

A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and Performance

Iraide Unanue¹, Iñigo Urteaga², Ronaldo Husemann³, Javier Del Ser⁴,
Valter Roesler⁵, Aitor Rodríguez⁶ and Pedro Sánchez⁷

^{1,2,4}TECNALIA RESEARCH & INNOVATION, P. Tecnológico, Zamudio,

^{3,5}UFRGS - Instituto de Informática. Av. Bento Gonçalves, Porto Alegre,

^{6,7}IKUSI-Ángel Iglesias, S. A., Paseo Miramón, Donostia-San Sebastian

^{1,2,4,6,7}Spain

^{3,5}Brazil

1. Introduction

The evolution of digital video technology and the continuous improvements in communication infrastructure is propelling a great number of interactive multimedia applications, such as real-time video conference, web video streaming and mobile TV, among others. The new possibilities on interactive video usage have created an exigent market of consumers, which demands the best video quality wherever they are and whatever their network support is (Schwarz et al., 2006). On this purpose, the transmitted video must match the receiver's characteristics such as the required bit rate, resolution and frame rate, thus aiming to provide the best quality subject to receiver's and network's limitations. Besides, the same link is often used to transmit to either restricted devices such as small cell phones, or to high-performance equipments, e.g. HDTV workstations. In addition, the stream should adapt to wireless lossy networks (Ohm, 2005). Based on this reasoning, these heterogeneous and non-deterministic networks represent a great problem for traditional video encoders which do not allow for on-the-fly video streaming adaptation.

To circumvent this drawback, the concept of scalability for video coding has been lately proposed as an emergent solution for supporting, in a given network, endpoints with distinct video processing capabilities. The principle of a scalable video encoder is to break the conventional single-stream video in a multi-stream flow, composed by distinct and complementary components, often referred to as *layers* (Huang et al., 2007). Figure 1 illustrates this concept by depicting a transmitter encoding the input video sequence into three complementary layers. Therefore, receivers can select and decode different number of layers – each corresponding to distinct video characteristics – in accordance with the processing constraints of both the network and the device itself.

The layered structure of any scalable video content can be defined as the combination of a base layer and several additional enhancement layers. The base layer corresponds to the lowest supported video performance, whereas the enhancement layers allow for the refinement of

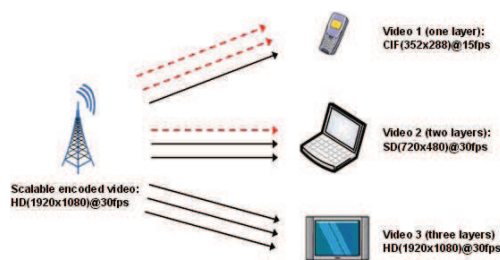


Fig. 1. Adaptation in scalable video encoding.

the aforementioned base layer. The adaptation is based on a combination within the set of selected strategies for the spatial, temporal and quality scalability (Ohm, 2005).

In the last years, several specific scalable video profiles have been included in video codecs such as MPEG-2 (*MPEG-2 Video*, 2000), H.263 (*H.263 ITU-T Rec.*, 2000) and MPEG-4 Visual (*MPEG-4 Visual*, 2004). However, all these solutions present a reduced coding efficiency when compared with non-scalable video profiles (Wien, Schwarz & Oelbaum, 2007). As a consequence, scalable profiles have been scarcely utilized in real applications, whereas widespread solutions have been strictly limited to non-scalable single-layer coding schemes. In October 2007, the scalable extension of the H.264 codec, also known as H.264/SVC (Scalable Video Coding) (*H.264/SVC*, 2010), was jointly standardized by ITU-T VCEG and ISO MPEG as an amendment of the H.264/AVC (Advanced Video Coding) standard. Among several innovative features, H.264/SVC combines temporal, spatial and quality scalabilities into a single multi-layer stream (Rieckl, 2008).

To exemplify the temporal scalability, Figure 2(a) presents a simple scenario where the base layer consists of one subgroup of frames and the enhancement layer of another. A hypothetical receiver in a slow-bandwidth network would receive only the base layer, hence producing a jerkier video (15 frames per second, hereafter labeled as *fps*) than the other. On the contrary, the second receiver (that would benefit from a network with higher bandwidth) would be able to process and combine both layers, thus yielding a full-frame-rate (30 *fps*) video and ultimately a smoother video reproduction. Thereafter, Figure 2(b) illustrates an example of spatial scalability, where the inclusion of enhancement layers increases the resolution of the decoded video sample. As shown, the more layers are made available to the receiver, the higher the resolution of the decoded video is. Finally, Figure 2(c) show the concept of quality scalability, where the enhancement layers improve the SNR quality of the received video stream. Once again, the more layers the receiver acquires, the better the user's quality of experience is.

On top of the benefits of the above introduced scalabilities, there are several other advantages furnished by H.264/SVC. One of such remarkable features of H.264/SVC is the support for video bit rate adaptation at NAL (Network Application Layer) packet level, which significantly increases the flexibility of the video encoder. Alternative scalable solutions, however, only support adaptation at the level of slices or entire frames (Huang et al., 2007). Furthermore, H.264/SVC improves the compression efficiency by incorporating an enhanced and innovative mechanism for inter-layer estimation, called ILP (Inter-Layer Prediction). ILP reuses inter-layer motion vectors, intra texture and residue information among subsequent layers (Husemann et al., 2009).

