

Scalable Video Coding for IPTV Services

Thomas Wiegand, *Senior Member, IEEE*, Ludovic Noblet, and Fabrizio Rovati

Abstract—Scenarios for the use of the recently approved Scalable Video Coding (SVC) extension of H.264/MPEG4-AVC in IPTV services are presented. For that, a brief technical overview of SVC when deployed in IPTV services is provided. The coding efficiency of the various scalability types of SVC is demonstrated followed by an analysis of the complexity of the various SVC tools. Based on this technical characterization, it is described how the different SVC features such as efficient methods for graceful degradation, bit rate adaptation, and format adaptation, can be mapped to application requirements of IPTV services. It is discussed how such mappings can lead to improved content portability, management and distribution as well as an improved management of access network throughput resulting in better quality of service and experience for the users of IPTV services.

Index Terms—H.264, IPTV, MPEG4-AVC, SVC.

I. INTRODUCTION

SCALABLE video coding (SVC) has been an active research and standardization area for at least 20 years. Prior international video coding standards such as H.262 | MPEG-2 Video [1], H.263 [2], and MPEG-4 Visual [3] already include several tools by which scalability can be supported. However, the scalable profiles of those standards have rarely been used. Reasons for that include the limitations of traditional video transmission systems as well as the fact that the spatial and quality scalability features came along with a significant loss in coding efficiency as well as a large increase in decoder complexity as compared to the corresponding non-scalable profiles.

The recently standardized SVC extension of H.264/MPEG4-AVC [4], [5] offers a significantly improved bit rate, distortion and complexity trade-off as described in this paper. But the deployment of a new video coding technology is also dependent on the business environment in which it is planned to be used. SVC offers a number of features that include efficient methods for graceful degradation, bit rate adaptation, and format adaptation. However, these features were not needed in traditional television broadcast via satellite, cable, or terrestrial channels. In traditional broadcast, the transmission channel either works perfectly or is not working at all, making graceful degradation and (to some extent) also bit rate

adaptation unnecessary. Moreover, the only deployed television format is standard definition, rendering format adaptation is not required.

The recent introduction of television (TV) via IP-based access networks (IPTV) associated with heterogeneous terminal resolutions (PC, SD, HD-ready and Full-HD TV) create the need for scalability. In this paper, we will characterize IPTV and show that the features offered by SVC (graceful degradation, bit rate adaptation, and format adaptation) could be very beneficial. More precisely, we describe how the SVC features can be used to provide content embedded scalability for achieving content-everywhere IPTV services. The results are improved

- content portability,
- content management and distribution,
- management of access network throughput, and
- quality of service and experience.

This paper is organized as follows. In the next section, a brief overview of SVC with emphasis on the applicability for IPTV services is provided. Section III gives results for coding efficiency of SVC showing its suitability for efficiently achieving graceful degradation, bit rate adaptation, and format adaptation. Section IV provides an assessment of complexity associated with the various SVC tools. Section V then describes how the SVC features can be beneficially used in IPTV applications.

II. OVERVIEW OF SVC

In general, a video bit stream is called scalable when parts of the stream can be removed in a way that the resulting sub-stream forms another valid bit stream for some target decoder, and the sub-stream represents the source content with a reconstruction quality that is less than that of the complete original bit stream but is high when considering the lower quantity of remaining data. Bit streams that do not provide this property are referred to as single-layer bit streams. The usual modes of scalability are temporal, spatial, and quality scalability. Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) and frame rate (temporal resolution), respectively. With quality scalability, the sub-stream provides the same spatio-temporal resolution as the complete bit stream, but with a lower fidelity—where fidelity is often informally referred to as signal-to-noise ratio (SNR).

The basic SVC design can be classified as layered video codec. In general, both the coder structure and the coding efficiency depend on the type of scalability that is required by an application. As an important feature of the SVC design, most components of H.264/MPEG4-AVC are re-used as specified in the standard. The base layer of an SVC bit stream is coded in compliance with H.264/MPEG4-AVC, and each standard compliant H.264/MPEG4-AVC decoder is capable of decoding this base layer when it is provided with an SVC bit stream.

Manuscript received August 05, 2008; revised February 17, 2009. First published May 05, 2009; current version published May 22, 2009.

T. Wiegand is with the Technical University of Berlin, 10587 Berlin, Germany. He is also with the Fraunhofer Institute for Telecommunications-Heinrich Hertz Institute (HHI), 10587 Berlin, Germany (e-mail: wiegand@hhi.de).

L. Noblet is with Orange Labs, 92794 Issy Moulineaux, France (e-mail: ludovic.noblet@orange-ftgroup.com).

F. Rovati is with STMicroelectronics' Advanced System Technology Labs, 20041 Agrate Brianza, Italy (e-mail: fabrizio.rovati@st.com).

Digital Object Identifier 10.1109/TBC.2009.2020954

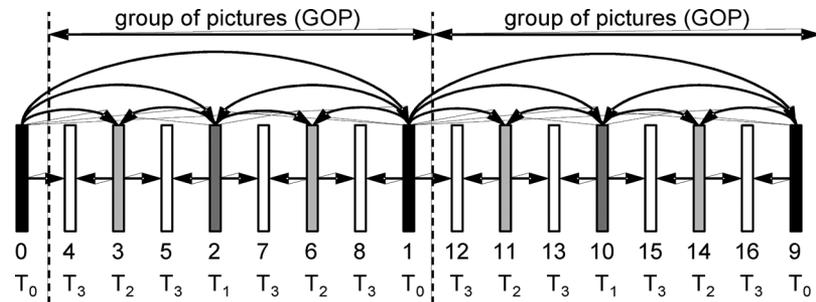


Fig. 1. Hierarchical B pictures enabling temporal scalability.

New compression tools are only added for supporting spatial and SNR scalability.

For a detailed description of SVC, please refer to the standard itself [4] or to [5]. In this paper, we will concentrate on a description relevant to IPTV services.

A. Profiles

The SVC Amendment of H.264/MPEG4-AVC specifies three profiles for scalable video coding [4], [5]: Scalable Baseline, Scalable High, and Scalable High Intra profile. The Scalable Baseline profile is mainly targeted for conversational and surveillance applications that require a low decoding complexity. Scalable Baseline profile has restrictions on the types of spatial scalability ratios and cropping it supports. In the Scalable High profile, which was designed for broadcast (including IPTV), streaming, and storage applications, spatial scalable coding with arbitrary resolution ratios and cropping parameters is supported. Quality and temporal scalable coding are supported without any restriction in both the Scalable Baseline and the Scalable High profile. Bit streams conforming to the Scalable Baseline and Scalable High profile contain a base layer bit stream that conforms to the restricted Baseline profile and the High profile of H.264/MPEG4-AVC [6], respectively. For IPTV services, it is expected that the Scalable High profile will be used and is therefore described in the next sections.

Bit streams conforming to the Scalable High Intra profile, which was mainly designed for professional applications, contain only IDR pictures (for all layers). Besides that, the same set of coding tools as for the Scalable High profile are supported.

B. Temporal Scalability

A bit stream provides temporal scalability when the set of coded pictures can be partitioned into a temporal base layer and one or more temporal enhancement layers with the following property. Let the temporal layers be identified by a temporal layer identifier i , which starts from 0 for the base layer and is increased by 1 from one temporal layer to the next. Then for each natural number k , the bit stream that is obtained by removing all access units of all temporal layers with a temporal layer identifier i greater than k forms another valid bit stream for the given decoder.

