden of adaptation from network modules, specifically conceived for such a task, to the content provider</u>.
- SVC adaptation is made possible by the hierarchical structure of the SVC stream, that permits to extract only a subset of the data contained in the bit stream, without the need for additional resource-consuming operations.
- SVC video content adaptation can be performed by either:
  - dropping enhancement layers;
  - **transcoding** to other video coding standard, e.g. H.264/AVC.
- We mainly focused on the SVC transcoding, by highlighting the status of the art, our devised approach to transcoding, and test results.

Conclusions - Transcoding

- Two transcoding solutions have been discussed, namely the spatial-domain transcoding architecture (**SDTA**) and the frequency domain transcoding architecture (**FDTA**).
- These general approaches have been mapped to **SVC-2-AVC transcoding**, where the re-encoding of inter-layer predicted macroblocks is a critical step to eliminate inter-layer dependencies.
- The rate distortion performance of **CGS rewriting** has been studied in comparison to H.264/AVC single-layer coding, brute force transcoding, and SVC scalable coding with and without rewriting constraints.
- The tests shown that the CGS rewriting can deliver a better rate distortion performance than the SVC scalable coding and is thus proven very effective in merging CGS layers into an H.264/AVC compatible target layer.

THANK YOU!

Comments, questions, etc. are welcome.