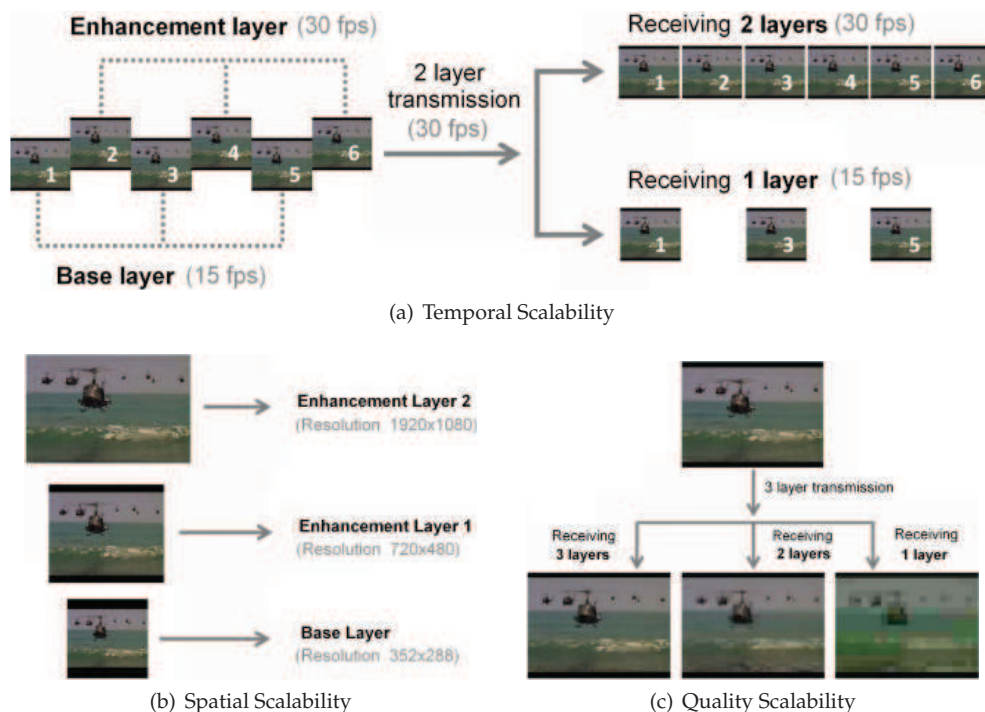


Fig. 2. Illustrative example of scalability approaches in H.264/SVC.

As a consequence of all these aspects, the H.264/SVC standard is currently considered the state-of-the-art of scalable video codecs. As opposed to prior video codecs, H.264/SVC has been designed as a flexible and powerful scalable video codec, which provides – for a given quality level – similar compression ratios at a lower decoding complexity with respect to its non-scalable single-layer counterparts. So as to corroborate this design principle, let us briefly compare H.264/SVC to non-scalable profiles of previous codecs, namely, MPEG-4 Visual (MPEG-4 Visual, 2004), H.263 (H.263 ITU-T Rec., 2000) and H.264/AVC (H.264/AVC, 2010). Codec performance has been analyzed in terms of both compression efficiency and video quality (focusing on the Peak Signal-to-Noise Ratio PSNR of the luminance component). In this analysis, three different video sequences (further details of these video sequences are included in Section 3) have been encoded, based on equivalent configurations and appropriate bit rates for each one, with the following implementations of the aforementioned codecs: H.263 (*Ffmpeg project*, 2010), MPEG-4 Visual (*Ffmpeg project*, 2010) and H.264/AVC (*JVT reference software*, 2010).

As shown in Figure 3(a), the real encoded file size is different for each codec, even if the same theoretical encoding bit rate has been set. The reason for this dissimilarity lies on the performance of the tested codec implementations, which loosely adjust the encoding process to the specified bit rate. From both Figures 3(a) and 3(b), it is clear that H.264/SVC and H.264/AVC are those codecs generating the lowest file size while achieving similar quality (e.g. 36.61 dB by H.264/AVC and 36.41 dB by H.264/SVC for the CREW video

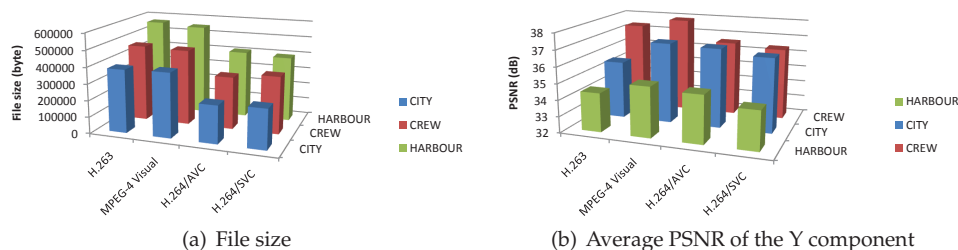


Fig. 3. Performance of different codecs over several video sequences.

sequence). Based on these simulations, it is concluded that H.264/SVC outperforms previous non-scalable approaches, by supporting three types of scalabilities at a high coding efficiency. These results not only evaluate the theoretical behavior of each analyzed codec, but also elucidate the outstanding performance of H.264/SVC with respect to other coding approaches when applied on a given video sample.