A typical inter-frame prediction structure that is used in the context of H.264/MPEG4-AVC is called hierarchical B pictures [8] and is illustrated in Fig. 1. While the pictures are depicted in

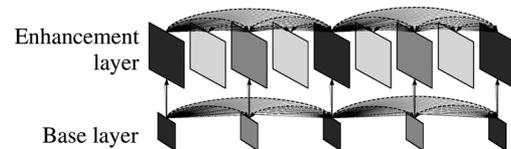


Fig. 2. Multi-layer structure with additional inter-layer prediction for enabling spatial scalable coding.

display order, the numbers below the pictures give the decoding/transmission order. The temporal layering is given by the T_i symbols with T_0 being the temporal layer with the highest importance (i.e. the temporal base layer) and T_3 being the temporal layer with the lowest importance (i.e. the highest temporal layer). In addition to enabling temporal scalability, the hierarchical prediction structures also provide improved coding efficiency when compared to classical IBBP coding at the cost of an increased encoding-decoding delay.

For hybrid video codecs, temporal scalability can generally be enabled by restricting motion-compensated prediction to reference pictures with a temporal layer identifier that is less than or equal to the temporal layer identifier of the picture to be predicted. Prior video coding standards such as MPEG-1 [7], H.262 | MPEG-2 Video [1], H.263 [2], and MPEG-4 Visual [3] all support temporal scalability to some degree. H.264/MPEG4-AVC [6] provides a significantly increased flexibility for temporal scalability because of its reference picture memory control. It allows the coding of picture sequences with arbitrary temporal dependencies, which are only restricted by the maximum usable Decoded Picture Buffer (DPB) size. Hence, for supporting temporal scalability with a reasonable number of temporal layers, no changes to the design of H.264/MPEG4-AVC were required. The only related change in SVC refers to the signaling of temporal layers and temporal layer switching points.

C. Spatial Scalability

For supporting spatial scalable coding, SVC follows the conventional approach of multi-layer coding, which is also used in H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual. Each spatial layer corresponds to a supported spatial resolution. In each spatial layer, motion-compensated prediction and intra prediction are employed as for single-layer coding. But in order to improve coding efficiency in comparison to simulcasting different spatial resolutions, additional so-called *inter-layer prediction* mechanisms are incorporated as illustrated in Fig. 2. Simulcasting refers to the provisioning and transmission of different

quality levels with multiple single-layer coded bit streams. Each quality level has assigned an independent bit stream.

In order to restrict the memory requirements and decoder complexity, SVC specifies that the same decoding order be used for all supported spatial layers. However, as illustrated in Fig. 2, lower layer pictures do not need to be present in all access units, making it possible to combine temporal and spatial scalability.

1) *Inter-Layer Prediction*: In order to improve the coding efficiency of the enhancement layers in comparison to simulcast, inter-layer prediction mechanisms have been introduced in SVC. These prediction mechanisms are switchable, so that an encoder can freely choose which lower layer information should be exploited for efficient enhancement layer coding. The main goal when designing inter-layer prediction tools is to enable the usage of as much information of the lower layers as possible for improving coding efficiency of the enhancement layers.

Inter-layer motion prediction: In order to employ reference layer motion data for spatial enhancement layer coding, additional macroblock modes have been introduced in spatial enhancement layers. The macroblock partitioning is obtained by upsampling the partitioning of the co-located 8×8 block in the lower resolution reference layer. The reference picture indices are copied from the co-located reference layer blocks, and the associated motion vectors are scaled by a factor of 2 in case of dyadic scalability. These scaled motion vectors are either used unmodified or refined by an additional quarter-sample motion vector refinement. Additionally, a scaled motion vector of the lower resolution can be used as motion vector predictor for the conventional macroblock modes.

Inter-layer residual prediction: The usage of inter-layer residual prediction is signaled by a flag that is transmitted for all inter-coded macroblocks. When this flag is set to true, the base layer signal of the co-located block is upsampled block-wise and used as prediction for the residual signal of the current macroblock, so that only the corresponding difference signal is coded.

Inter-layer intra prediction: Furthermore, an additional intra macroblock mode is introduced, in which the prediction signal is generated by upsampling the co-located reconstruction signal of the lower layer. For this prediction it is generally required that the lower layer is completely decoded. This includes the computationally complex operations of motion-compensated prediction and deblocking. However, as shown in [9] this problem can be circumvented if the inter-layer intra prediction is restricted to those parts of the lower layer picture that are intra-coded. With this restriction, each supported target layer can be decoded with a single motion compensation loop.

Generalized scalability and interlaced coding: SVC supports spatial scalable coding with arbitrary resolution ratios. The only restriction is that neither the horizontal nor the vertical resolution can decrease from one layer to a higher layer. The SVC design further includes the possibility that an enhancement layer picture represents only a selected rectangular area of its corresponding reference layer picture, which is coded with a higher or identical spatial resolution. Alternatively, the enhancement layer picture may contain additional parts beyond the borders of the reference layer picture. This so-called reference and enhancement layer *cropping* may also be combined and can even be modified on a picture-by-picture basis.

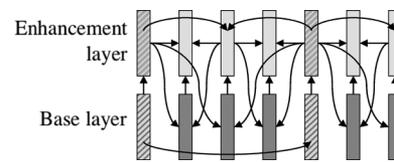


Fig. 3. Key picture concept of SVC for hierarchical prediction structures, where key pictures are marked by the hatched boxes.

Furthermore, the SVC design also includes tools for spatial scalable coding of interlaced sources. For both extensions, the generalized spatial scalable coding with arbitrary resolution ratios and cropping as well as for the spatial scalable coding of interlaced sources, the three basic inter-layer prediction concepts are maintained. But especially the derivation process for motion parameters as well as the design of appropriate upsampling filters for residual and intra blocks needed to be generalized. For a detailed description of these extensions the interested reader is referred to [10] and [11].

D. Quality Scalability

Quality scalability can be considered as a special case of spatial scalability with identical picture sizes for base and enhancement layer. For quality scalability, coarse-grain scalability (CGS) and medium-grain scalability (MGS) are distinguished.

Coarse-grain scalability: CGS coding is achieved using the concepts for spatial scalability. The only difference is that for CGS the upsampling operations of the inter-layer prediction mechanisms are omitted. Note that the restricted inter-layer prediction that enables single-loop decoding is even more important for CGS than for spatial scalable coding.

Medium-grain SNR scalability: For MGS coding in SVC, two new approaches to quality scalability have been introduced: motion-compensated prediction for the base layer from the enhancement layer and *key pictures* [12]. These approaches are illustrated in Fig. 3.

The hatched boxes mark the key pictures. All layers of these pictures, i.e., the base and the enhancement layers, are exclusively motion-compensated from other base layer pictures at encoder and decoder. This has the advantage that if any enhancement layer data are lost, no drift occurs. The disadvantage is a reduced coding efficiency of these key pictures since motion compensation references lower quality pictures. Hence, the number of key pictures should be kept small.

For all other pictures (that are not key pictures) motion-compensated prediction at the encoder is referencing the highest quality picture available. If the highest quality reference pictures are available at the decoder, they are also used for motion compensation. Otherwise, their lower quality counterparts are used for motion compensation at the decoder. The drift that is introduced by referencing a different picture at encoder and decoder is bound by the hierarchical B picture prediction structure, i.e., the key pictures close the recursive DPCM prediction loop while the non-key pictures show a very limited recursive prediction dependency.

Another important difference between CGS and MGS is that the transform coefficient refinement signal is accumulated in the

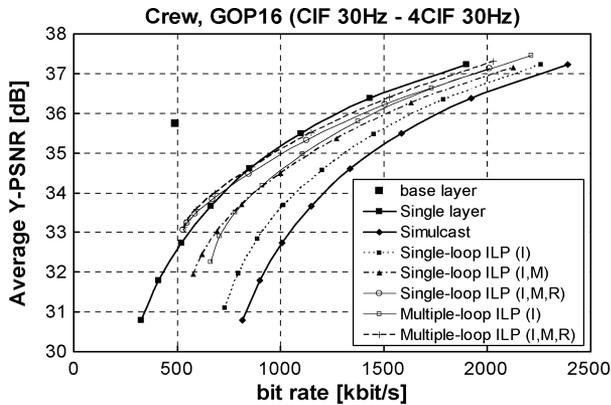


Fig. 4. Efficiency analysis of the inter-layer prediction concepts in SVC for different sequences and prediction structures. The rate-distortion point for the base layer is plotted as a solid rectangle inside the diagrams, but it should be noted that it corresponds to a different spatial resolution.

frequency domain for MGS as opposed to accumulating it in the spatial domain for CGS (and spatial scalability).