In this line of research, this chapter delves into the roots of H.264/SVC by analyzing, through practical experiments, its tradeoff between quality, coding efficiency and performance. First, Section 2 introduces the reader to the details of the H.264/SVC standard by thoroughly describing the functional structure of a H.264/SVC encoder and its supported scalabilities. Next, several applied experiments are provided in Section 3 in order to evaluate the real requirements of a practical H.264/SVC video coding solution. These experiments have all been performed using the official H.264/SVC reference implementation: the JSVM (Joint Scalable Video Model) software (*JSVM reference software*, 2010). Obviously, the scalable nature of this new video coding standard requires a rigorous analysis of its temporal, spatial and quality processing capabilities. Consequently, three scenarios of experiments have been defined to specifically address each type of scalability:

- First, Subsection 3.1 presents the scenario utilized for evaluating the temporal scalability, where the effects of the GOP (Group of Pictures) size parameter and the frame structure are analyzed on practical H.264/SVC encoding procedures. Since the arrangement of the frames within a GOP impacts directly on the performance of the video codec, it is deemed essential to evaluate the advantages and disadvantages of different GOP sizes and structures in the overall encoding and decoding process (Wien, Schwarz & Oelbaum, 2007).
- A second scenario is next included in Subsection 3.2 aimed at evaluating the spatial scalability of H.264/SVC. This subsection analyzes the performance of both video encoder and decoder, emphasizing on distinct relations between screen resolutions of consecutive video layers. Two main algorithms are supported by H.264/SVC: the traditional dyadic solution (only when a resolution ratio of 2:1 among consecutive layer is used) or non-dyadic solution (when any other resolution ratio is possible).
- Subsection 3.3, which comprises the third scenario, analyzes the quality scalability of the H.264/SVC over different configurations. First, the fidelity of the H.264/SVC codec is examined by focusing on the influence of the quantization parameter and the relationship between quality enhancement layers. Besides, the evaluation of the coding efficiency of the H.264/SVC prediction structure between quality layers is also covered. This subsection

concludes by presenting a practical comparison between coarse and medium quality granularity.

Subsequently in Subsection 3.4, other equally-influential features of this scalable codec are scrutinized. On one hand, this final set of experiments investigate the complexity load rendered by different motion-search algorithms and related configurations on practical video encoding procedures. Particularly, the influence in the prediction module of relevant parameters such as the search-window size and the block-search algorithm is evaluated. On the other hand, the benefits of applying distinct deblocking filter types in the encoding and decoding process is examined. Deblocking filters are applied to block-coding based techniques to blocks within slices, looking for the prediction performance improvement by smoothing potentially sharp edges formed between macroblocks (Marpe et al., 2006). Finally, this subsection concludes with the evaluation of the Motion-Compensated Temporal pre-processing Filter (MCTF) included in the H.264/SVC standard.

Based on all the results presented through the chapter, optimized H.264/SVC configurations are suggested in Section 4. These configurations are specifically designed to improve either the efficiency of the encoder or the encoded video quality, which yield significant gains when compared to conventional H.264/SVC solutions. Finally, Section 5 brings up our final considerations.

2. Overview of H.264/SVC

The sophisticated architecture of the H.264/SVC standard is particularly designed to increase the codec capabilities while offering a flexible encoder solution that supports three different scalabilities: temporal, spatial and SNR quality (Wien, Cazolot, Graffunder, Hutter & Amon, 2007). Figure 4 illustrates the structure of a H.264/SVC encoder for a basic two-spatial-layer scalable configuration.

In H.264/SVC, each spatial dependency layer requires its own prediction module in order to perform both motion-compensated prediction and intra prediction within the layer. Besides, there is a SNR refinement module that provides the necessary mechanisms for quality scalability within each layer. The dependency between subsequent spatial layers is managed by the inter-layer prediction module, which can support reusing of motion vectors, intra texture or residual signals from inferior layers so as to improve compression efficiency. Finally, the scalable H.264/SVC bitstream is merged by the so-called multiplex, where different temporal, spatial and SNR levels are simultaneously integrated into a single scalable bitstream.

The following subsections present each scalability type individually, describing their features according to the standardized specifications of the H.264/SVC video codec.

2.1 Temporal scalability

The term “temporal scalability” refers to the ability to represent video content with different frame rates by as many bitstream subsets as needed (Figure 2(a)). Encoded video streams can be composed by three distinct type of frames: I (intra), P (predictive) or B (Bi-predictive). I frames only explore the spatial coding within the picture, i.e. compression techniques are applied to information contained only inside the current picture, not using references to any other picture. On the contrary, both P and B frames do have interrelation with different pictures, as they explore directly the dependencies between them. While in P frames inter-picture predictive coding is performed based on (at least) one preceding reference

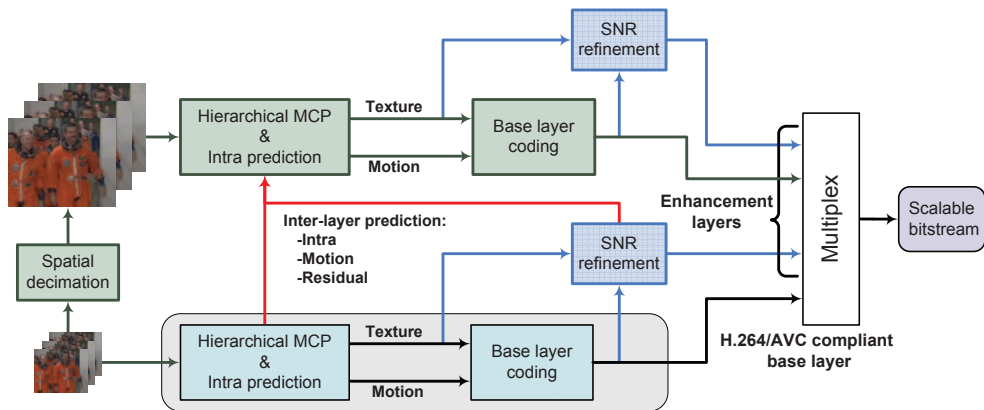


Fig. 4. Block diagram of a H.264/SVC encoder for two spatial layers.

picture, B frames consist of a combination of inter-picture bi-predictive coding (i.e. samples of both previous and posterior reference pictures are considered for the prediction). In addition, the H.264 standard family requires the first frame to be an Instantaneous Decoding Refresh (IDR) access unit, which corresponds to the union of one I frame with several critical non-data related information (e.g. the set of coding parameters). Generally speaking, the GOP structure specifies the arrangement of those frames within an encoded video sequence.