III. RATE DISTORTION PERFORMANCE OF SVC

A. Rate Distortion Performance of Spatial Scalability

The efficiency of the SVC inter-layer prediction techniques for spatial scalable coding has been evaluated in comparison to single-layer coding and simulcast. For this purpose, the base layer was coded at a fixed bit rate, whereas for encoding the spatial enhancement layer, the bit rate as well as the amount of enabled inter-layer prediction mechanisms was varied. Additional simulations have been performed by allowing an unconstrained inter-layer intra prediction and hence decoding with multiple motion compensation loops is required. Only the first access unit was intra-coded and Context-Adaptive Binary Arithmetic Coding (CABAC) was used as the entropy coding method. Simulations have been carried out for a Group Of Pictures (GOP) size of 16 pictures. All encoders have been rate-distortion optimized according to [13]. For each access unit, first the base layer is encoded, and given the corresponding coding parameters, the enhancement layer is coded [14]. The inter-layer prediction tools are considered as additional coding options for the enhancement layer in the operational encoder control. The lower resolution sequences have been generated following the method in [13]. The simulation results for the sequence “Crew” with spatial scalability from CIF (352×288) to 4CIF (704×576) and a frame rate of 30 Hz are depicted in Fig. 4. As can be seen, the inter-layer prediction (ILP) tools, given as intra (I), motion (M), and residual (R) prediction, improve the coding efficiency when compared to simulcast.

The results can be further improved by applying a multi-layer encoder control which takes into account the impact of the base layer coding decisions on the rate-distortion efficiency of the enhancement layers as presented in [15]. The algorithm determines the base layer coding parameters using a weighted sum of the Lagrangian costs for base and enhancement layer. Via the corresponding weighting factor it is possible to trade-off base and enhancement layer coding efficiency. In Fig. 5, an example result for spatial scalable coding with hierarchical B pictures

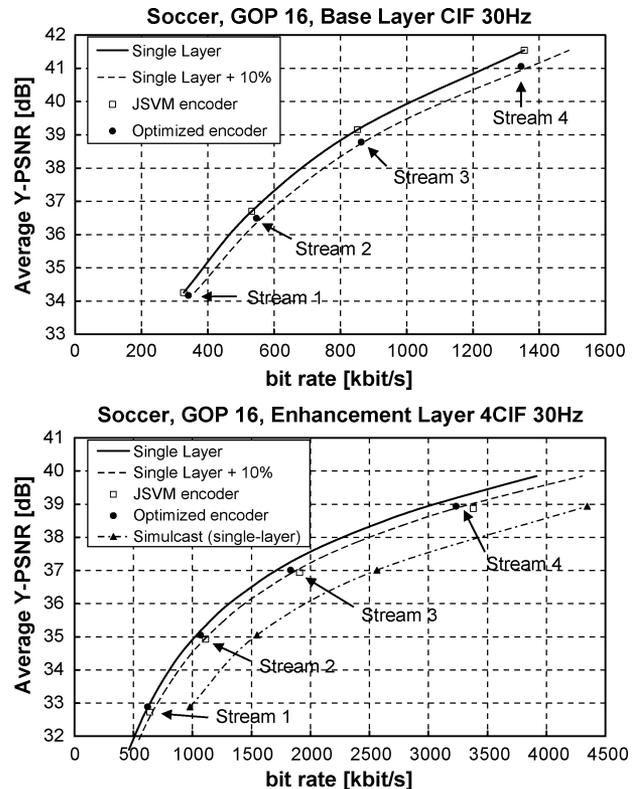


Fig. 5. Experimental results for spatial scalable coding (from CIF to 4CIF, 30 Hz) of the sequence “Soccer” using an optimized encoder control.

and a GOP size of 16 pictures is shown. Four scalable bit streams have been coded with both the JSVM and the optimized encoder control. The quantization parameter QP_E for the enhancement layer was set to $QP_B + 4$, with QP_B being the quantization parameter for the base layer. With the optimized encoder control the SVC coding efficiency can be controlled in a way that the bit rate increase relative to single layer coding for the same fidelity is always less than or equal to 10% for both the base and the enhancement layer.

The coding efficiency was tested in a subjective test in MPEG as described in [16], [17] using a Single Stimulus procedure namely the Single Stimulus MultiMedia (SSMM) method. This means that no reference sequence is available and the test subjects do only see the coded video before they give a vote. The following sequences have been encoded for subjective tests of spatial scalability performance: Crew, Crowd Run, Seeking, and Soccer. The resolutions were QVGA at 15 or 12.5fps and VGA at 30 or 25fps. The actual bit rates for the SVC encoding have been set to

- Crew, QVGA: 384 kbps, VGA: 1152 kbps
- Crowd Run, QVGA: 1024 kbps, VGA: 3072 kbps
- Seeking, QVGA: 512 kbps, VGA: 1536 kbps
- Soccer, QVGA: 256 kbps, VGA: 768 kbps

The bit rate for the H.264/MPEG4-AVC single-layer encoding was chosen to be 10/11th of the bit rate of the SVC encodings. The intra refresh period was set to 1.6 and 1.92 seconds for the 30Hz and 25Hz sequences (at highest temporal resolution), respectively. All encodings were done with a similar degree of encoder optimization following [13]. For

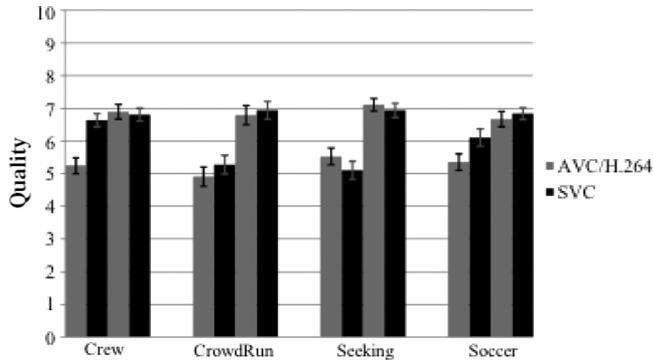


Fig. 6. Subjective test results for spatial scalable coding (from QVGA to VGA, 30 Hz) of the sequences “Crew”, “Crowd Run”, “Seeking”, and “Soccer”. For each sequence, the two left-hand side bars show subjective quality for QVGA and the right-hand side bars show subjective quality for VGA.

the SVC bit streams we additionally employed the multi-layer encoding concept presented in [15] in order to distribute the coding efficiency losses relative to single layer coding between base and enhancement layer. The MOS results for the sequences as depicted in Fig. 6 show that also for subjective results the spatial scalability in SVC comes with a bit rate overhead of less than 10% for both the base and enhancement layer.

B. Rate Distortion Performance of Quality Scalability

In a first set of experiments, the CGS and MGS methods for providing quality scalability are evaluated. In Fig. 7, the coding efficiency of CGS coding and MGS coding with key pictures is compared to that of single-layer coding for hierarchical B pictures with a GOP size of 16 pictures. For the quality scalable bit streams, the bit rate interval between the lowest and highest supported rate point corresponds to a QP difference of 12, i.e., the enhancement layer quantization step is equal to 1/4th of the base layer quantization step size. By comparing different CGS configurations with different choices of δQP (DQP), which is the numerical difference between the QP values of two successive layers, it can be seen that coding efficiency generally decreases with an increasing number of supported rate points, i.e., with decreasing DQP. The diagrams also contain rate-distortion curves for CGS with multiple-loop decoding, which is not supported by the SVC design. As already observed for spatial scalable coding, multiple-loop decoding for CGS increases coding efficiency only slightly and therefore, it does not justify the corresponding increase in decoder complexity relative to single-loop decoding. Additionally Fig. 7 also shows the coding efficiency of the more flexible MGS coding with the usage of the key picture concept and a DQP of 6. The improved coding efficiency at the highest rate point and the reduced coding efficiency at the lowest rate point for the MGS runs in comparison to the CGS runs with DQP equal to 6 are a result of the improved encoder control for MGS.

In order to verify the coding efficiency of quality scalability, a similar set of experiments using multi-layer encoder control has been conducted as for spatial scalability. Fig. 8 shows how the coding efficiency of quality scalable coding can be improved by employing the optimized encoder control of [15]. For this

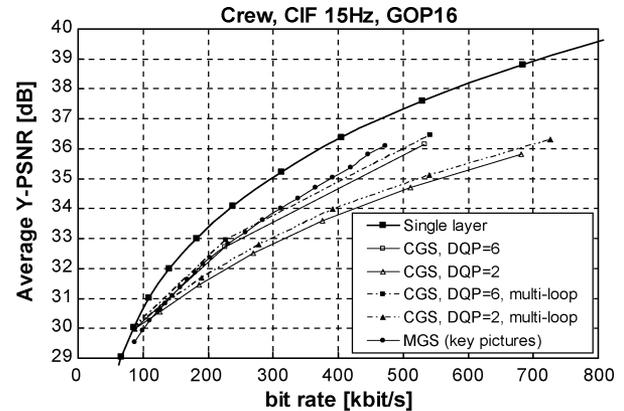


Fig. 7. Comparison of coarse-grain and medium-grain quality scalable coding with different configurations for the sequences “City” and “Crew” in CIF resolution and a frame rate of 15 Hz.

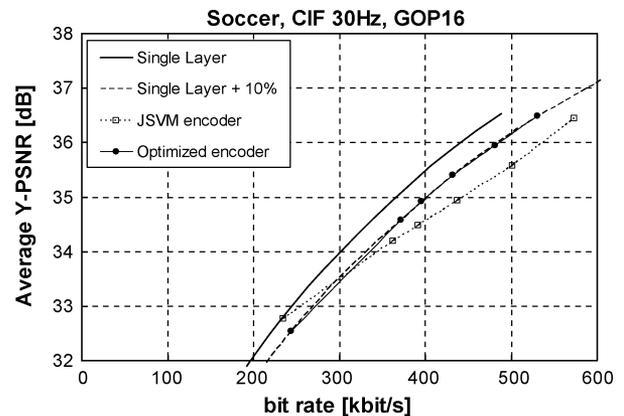


Fig. 8. Experimental results for quality scalable coding of the sequence “Soccer” (CIF resolution, 30 Hz) using an optimized encoder control.

simulation, hierarchical B pictures with a GOP size of 16 pictures were used. Quality scalability is achieved by MGS coding without using key pictures. The depicted rate points have been obtained by successively discarding the largest temporal levels of the MGS enhancement layer. It can be seen that coding efficiency can be significantly improved at the high-rate end by tolerating a coding efficiency loss for the lowest rate point. With the optimized encoder control it is possible to limit the bit rate increase compared to single-layer coding at the same fidelity to about 10% over the entire supported bit rate range.

The coding efficiency was again tested as described in [16], [17] using a Single Stimulus procedure namely the SSMM method. For that, the same set of sequences as above was coded at QVGA resolution with 15 or 12.5fps and the following bit rates:

- Crew, Soccer, low quality: 192 kbps, high quality: 384 kbps
- Crowd Run, Seeking, low quality: 384 kbps, high quality: 768 kbps

The bit rate for the H.264/MPEG4-AVC single-layer encoding was again chosen to be 10/11th of the bit rate of the SVC encodings while all other conditions were equal or as described for spatial scalability.

The results again verify what was already observed from the results in Fig. 9: SNR scalability can be obtained over a range

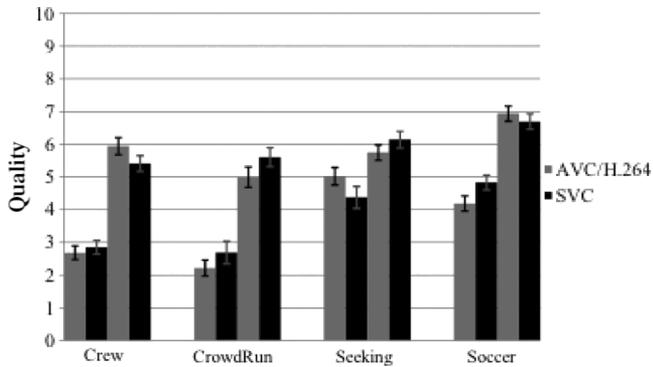


Fig. 9. Subjective test results for quality scalable coding (QVGA@15/12.5 fps) of the sequences “Crew”, Crowd Run”, Seeking, and Soccer. For each sequence, the two left-hand side bars show subjective quality for the lower bit rates and the right-hand side bars show subjective quality for the higher bit rates.

of factor two in bit rate with less than or equal to 10% bit rate overhead.

IV. COMPLEXITY ANALYSIS OF SVC

In this section we analyze the complexity analysis of SVC decoders, encoder complexity is not further addressed as it is heavily implementation-dependent and also out of the scope of the standard. Our complexity analysis of SVC decoders includes computational load (and proportionally, silicon implementation area), and data footprint/movement/formatting (and proportionally, external memory size/bandwidth requirements). The complexity evaluations are performed via assessing the relative complexity incurred when upgrading commercially available H.264/MPEG4-AVC decoders to fully support the newly introduced SVC tools.

First commercial applications around SVC have already been announced and several prototype applications have been showcased. As mentioned before, full SVC support starting from current H.264/MPEG4-AVC decoders will require substantial rework of available designs, but not a complete restart from scratch. This is descending from the original choice in the standardization committee to extend the current H.264/MPEG4-AVC coding scheme rather than using other scalable coding approach known in the art, as mentioned in Section II. All considerations below are contingent on the supported maximum resolution/frame rate and this can vary from one application to another. What can be impractical at a given resolution point can be technically feasible at a lower resolution and vice versa. We shall refer in this analysis to the use cases/resolutions as defined in Section V.

Table I summarizes all novel SVC coding tools and their respective implementation complexity for a typical mobile or Set Top Box (STB) device. Target device (form factor, power supply availability) and service (resolution, bit rate) differences determine distinctive complexity evaluation for the two cases.

In particular, if we take into account the Scalable Baseline profile, additional coding tools to be supported with regards to H.264/MPEG4-AVC Baseline profile are EI, EP, EB slices, 8×8 transform/prediction, weighted prediction, QP scaling matrixes, CABAC, MGS, spatial scalability (CGS, Dyadic, Extended Spatial Scalability with 1.5 ratio). Home devices,

TABLE I
COMPLEXITY ASSESSMENT FOR THE MOST RELEVANT SVC CODING TOOLS

SVC Profile	Tool	Complexity	
		Mobile	STB
Scalable Baseline & Scalable High Profile common tools	EI, EP slices	Easy	Easy
	EB slices	Medium	Easy
	8×8 transform/ prediction	Easy	Easy
	Weighted prediction	Easy	Easy
	QP scaling matrixes	Easy	Easy
	CABAC	Medium	Easy
	MGS (Unconstrained)	Hard	Hard
	Spatial scalability, $x1$ (CGS)	Easy	Easy
	Spatial scalability, dyadic	Easy	Easy
	Spatial scalability, ESS $x1.5$	Medium	Easy
Scalable High additional tools	Spatial scalability, ESS arbitrary ratio	Hard	Medium
	Interlaced PICAFF support	Not relevant	Medium
	Interlaced MBAFF support	Not relevant	Medium

such as media centers and STBs should be capable of supporting all these features at a small incremental cost, as most of these tools are already implemented anyhow in most of the products available on the market, as they normally support H.264/MPEG4-AVC Main or High profile. Spatial scalability features are the major missing functions, but those could be implemented in dedicated HW blocks based on up-sampling filters, in particular as this technology is well understood in the industry. MGS could potentially be more challenging to be implemented due to the double loop deblocking filtering. Nevertheless, this drawback might be overcome by determining a maximum frequency of key pictures in the bitstream.