Certainly, the singular dependency and predictive characteristics of each frame type imply divergent coded video stream features. In previous scalable standards (e.g. MPEG-2, H.263 and MPEG-4 Visual), the temporal scalability was basically performed by segmenting layers according to different frame types. For example, a video composed by a traditional "IBBP" format (one I frame followed by two B frames and one P frame) could be used to build three temporal layers: base layer (L_0) with I frames, first enhancement layer (L_1) with P frames and the second enhancement layer (L_2) with B frames. This dyadic approach (2:1 decomposition format) has been proven to be functional, although it provides limited bandwidth flexibility (i.e. the total bit rate required by I frames is significantly larger than that of P and B frames (Rieckl, 2008)). By contrast, in H.264/SVC the basis of temporal scalability is found on the GOP structure, since it divides each frame into distinct scalability layers (by jointly combining I, P and B frame types). As for the H.264/SVC codec, the GOP definition can be rephrased as the arrangement of the coded bitstream's frames between two successive pictures of the temporal base layer (Schwarz et al., 2007). It is important to recall that the frames of the temporal base layer do not necessarily need to be an I frame. Actually, only the first picture of a video stream is strictly forced to be coded as an I frame and to be included in the initial IDR access unit.

In order to increase the flexibility of the codec, the H.264/SVC standard defines a distinct structure for temporal prediction, where reference frames for each video sequence are reorganized in a hierarchical tree scheme. This tree scheme improves the distribution of information between consecutive frames and allows for both a dyadic and a non-dyadic temporal scalability. Figure 5(a) exemplifies this hierarchical temporal decomposition for a 2:1 frame rate relation in a four-layer encoded video. In this example, the base layer L_0 , which is constituted by I or P frames, permits to reconstruct one picture per GOP. The first enhancement layer L_1 , usually composed by B frames, extracts one additional picture per

GOP in addition to that of L_0 . The second enhancement layer L_2 , which is comprised by B frames, further extracts two additional pictures per GOP jointly with those of previous layers. Finally, the third enhancement layer L_3 allows recovering eight pictures.

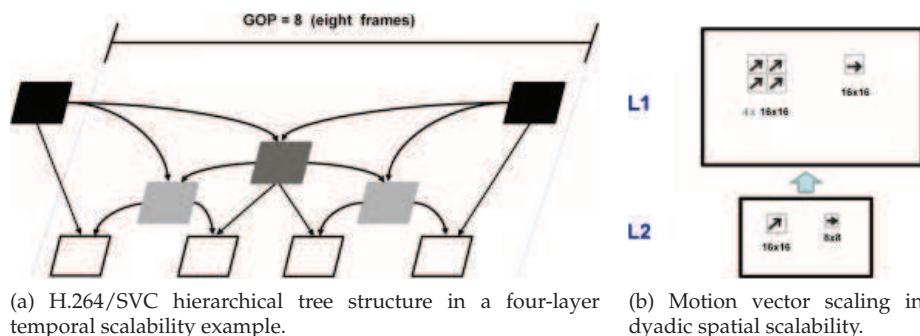


Fig. 5. Graphical support examples for H.264/SVC temporal and spatial scalabilities.

On top of this, H.264/SVC suggests the inclusion of a pre-processing filter before the motion prediction module, which can improve the data information distribution and eliminate redundancies between consecutive layers. The proposed algorithm is referenced as MCTF. This additional filter, when applied over the original data, performs motion aligned decomposition processing. As a result, the correlation between filtered layers is improved, while the overall complexity of the encoder is increased (Schafer et al., 2005).

2.2 Spatial scalability

The spatial scalability is based on representing, through a layered structure, videos with distinct resolutions, i.e. each enhancement layer is responsible for improving the resolution of lower layers (as in Figure 2(b)). The most common configuration (i.e. dyadic) adopts the 2:1 relation between neighbor layers, although H.264/SVC also contemplates non-dyadic ratios (Segall & Sullivan, 2007). This last solution demands the inclusion of a new class of algorithm called Extended Spatial Scalability (ESS) (Huang et al., 2007).

The approaches of previous scalable encoders basically consist of reusing motion prediction information from lower layers in order to reduce the global stream size. Unfortunately, the image quality obtained by this methodology is quite limited. On the contrary, and in order to improve its efficiency, the H.264/SVC encoder introduces a more flexible and complex prediction module called Inter-Layer Prediction (ILP). The main goal of the ILP module is to increase the amount of reused data in the prediction from inferior layers, so that the reduction of redundancies increases the overall efficiency. To this end, three prediction techniques are supported by the ILP module:

- **Inter-Layer Motion Prediction:** the motion vectors from lower layers can be used by superior enhancement layers. In some cases, the motion vectors and their attached information must be rescaled (see Figure 5(b)) so as to adjust the values to the correct equivalents in higher layers (Husemann et al., 2009).
- **Inter-Layer Intra Texture Prediction:** H.264/SVC supports texture prediction for internal blocks within the same reference layer (intra). The intra block predicted in the reference layer can be used for other blocks in superior layers. This module up-samples the

resolution of inferior layer's texture to superior layer resolutions, subsequently calculating the difference between them.