As already pointed out, the possibility of employing inter-layer intra prediction is restricted to selected enhancement layer macroblocks, although coding efficiency can typically be improved by allowing this prediction mode in an enhancement layer, as it was done in the initial SVC design [18]. In [12], [19] and [20], however, it was shown that decoder complexity can be significantly reduced by constraining the usage of inter-layer intra prediction. The idea behind this so-called *constrained inter-layer prediction* is to avoid the computationally complex and memory access intense operations of motion compensation and deblocking for inter-coded macroblocks in the reference layer. Consequently, the usage of inter-layer intra prediction is only allowed for enhancement layer macroblocks, for which the co-located reference layer signal is intra-coded. It is further required that all layers that are used for inter-layer prediction of higher layers are coded using constrained intra prediction, so that the intra-coded macroblocks of the reference layers can be constructed without reconstructing any inter-coded macroblock.

Under these mandatory SVC restrictions, each supported layer can be decoded with a *single motion compensation loop*. Thus, the overhead in decoder complexity for SVC compared to single-layer coding is smaller than that for prior scalable video coding standards, which all require multiple motion compensation loops at the decoder side. Additionally, it should

be mentioned that each quality or spatial enhancement layer Network Abstraction Layer (NAL) unit can be parsed independently of the lower layer NAL units, which provides further potentials for reducing the complexity of decoder implementations [21].

Mobile implementations should put additional care into CABAC and EB slices decoding, as these are normally not implemented on current devices, even if the trend, at least for high-end devices, will be more and more to support HD High profile streams on mobile devices as well. Please note that CABAC decoding is only required for decoders capable of decoding bit streams conforming to level 2.2 and higher, i.e. resolutions typically above CIF resolution.

In Scalable High profile, additional tools need to be supported including Extended Spatial Scalability with arbitrary spatial ratios and interlaced coding. Both have a non-negligible impact on the implementation complexity, as they involve retrieving and reformatting several macroblocks of input data for each macroblock to be decoded. Arbitrary ratio ESS will generate unpredictable and potentially totally unaligned memory access patterns due to the oversampled grids, which will require more complex Direct Memory Access (DMA) engines. Interlaced support will also introduce some extra complexity, especially if MB-level adaptive frame/field (MBAFF) is used, as for each inter-layer prediction, macroblock pairing will have to be potentially undone and then redone at the upsampled macroblock grid level for field-to-frame and frame-to-field conversion. Since interlaced is not used in the mobile space, the Scalable High Profile is also unlikely to be deployed there.

Given all the above considerations, the evolutionary nature of the SVC will allow the design of products supporting the standard at non-negligible but also not very high costs. Towards this goal, practical deployment of SVC services can also be, at least in some applications, curtailed in some parameters to favor adoption at cost-effective points. Such restrictions could for example comprise the restriction of the number of key pictures in MGS, or on the different combination of field-frame consecution in successive spatial layer, along with the standard constraints on processing a maximum number of macroblocks or maximum bit rates that must be supported according to the level definitions in the SVC standard.

As already indicated, all these complexity considerations only apply to decoders. Encoder implementations have much more freedom in deciding which SVC feature to support and to which degree, just as any other hybrid video encoder. It is foreseeable that first commercial consumer implementations will be basic in supported features, letting competition and deployment requirements progressively enrich newly introduced products.

V. IPTV USE CASES AND SCALABILITY SCHEMES

Providing TV over IP networks can refer to many scenarios depending on the definition of TV and IP networks. Actually, the vast majority of "IPTV services" already offers both linear TV and video on demand services. The "IP network" is often referred to as both the delivery network (e.g. managed, unmanaged including the Internet) and the access network (fixed and

potentially also mobile). We will also use the term IP network with the same comprehensive meaning.

We observe the following trends in the IPTV customer/end user behavior:

- *the strong customer expectation for having the best quality of service and experience*: user dissatisfaction and frustration is a serious concern for IPTV service operators;
- *a customer appetite for everything on-demand and for personalization*: younger demographics being used to content consumption over the internet and familiar with content portability (e.g. music from PC to mp3 player) also desire the same features for TV and IPTV services;
- *a demand for timeless/immediate usage* with "time" being an important value for the customer. This leads to high expectations for the service richness with the challenge to not losing the interaction simplicity of existing TV services;

These trends on the customer side are coupled with the following trends on the service and technology provider sides:

- *a hyper-competitive market situation* not only for content and service providers but also for the consumer electronics industry with an amazing market rhythm for the introduction of innovative retail devices (e.g. flat screens and HDTV, portable media players (PMPs), Internet connectivity) and key interoperability issues to be solved;
- *a steady increase of average access bit rate* allowing the end-customer to easily get video content over the Internet,
- *the internet as a de facto convergent video platform* with flat network access subscription models and the availability of mobile phones with broadband-like Internet connectivity,
- *the legal obligations about network neutrality for QoS are still unclear*, i.e. no clear boundary between open Internet, Internet service provider and walled gardens, between managed and unmanaged delivery networks (a trend that is becoming increasingly complex with high penetration rates for home and mobile broadband connections)
- *a growing "long tail"* with more and more contents reaching smaller groups and a resulting audience fragmentation leading to key challenges and opportunities for advertising strategies;
- *provisioning of free/devalued contents over the open Internet* in order to attract customers to the walled garden world.

The trends described above occur in a context where there is more and more diversity in:

- *network access technologies* (residential with ADSL, VDSL, FTTH or even wireless with WiMax, mobile with 3G, 3G+ and 4G tomorrow);
- *service architectures* (centralized, distributed, managed or unmanaged, peer-to-peer or hybrid);
- *consumer devices* (from small handheld terminals as mobile phones all the way to HDTV screens), all possibility connected to the home network through a home gateway (Ethernet, Power line communications, wireless);
- *contents* (professional contents and media but also user generated contents and media).

The above considerations lead to a new general trend in all IPTV services towards providing video content everywhere, to

any device and at any time resulting in models like linear TV, catch-up, or on-demand services. This trend is inherently linked to horizontal service approaches in contrast to the initial more vertical introduction of IPTV, mobile and web TV/video services. Hence, in the near future, these emerging services will move towards maturity and initial technology choices may have to be reconsidered.

This major change in the video and TV landscape is leading to considerable needs for efficient content delivery architectures. One consideration is content adaptation which today becomes a key issue for achieving the “content everywhere paradigm” and can be performed at different stages:

- at *content preparation* and off-line repurposing time: *multiple content format instances* leading to considerable costs increase while the revenue generated by the content might be low (long tail, free/devalued content);
- at *delivery* (down to the edge) with on-line content purposing (i.e. at the time the content is delivered) requiring cost effective solutions and thus dense and massively parallel processing capabilities for being practical;
- at *end-user* and at the time the content is on the user’s device: this implies end-user devices capable of content transcoding, i.e. decoding a stream and re-encoding it with different parameters. Transcoding results in key issues on interoperability and content protection management (transcoding requires protection to be temporary removed). Moreover, transcoding typically takes considerable time, degrades the content quality and is generally also not feasible for the end-user devices without increasing the device cost.