- **Inter-Layer Residual Prediction:** as a consequence of several coding process observations, it has been identified that when two consecutive layers have similar motion information, the inter-layer residues register high correlation. Based on this, in H.264/SVC the inter-layer residual prediction method can be used after the motion compensation process to explore redundancies in the spatial residual domain.

Supplementarily, the H.264/SVC standard supports any resolution, cropping and dimensional aspect relation between two consecutive layers. For instance, a certain layer may use SD resolution (4:3 aspect), while the next layer is characterized by HD resolution (16:9 aspect) (Schafer et al., 2005). The most flexible solution, which does not use a dyadic relation, is called ESS (Extended Spatial Scalability), where any relation between consecutive layers is supported.

2.3 SNR scalability

The SNR scalability (or quality scalability) empowers transporting complementary data in different layers in order to produce videos with distinct quality levels. In H.264/SVC, SNR scalability is implemented in the frequency domain (i.e. it is performed over the internal transform module). This scalability type basically hinges on adopting distinct quantization parameters for each layer. The H.264/SVC standard supports three distinct SNR scalability modes (Rieckl, 2008):

- **Coarse Grain Scalability (CGS):** in this strategy (Figure 6(a)), each layer has an independent prediction procedure (all references have the same quality level) in a similar fashion to the SNR scalability of MPEG-2. In fact, the CGS strategy can be regarded as a special case of spatial scalability when consecutive layers have the same resolution (Huang et al., 2007).
- **Medium Grain Scalability (MGS):** the MGS approach (Figure 6(b)) increases efficiency by using a more flexible prediction module, where both types of layer (base and enhancement) can be referenced. However this strategy can induce a drifting effect (i.e. it can introduce a synchronism offset between the encoder and the decoder) if only the base layer is received. To solve this issue, the MGS specification proposes the use of periodic key pictures, which immediately resynchronizes the prediction module.
- **Fine Grain Scalability (FGS):** this version (Figure 6(c)) of the SNR scalability aims at providing a continuous adaptation of the output bit rate in relation to the real network bandwidth. FGS employs an advanced bit-plane technique where different layers are responsible for transporting distinct subsets of bits corresponding to each data information. The scheme allows for data truncation at any arbitrary point in order to support the progressive refinement of transform coefficients. In this type of scalability, only the base layer casts motion prediction techniques.

As a means to understand each SNR scalability granularity mode of H.264/SVC, the internal correlation between layers for a two-layer video stream can be observed in Figure 6. Note that the black frames in Figure 6(b) represent key pictures with periodicity of 4 pictures.

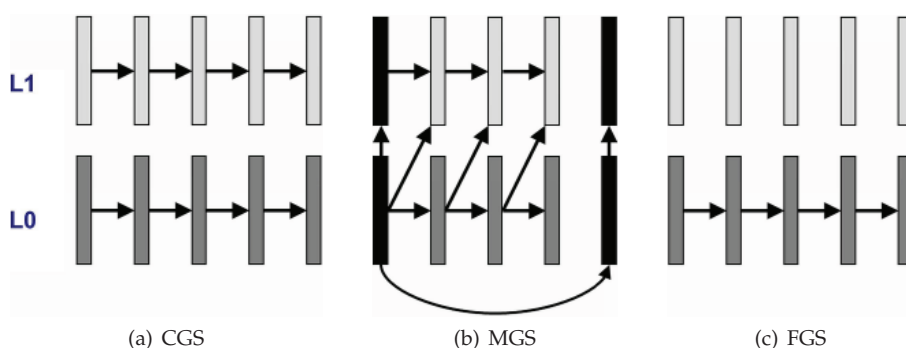


Fig. 6. H.264/SVC SNR scalability granularity mode for a two-layer example.

3. Performance experiments

Heretofore this tutorial has introduced the H.264/SVC video coding standard and its pivotal underlying concepts. This section delves into the description of several experiments evaluating the requirements of a practical H.264/SVC solution. As a consequence of the standardization process of H.264, the different entities involved in it (including the industry members, the ITU-T body and MPEG) formed the so-called Joint Video Team (JVT) which, among various duties, has developed the official H.264/SVC reference code. This reference implementation of the codec, coined as JSVM, undergoes continuous developments so as to track the numerous features of this standard. For the purpose of the experiments later detailed, JSVM version 9.19.4 (*JSVM reference software*, 2010) has been used, which even if not necessarily efficient or optimized, guarantees full compliance with the standard. Since the goal of this section is to provide an overview of the practical characteristics of this scalable codec, it is considered mandatory to tackle every tests from a generic video-sample-agnostic approach. Consequently, experiments have been repeated with different video sequences, thus the performance of the codecs is evaluated over video samples of diverse characteristics: miscellaneous motion patterns, various spatial complexities, shapes, etc.

Specifically, the tested video samples are the conventional CREW, CITY and HARBOUR sequences (*YUV video repository*, 2010). These video sequences cover a wide range of dynamism scales: CREW presents a spatial craft crew walking quickly (i.e. constant object movement); CITY is a 360-degree view of a skyscraper recorded by a slow-motion camera (slow panning motion); finally, HARBOUR shows the filming from a fixed camera in a sailboat race (high dynamism). In addition to the different attributes of each video sequence, diverse resolutions and frame rates have been further considered: 176x144 pixels (QCIF) at 15 fps, 352x288 pixels (CIF) at 30 fps and 704x576 pixels (4CIF) at 60 fps.

For the performance evaluation of the H.264/SVC codec, the following metrics have been used for all the experiments (unless specifically indicated): encoding complexity (measured as the time in seconds required to encode a 10-second video sample), encoding efficiency (defined as the size of the encoded video sequence), decoding complexity (as the number of seconds to decode a 10-second encoded video sequence) and, finally, the objective video-quality resulting from the encoding and decoding process (i.e. the PSNR value of the luma component of the video sequence). The description, results and conclusions of the

different experiments provided in the following sections permit to evaluate the key features of H.264/SVC.

3.1 Temporal scalability

As explained in Section 2.1, the frame structure imposed on the GOP (Group of Pictures) is essential not only for the temporal scalability offered by this scalable codec, but also for the features of the resulting video stream. In fact, changing the GOP size directly affects the number of temporal layers contained in the encoded bitstream. For example, in a temporal dyadic approach, a video stream encoded with GOP size equal to 16 generates the following five temporal layers: T_0 (1 frame per GOP), T_1 (2 frames per GOP), T_2 (4 frames per GOP), T_3 (8 frames per GOP) and T_4 (16 frames per GOP). However, encoding the same video with GOP size equal to 8 renders four temporal layers: T_0 (1 frame per GOP), T_1 (2 frames per GOP), T_2 (4 frames per GOP) and T_3 (8 frames per GOP). Finally, defining a GOP size of 4 produces only three temporal layers: T_0 , T_1 and T_2 . Therefore, it may be concluded that the flexibility of a temporal scalable solution (in terms of the number of layers) is directly proportional to the selected GOP size. Nevertheless, increasing the GOP size does have some implicit collateral effects: it influences the overall encoding efficiency, as it imposes a variation in the number of I, P and B frames per GOP.

In order to prove this effect, several experiments have been performed by changing the GOP size parameter while the output bit rate is kept constant. Figure 7 show the obtained results in terms of the quality for the upper and base layer.

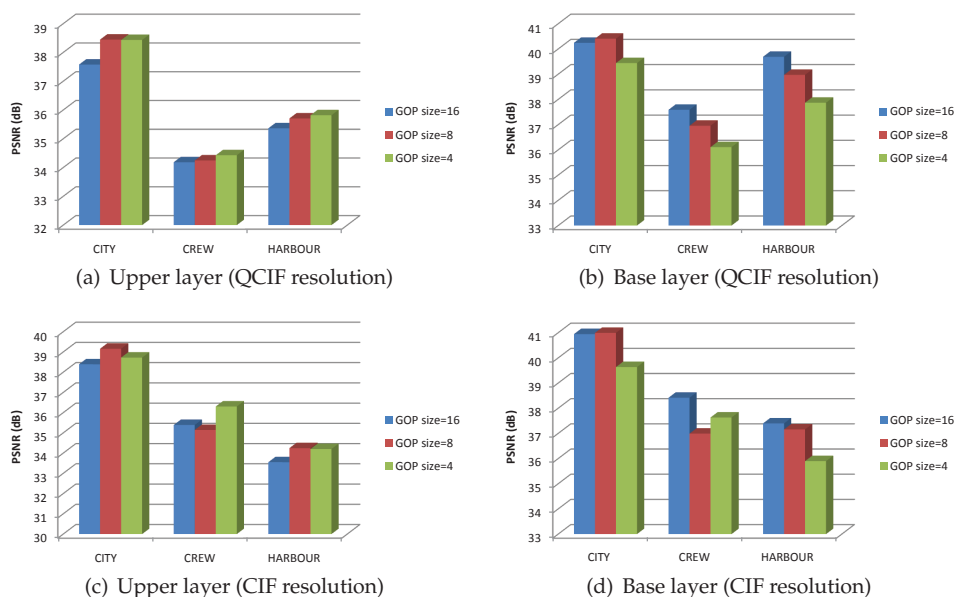


Fig. 7. Impact of the GOP size on the H.264/SVC quality for different video sequences.

By taking a closer look at Figures 7(a) and 7(c) the reader may notice that there is no significant quality difference in the final recovered video (i.e. upper layer) when increasing the GOP size. Nevertheless, the behavior of the quality of the base layer lightly varies depending

on both the particularly used video samples and the selected resolutions, as can be seen in Figures 7(b) and 7(d). An increment of the GOP size entails an increment of the quality of the base layer for CREW-QCIF, HARBOUR-QCIF and HARBOUR-CIF video sequences whereas, for instance, such a direct relation in the CREW-CIF video sample is not so evident. This variability in the quality performance can be, in part, induced by the particularities of the scalable prediction module (H.264/SVC ILP). Theoretically speaking, a GOP size increment should imply a quality improvement, as the number of B frames rises while contributing to an efficient encoding.

On the contrary, the complexity of the encoder is clearly influenced by the GOP size parameter, i.e. the increase in the number of layers (and therefore B frames) implies higher requirements for the encoder prediction module. Such an encoding complexity increase (measured in terms of the encoding execution time) is depicted in Figure 8. For instance, an increment around 20% in encoding time is obtained when comparing GOP sizes of 4 and 16 for the CITY video sequence at QCIF resolution.

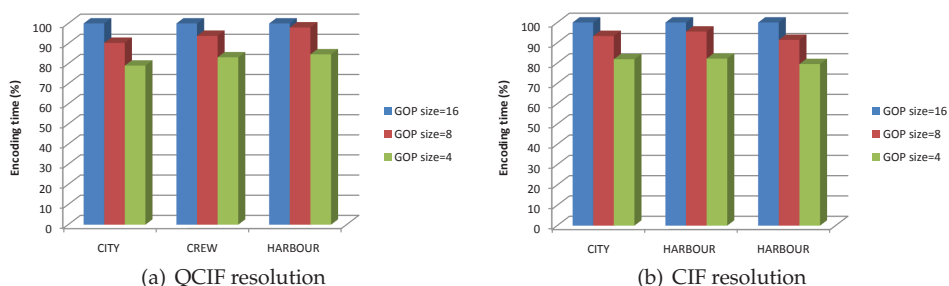


Fig. 8. GOP size impact in H.264/SVC encoding time for different video sequences.