Because of all these considerations, *scalability and flexibility are key points* for the near future of IPTV services, regardless whether these are new services or an evolution of existing ones. Such scalability is needed not only at the architecture and infrastructure levels but also at content level.

While many different video codecs are still used (depending on the targeted device for a given service), there is today a strong trend towards H.264/MPEG4-AVC. H.264/MPEG4-AVC has actually not only been widely implemented in STBs, it is also expected to be common in mobile devices (smart phones, mobile Internet devices) as well as in PMPs; it has even been recently introduced in Flash Player 9 and QuickTime players.

As it was previously mentioned, providing content everywhere services implies the implementation of content repurposing bricks somewhere in the service architecture for transcoding between formats (QVGA, VGA, WVGA, SDTV, HDTV, etc . . .), bit rates (access networks dependency) and for many output codecs. With H.264/MPEG4-AVC generalization, transcoding will soon be limited to format and bit rate adaptation and there will be less need for providing several output codecs. At the same time, the picture resolution both in terms of available contents and devices capabilities will considerably increase. These are also key points in considering the SVC extension of H.264/MPEG4-AVC.

This section describes use cases for SVC as a technology providing content embedded scalability for achieving content everywhere services and considering the panel of previously mentioned stakes:

- Content portability
- Optimized content management and distribution
- Smart management of access network throughput
- Improved quality of service and experience

A. Content Portability

With the multiplication of personal devices (personal computer, mobile smart phone, PMPs, etc.), content portability becomes a key feature for multi-screen video services in order to allow side loading/viewing between the screens: this refers to the ability for a consumer to watch video content, either downloaded or recorded, on a different device than the one it has been downloaded or recorded on. This also abstracts the notion of media library/repository in terms of device dependencies. Each of these personal devices may have different capabilities (display size, processing power). But, independently from these technical considerations, the major stake is probably interoperability among the devices including content protection. This consideration is one of the motivations for the SVC extensions of H.264/MPEG4-AVC. There are different ways for implementing content portability, from bottom level approaches (same to any) to content adaptation at the time the content is transferred from one device to the other. The first one implies bottom level value for the end-user and thus for the content service provider. The second one appears as the best and most flexible approach for valuing the service but it has to take into account the fact that time and simplicity are important values for the end-user.

Typically, content adaptation can be performed through a transcoding operation executed at the time the content is stored (but would require a pre-definite list of possible targeted devices) or at the time the content is exported from the media library to a given device. For the first option, this implies that the content is not available before it has been transcoded and wastes storage. For the second one, the transcoding operation may require to be performed at a much greater speed than real-time and typically at a speed consistent with the transfer capacity of the interface used for transferring the content (USB, Bluetooth, WiFi, etc . . .). For both cases, the transcoding operation would require significant processing power considering that this operation would most often be associated with resolution reduction. Thus, transcoding may require significant processing time for the end-user. In terms of simplicity, the end-user would have to configure the conversion operation unless the device implements an agent capable of describing its technical characteristics for automatically configuring the conversion process. It worth mentioning that prior to any transcoding operation, any content protection has to be removed which may not comply to content rights owner interests. Moreover, transcoding also typically results in degraded picture quality.

In this context, it is interesting to notice that, at the same time H.264/MPEG4-AVC is being generalized, the devices resolution capabilities are tending to more variety. Typically, those resolutions can be categorized as:

- QVGA up to 25/30 frames per second for smart phones, low-end mobile Internet devices. A sufficient resolution for PCs in the case of display within a web page.

- VGA and even WXGA up to 25/30 frames per second for high-end mobile Internet devices, PMPs. Default resolution for full screen display on PCs.
- SDTV up to 25/30 frames per second for set top boxes, digital video disc players and portable media players.
- HDTV up to 50/60 frames per second for set top boxes, digital video disc players, game consoles and PCs.

The combination between SVC spatial and temporal scalability allows the efficient implementation of content portability:

- between PCs, PMPs, smart phones and mobile Internet devices
 - QVGA base layer at 25/30 frames per second for the smart phone and the PC (windowed display mode)
 - VGA enhancement layer at 25/30 frames per second for the PC (full screen mode), the PMP and the mobile Internet device
- between STBs (or even the video disc player) and PMPs
 - SDTV base layer for low end PMPs in the case of an 720p/1080i content or 720p base layer for high-end PMPs in the case of an 1080p content
 - HDTV (720p/1080i or 1080p) enhancement layer

Using SVC to provide content portability addresses different key requirements, namely simplicity, instantaneous availability, content rights preservation as well as no picture quality degradation.

B. Optimized Content Management and Distribution

The increasing consumer appetite for on-demand content (VOD, catch-up TV) leads to a considerable growth of the long tail: smaller amount of popular content consumed by many people, larger amount of content individually consumed by few people. But the long tail is an important argument for service differentiation. Actually, the overall number of people consuming the long tail is probably more important than the overall number of people consuming the most popular contents. With multi-screen programming, the on-demand catalog has to be prepared for consumption on different devices leading to as many instances as there are targeted devices if conventional single layer codecs are used. While it can be easily understood that the content preparation cost can be efficiently amortized for the most popular content, it is less obvious for the long tail. The above mentioned costs correspond to the following tasks needed in the general process of providing on-demand content:

- content capture (from tape, from turn-around, from file, etc.), editing, may require indexing means as well as meta-data extraction/generation,
- content preparation (encoding, transcoding) and metadata processing,
- content integrity checking,
- content packaging (including protection),
- content provisioning,
- content ingest.

Content preparation and checking are important costs within this process. Multiplying the instances to be generated in order to address multiple devices through multiple access networks increases the overall on-demand process cost model, especially for long-tail content and even with automated workflows.

At the moment, some solutions are proposed which consist in preparing a single content instance and transcoding it at the time it is served (on-the-fly transcoding), depending on contextual usage information (typically the device capability). This is typically the case for making Internet content available to mobile devices. In order to achieve cost efficiency, the associated architecture requires high density and processing power to perform massively parallel content adaptation. These architectures become realistic with low complexity codecs and relatively small input and output picture resolutions (QVGA). It is not proven that they would still be feasible with higher complexity codecs such as H.264/MPEG4-AVC and with the higher input and output resolutions (typically starting from VGA) that the market is now requiring. Another point to make about IPTV and content distribution is about the last mile in the case of ADSL services and both HDTV channels and multi-room sets. While there is a strong consumer expectation for having HDTV content to be displayed on their (potentially large) flat screen, there are some cases for which the consumer may not be able to receive native HDTV content in real-time streaming, typically caused by network throughput limitations (access network, home network). This may lead to an unnecessary frustration since the content by itself might be of interest independently from its picture format. One solution could be to simultaneously provide an SDTV resolution of the content or even an intermediate HDTV one (typically half horizontal resolution). Using single layer codecs results in simulcast and storage capacities inflation (typically when catch-up TV is implemented). An intermediate instance (half HD for instance) could optionally be provided so that everyone can be served at the best of his/her network connection capabilities. In this case, SVC could considerably improve both the capital and operational expenditures for achieving this.

From what was described in the previous section about content portability, SVC with the same scalability scenario (spatial scalability) would considerably help to improve content management and distribution for on-demand services including situations where on-the-fly transcoding is still required for providing content to legacy devices not able of decoding SVC and H.264/MPEG4-AVC streams. In such a case, for the vast majority of the situations, only the base layer would have to be transcoded, preserving the implemented architecture.

C. Smart Management of Access Network Throughput

While the access network can be seen as a single pipe, its throughput for a given direction (upstream, downstream) is most often split into different logical or virtual channels, each one used for providing different classes of services in terms of quality of service and traffic priority. But this distribution is often static and leads to a considerable waste of bit rate in situations where all services are not being transmitted simultaneously. In the IPTV scenario, an ADSL subscriber may watch TV with guaranteed QoS in the living room while Internet video is being downloaded or streamed over the PC in another room. This subscriber may also watch an HDTV channel in the living room on an HDTV set while another channel is watched in another room on another SDTV-only set. The described usage scenarios may or may not occur at the same time. The

basic idea is to use SVC capabilities for adapting the distributed content to the context while optimizing the available access network resources. SVC SNR scalability provides efficient means for achieving a *better simultaneous walled garden IPTV and Internet experience* for the end user by *adapting the walled garden IPTV video bit rate depending on the throughput that is necessary for achieving the appropriate user experience* on the Internet side, within reasonable limits. Moreover, if the Internet video is encoded with SVC SNR scalability, it can also be adapted so that the impact over the walled garden IPTV stream is not perceptible (including fade-in/fade-out mechanisms at the transitions). This adaptation can be either dynamically managed on the network side or under full user control. With single layer video technologies, the implementation of such dynamic mechanisms would require a transrating operation to be performed somewhere in the network, most probably at the edge, and would lead to very complex architectures in terms of density for massive processing and cost efficiency. SVC based adaptation just requires dropping of selected IP packets and therefore allows significantly simpler mechanisms and more cost efficient implementations.

An extension of this use case is throughput dynamic adaptation, for a second TV set turned on (SDTV) when a first one is already being used (HDTV). In this case, *the available throughput can be shared between both programs, providing optimal trade-off in quality to each of them* depending on their properties and on the target terminal characteristics.

D. Improved Quality of Service and Experience

While managed content delivery (walled garden IPTV) is supposed to provide a defined level of Quality of Service (QoS), unmanaged content delivery (typically open Internet) mostly refers to best effort conditions delivery.

Considering managed networks, QoS is affected by packet loss (typically impulsive noise over an ADSL access network) and not by bit errors and network congestion. Actually, current solutions for QoS improvement over managed networks rely on Forward Error Correction (FEC) and/or packet retransmission. These solutions can be used with single layer or scalable video codecs. However scalable solutions can lead to more efficient schemes:

- by providing stronger FEC protection to the base layer, and lower ones to enhancement layers,
- by allowing retransmission without extra bit rate overhead: when one or more packets have to be retransmitted, an enhancement layer packet can be removed.

Still for managed networks, the use of SVC properties can provide better user experience in terms of channel change delay (rely on scalability) without requiring extra throughput provision, as well as for improving content navigation (trick modes) when streaming on-demand content from an IPTV VOD server over a bit rate constrained access network.

While IPTV emerged with SDTV, it is important to consider that more and more HDTV will be offered in the coming years and that some technical difficulties met with SDTV will have to be reconsidered including fast channel change for linear TV and trick modes for on-demand services. And SVC can be considered into this HDTV projection and not only the current context.

Actually, when changing from one TV channel to another one in IPTV, the process does not only include program identification and video synchronization as with broadcast but also a dialog between the end-user device and the service platform. This results in an increase of the channel change delay compared to broadcast and most IPTV operators are looking at reducing this delay for providing a more satisfactory user experience. Some solutions propose improvement at the video level which consists in implementing both a circular buffer for always providing a random access point after the channel change command has been executed and then temporarily bursting the served bit stream for reducing the buffer filling delay. This burst mechanism typically requires extra bit rate which can be acceptable for SDTV services but may not be possible when providing HDTV services. But when an IPTV service implements SVC for delivering HDTV channels including an SDTV base layer, it would be possible to temporarily burst an SDTV base layer at the overall throughput resulting in faster video buffer filling and thus much shorter delay at no extra throughput.

Concerning trick modes with HDTV VOD, most of the mechanisms rely on using intra pictures: intra pictures are picked from the bit stream at intervals which increase with the trick mode speed. The proportion between the intra picture size and the service bit rate is a weighting factor over this distance. With high speed trick modes, it is very difficult for the end user to have a satisfactory navigation experience as successively displayed pictures may be from different scenes. If an IPTV service would implement SVC for providing HDTV with an SDTV base layer, the fast forward/rewind fluidity would be considerably improved by relying on base layer intra pictures on fast speed trick modes.

For open Internet video streaming over unmanaged networks, scalable video coding can further bring means for graceful degradation in order to handle both congestion situation and access network constraints. But this may imply video traffic control and, in some cases where this control is not performed either by the end-user device or the video server, these mechanisms may not be acceptable regarding regulation about net neutrality (becomes managed delivery if the adaptation is controlled within the network).

Another use case for the PC decoding, and probably more for the open Internet video, is about considering whether the display is performed in a windowed mode or a full screen one. For the first case, it would be useful if the bit stream to be decoded could be adapted either in order to minimize the used resources (PC processing, bandwidth) or to fit within the available resources (as these resources might be used for competing video decoding of several streams for either picture-in-picture usage or mosaics of videos). Actually, for the windowed mode, it is not necessary to decode the video at full resolution or quality, a lower one is sufficient in terms of user experience. If full quality is required even in a windowed mode then this can rely on a user action. This use case can be combined with the content portability one between the PC and the PMP as well as the mobile device.

In that context, SNR scalability is probably the most promising feature for QoS management, whereas spatial scalability would be useful for the previously mentioned quality of experience cases.

VI. CONCLUSION

The new technical design of the scalable video coding (SVC) extension of H.264/MPEG4-AVC offers a number of significant improvements. All types of scalability, i.e. spatial, quality (or SNR) and temporal scalability as well as their combination are provided. The structure of the bit stream is similar to a data partitioning with a set of base layer packets being present that offer a backwards-compatible bit stream to H.264/MPEG4-AVC. The decoding operation is always very close to H.264/MPEG4-AVC decoding with the data being distributed over multiple packets.

The SVC design therefore allows for simple adaptation by dropping the corresponding packets providing the feature that the same content is represented at multiple resolutions (spatial or temporal) or multiple qualities. For the experimental results provided in this paper it can be seen that each of the content representations is coded with an efficiency being very close to a direct single-layer coding of this content using H.264/MPEG4-AVC.

It is also estimated that technical complexity of the design of the novel SVC features in Silicon System-on-Chip (SoC) for consumer end devices, will not be a too high burden to prevent commercial exploitations. Of course all infrastructure costs will have to be taken into consideration by service providers to warrant scalable services roll-out. But these costs need to be considered not only at a compression level but also at a more global level taking into account new distributed content delivery networks as well as convergence infrastructure for providing IPTV services.

Based on the technical description of SVC the application of SVC to IPTV services is analyzed. SVC offers benefits for features such as efficient graceful degradation as well as bit rate and format adaptation. As it is described, these features help achieving IPTV content everywhere services with:

- Improved content portability
- Optimized content management and distribution
- Smart management of access network throughput
- Improved quality of service and experience

For these improvements, in particular spatial scalability appears to be the key feature of SVC making it a viable candidate for emerging and future IPTV services.

ACKNOWLEDGMENT

The authors thank Heiko Schwarz and Detlev Marpe at HHI for providing simulation results and many discussions, Tobias Oelbaum for conducting the subjective tests; Daniele Alfonso and Iljia Materic at STMicroelectronics for their insight on SVC implementation complexity evaluation; Maryline Clare, Sylvain Kervadec and Gilles Teniou at Orange Labs for their investigation of IPTV use cases.