It is also interesting to analyze the advantages of using higher GOP sizes for the temporal scalability, as an increment in the GOP size augmentates the number of available temporal layers and ultimately, enhances the flexibility of the video stream. As aforementioned in Section 2.1, three frames types are generally considered to encode a video picture: I, P and B frames. The difference between those frame types mainly resides on the references used by them for the predictive coding. Certainly, the singular dependency and predictive characteristics of each frame type lead to divergent encoded video stream features. Furthermore, the arrangement of the frames within a GOP directly impacts on the codec performance as well. In this context, Figure 9 shows how different GOP structures influences the encoding and decoding complexity, while maintaining a similar video quality. The evaluated GOP structures are:

- **B**: an initial P frame and 15 consecutive B frames form the GOP structure.
- **B_I**: the GOP is composed by an initial I frame and 15 consecutive B frames.
- **B_IDR**: the GOP arrangement corresponds to an initial IDR frame, followed by 15 B frames.
- **NoB**: only P frames (16) are used in the whole GOP.
- **NoB_I**: the GOP is composed by an initial I frame, followed by 15 P frames.
- **NoB_IDR**: an initial IDR frame followed by 15 P frames form the GOP structure.

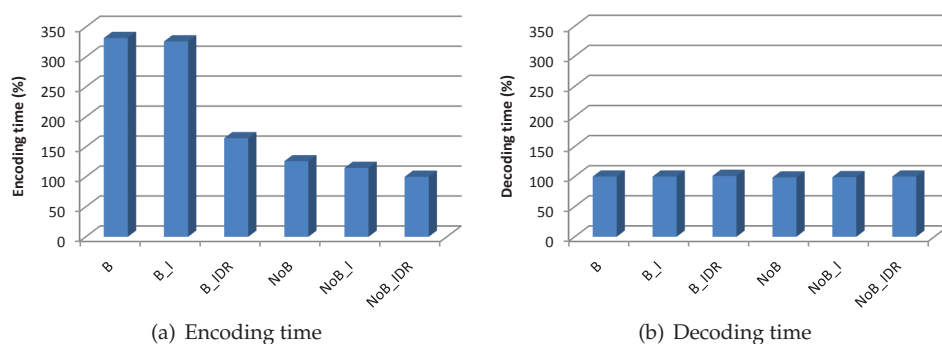


Fig. 9. GOP's structure impact in H.264/SVC codec for the HARBOUR video sequence.

This experiment clearly stresses on the influence of B frames within a GOP, since they impose a significant coding complexity increase. However, their inclusion does not provide any comparable advantage, as quality remains almost equal – differences of less than 0.5 dB were obtained in performed experiments – at the cost of a small bit rate variation. Similar results have been observed for other experiments based on different GOP sizes and video sequences, which are not included here for the sake of space. Regarding the influence of I and IDR pictures, further tests indicate that the quality, complexity and bit rate behaviors are similar for both type of frames. Figure 10 supports this claim for different I and IDR inclusion periods (a stream encoded only with P frames has been employed as a reference).

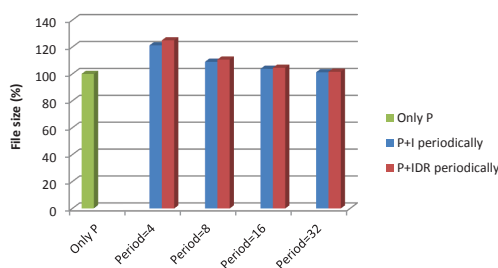


Fig. 10. GOP structure's (I Vs IDR) impact in H.264/SVC codec.

Along with the implications on video bit rate, the determination of the intra-frame frequency also plays an important role when dealing with packet losses in real video streaming applications, which may be due to different phenomena, e.g. congestion, wireless communication losses or handovers (Unanue et al., 2009). As exemplified in Figure 11, video-quality recovery is directly influenced by the GOP structure and particularly, by the reception of an intra-type frame. Due to the intrinsic features of intra-type frames, the sooner an intra-type frame is received, the sooner the video quality is recovered. Based on this rationale and referring to the plotted example, the video quality recovery for H.264/SVC sequences including intra-type frames is much faster (maroon line in Figure 11) than that corresponding to streams without intra-type frames (green line in Figure 11). It is important to remark that with the reception of an intra-type frame, the quality of the received video is almost immediately recovered, whereas the intrinsic dependencies of P and B frames involve

a slower quality recovery when facing losses. In other words, due to the use of a predictive encoding structure, a frame loss not only affects the current GOP, but may have impact in preceding and subsequent GOPs as well.

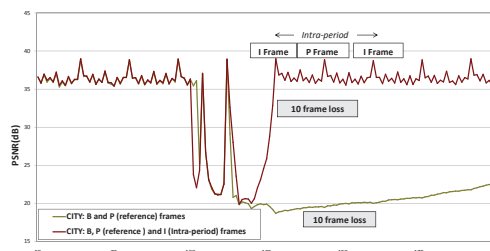


Fig. 11. Frame loss impact on H.264/SVC streams subject to different GOP structures.