REFERENCES

- [1] ITU-T and ISO/IEC JTC 1, "Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video," ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), Nov. 1994.
- [2] ITU-T, "Video Coding for Low Bit Rate Communication," ITU-T Recommendation H.263, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

- [3] ISO/IEC JTC 1, "Coding of Audio-Visual Objects—Part 2: Visual," ISO/IEC 14492-2 (MPEG-4 Visual), Version 1: Apr. 1999, Version 2: Feb. 2000, Version 3: May 2004.
- [4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits and Systems for Video Technology, Special Issue on Scalable Video Coding*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [5] Video Codec for Audiovisual Services at $p \times 64$ kbit/s ITU-T Recommendation H.261, Version 1: Nov. 1990, Version 2: Mar. 1993.
- [6] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [7] "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s—Part 2: Video," ISO/IEC 11 172-2 (MPEG-1 Video), Mar. 1993.
- [8] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," Joint Video Team, Doc. JVT-P014, Poznan, Poland, Jul. 2005.
- [9] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand, "Constrained inter-layer prediction for single-loop decoding in spatial scalability," in *Proc. of ICIP 2005*, Genova, Italy, Sep. 2005.
- [10] A. Segall and G. J. Sullivan, "Spatial scalability," *IEEE Trans. Circuits and Systems for Video Technology*, to be published.
- [11] E. François and J. Vieron, "Interlaced coding in SVC," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1136–1148, Sep. 2008.
- [12] H. Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, and T. Wiegand, "Technical Description of the HHI Proposal for SVC CE1," ISO/IEC JTC 1/SC 29/WG 11, doc. M11244, Palma de Mallorca, Spain, Oct. 2004.
- [13] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, Jul. 2003.
- [14] J. Reichel, H. Schwarz, and M. Wien, Eds., "Joint Scalable Video Model 11 (JSVM 11)," Joint Video Team, Doc. JVT-X202, Geneva, Switzerland, Jul. 2007.
- [15] H. Schwarz and T. Wiegand, "R-d optimized multi-layer encoder control for SVC," in *Proceedings of ICIP'07*, San Antonio, TX, USA, Sep. 2007.
- [16] MPEG Test Subgroup, "Subjective Test Results for the CfP on Scalable Video Coding Technology," ISO/IEC JTC 1/SC29/ WG11, Tech. Rep. N6383, Apr. 2004.
- [17] T. Oelbaum, H. Schwarz, M. Wien, and T. Wiegand, "Subjective quality evaluation of SVC," in *Proceedings of ICIP'08*, San Diego, CA, USA, Sep. 2008.
- [18] H. Schwarz, D. Marpe, and T. Wiegand, "SVC Core Experiment 2.1: Inter-Layer Prediction of Motion and Residual Data," ISO/IEC JTC 1/SC 29/WG 11, doc. M11043, Redmond, WA, USA, Jul. 2004.
- [19] H. Schwarz, D. Marpe, and T. Wiegand, "Constrained inter-layer prediction for single-loop decoding in spatial scalability," in *Proc. of ICIP'05*, Genoa, Italy, Sep. 2005.
- [20] H. Schwarz, D. Marpe, and T. Wiegand, "Further results on constrained inter-layer prediction," Joint Video Team, Doc. JVT-O074, Busan, Korea, Apr. 2005.
- [21] H. Schwarz, D. Marpe, and T. Wiegand, "Independent parsing of spatial and CGS layers," Joint Video Team, Doc. JVT-S069, Geneva, Switzerland, Mar. 2006.



Thomas Wiegand (M'05–SM'08) is Professor of Electrical Engineering at the Technical University of Berlin chairing the Image Communication group and is jointly heading the Image Processing department of the Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute, Berlin, Germany.

He received the Dipl.-Ing. degree in Electrical Engineering from the Technical University of Hamburg-Harburg, Germany, in 1995 and the Dr.-Ing. degree from the University of Erlangen-Nuremberg, Germany, in 2000. He joined the Heinrich Hertz

Institute in 2000 as the head of the Image Communication group in the Image Processing department and remains also active in this role. His research interests include video processing and coding, multimedia transmission, semantic image representation, as well as computer vision and graphics.

From 1993 to 1994, he was a Visiting Researcher at Kobe University, Japan. In 1995, he was a Visiting Scholar at the University of California at Santa Barbara, USA. From 1997 to 1998, he was a Visiting Researcher at Stanford University, USA and served as a consultant to 8×8 , Inc., Santa Clara, CA, USA. From

2006-2008, he served as a consultant to Stream Processors, Inc., Sunnyvale, CA, USA. He is currently a member of the technical advisory boards of the two start-up companies Vidyo, Inc., Hackensack, NJ, USA and Skyfire, Inc., Mountain View, CA, USA.

Since 1995, he is an active participant in standardization for multimedia with successful submissions to ITU-T VCEG, ISO/IEC MPEG, 3GPP, DVB, and IETF. In October 2000, he was appointed as the Associated Rapporteur of ITU-T VCEG. In December 2001, he was appointed as the Associated Rapporteur / Co-Chair of the JVT. In February 2002, he was appointed as the Editor of the H.264/AVC video coding standard and its extensions (FRExt and SVC). In January 2005, he was appointed as Co-Chair of MPEG Video.

In 1998, he received the SPIE VCIP Best Student Paper Award. In 2004, he received the Fraunhofer Award for outstanding scientific achievements in solving application related problems and the ITG Award of the German Society for Information Technology. In 2008 and 2009, he received the Primetime and Daytime Emmy Engineering Awards, respectively, that was awarded to the JVT/VCEG/MPEG standards committee by the Academies of Television Arts & Sciences for development of H.264/MPEG-4 AVC, for which Wiegand served as associated rapporteur/co-chair, editor, and technical contributor. In 2009, he received the EURASIP Group Technical Achievement Award "for active contributions to video coding research and standardization activities". Since January 2006, he is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.



Ludovic Noblet is the head of the advanced video compression lab at Orange.

He received the Electronics and Computing Systems Dipl.-Ing. from the "Ecole Polytechnique de l'Université de Nantes", France in 1992.

He started his career at Alcatel where he was in charge of introducing internet technologies within the Alcatel private network (Alcanet).

He then moved at Thomson Corporate Research in October 1994 where he has been designing and contributing to the development and success of three

successive generations of MPEG2 encoders and one generation of MPEG2 decoder. He was appointed Head of Applicative Video Compression Cores at Thomson Grass Valley (formerly Nextream) in 1997 and played key roles

in accompanying, as expert, Thomson Marketing and Sales teams. He proactively contributed to the very first IPTV field trials in 2001 and, in 2002, he started working on very first H.264/MPEG4-AVC encoding implementations for Thomson's first generations of SDTV and HDTV H.264/MPEG4-AVC encoders. He was member of the Thomson's H.264/MPEG4-AVC Mustang compression chip team in 2004.

He joined France Telecom in 2004 as an IPTV architect and senior technical advisor for the introduction of H.264/MPEG4-AVC and high-definition within the Orange TV service. In his role, he was also involved in anticipation activities and investigated optimized distribution schemes including the use of SVC for IPTV.

In September 2006, he was appointed head of the "Advanced Video Compression" team at Orange Labs and acknowledged as senior expert in video compression for services development and advice.

Since December 2004, he is the Orange representative at DVB in commercial and technical ad hoc groups for defining the use of A/V codecs in DVB applications. He represents Orange at the Open IPTV Forum for the same topics.



Fabrizio Simone Rovati received Electronic Engineering degree from Politecnico of Milano in 1996. He completed an MBA at Alma graduate school, University of Bologna in 2006. He has since 1995 joined STMicroelectronics (Bristol, U.K. and later Agrate B.za, Italy) working on algorithms and architectures for digital video encoding and decoding systems, and streaming technologies for adaptive and robust multimedia delivery. He is currently leading the Networked Multimedia research team in Advanced System Technology Labs, STMicroelectronics' corporate system R&D group.

During his career he has authored or co-authored 16 British, European and U.S. granted patents, and more than 30 international publications in conferences or technical journals. He has been contract professor at Politecnico of Pavia University during academic year 2001-02, teaching "Digital Electronics II". He gave several lectures at Politecnico of Milan on digital video compression algorithms and architectures.