Nevertheless, and besides the above proven fact that intra-frames provide faster quality recovery, the speed of video sequence's quality recovery not only depends on the GOP structure, but also on the particular video sequence characteristics. That is, for almost similar frame sequences (e.g. semi-static motion in CITY sequence), the coded P and B frames provide little information with respect to each other. Therefore, in those kinds of motion sequences, it is difficult to recover from the loss of previous frames unless intra-frames are included (Unanue et al., 2009). Consequently, it is deemed crucial to carefully determine the frequency of these type of frames – whether they are I or IDR – which poses a tradeoff between file size and recovery speed: a higher inclusion frequency accelerates the video-quality recovery in lossy environments at a penalty in file size. In summary, granting priority to the bit rate of the stream or to the recovery speed of the video quality is a decision to be taken as a function of the considered scenario. Similarly, the selection between I and IDR frames (or any combination of both) should be also left open to each particular application.

3.2 Spatial scalability

With spatial scalability, different layers within the same encoded video stream contain distinct video resolutions. To support this scalability, motion, texture and residual information from previous layers (after rescaling to the new resolution) can be reused at the H.264/SVC encoder. When the relation between layers is 2:1 (i.e. dyadic case), the rescaling algorithm in a H.264/SVC encoder is rather simple, since in this case the operation to rescale a layer reduces to a simple bit-shift operation. However, H.264/SVC also supports any other resolution ratio between subsequent layers (i.e. non-dyadic cases), for which more complex mathematical operations are necessitated.

In order to determine the real requirements of H.264/SVC's spatial scalability encoding, several practical experiments have been performed varying the resolution ratios between layers. In the first case, a QCIF resolution base layer and a CIF resolution enhancement layer (dyadic scenario) were used. In the second experiment, the enhancement layer is adjusted to 240x112 pixels, while keeping the same base layer (non-dyadic scenario). Please note that in order to simplify the comparison, the output bit rate has been adjusted to the same value in both cases.

On one hand, Figure 12(a) depicts the quality comparison for both experiments, where a slightly higher quality for the dyadic scenario can be observed. This phenomenon is explained by noticing that a 2:1 relation does not produce any rescaling distortion, which does not hold

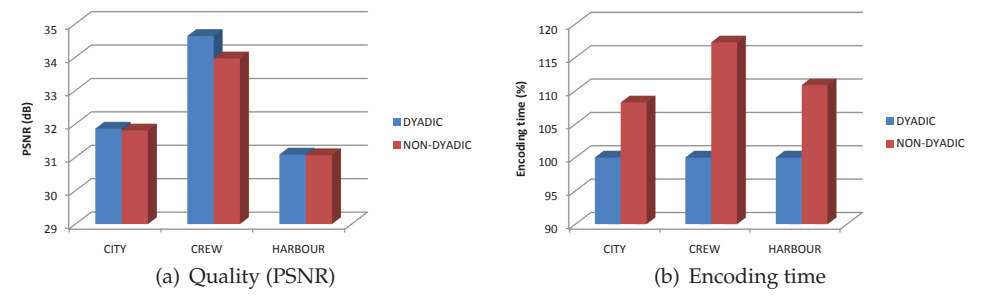


Fig. 12. Spatial scalability evaluation: dyadic and non-dyadic solutions.

for non-integer resolution ratios. On the other hand, when addressing non-dyadic cases the encoder complexity increases significantly, as shown in Figure 12(b). In other words, dyadic configurations can be processed with significant lower encoding time than the non-dyadic ones, e.g. the non-dyadic approach increases the encoding load up to approximately 18% for the CREW video sequence.

3.3 SNR scalability

The SNR scalability implicates several techniques in order to create layers of different quality levels within the same encoded bitstream. In this regard, JSVM provides several options to specify the desired quality not only for each particular layer, but also for the overall encoded stream. First, this subsection focuses on the so-called Quantization Parameter (QP), which is directly related to the quantization process of the original video sequence. Then, the specific properties of two of the distinct SNR scalability modes of H.264/SVC are analyzed, namely, CGS and MGS. The FGS mode has not been included in these experiments since, as opposed to CGS and MGS, it does not allow personal configuration of relevant parameters, such as the number of layers or the value of quantization step per layer.

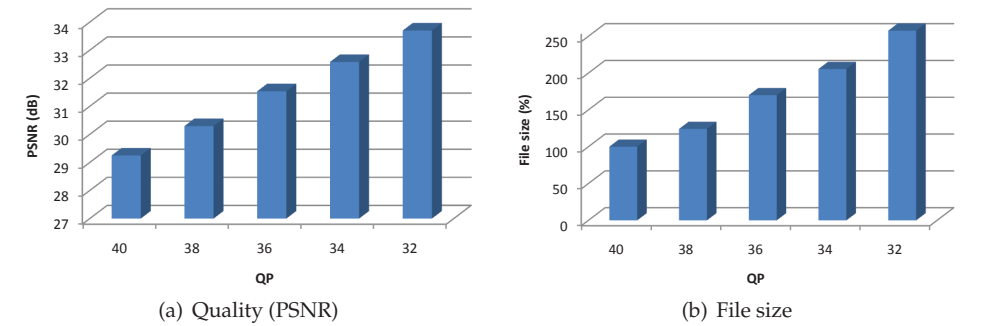


Fig. 13. Evaluation of the SNR scalability: impact of the quantization parameter QP.

In general lower quantization parameter values lead to both better PSNR level and higher bit rate for the encoded video stream. However, during the encoding process, the QP value is not maintained exactly equal for all the frames within the given stream, i.e. it varies slightly depending on the position of each frame within the GOP. The appropriate QP value for each particular scenario or multimedia application should be selected by not only taking into

account the desired quality, but also by analyzing the practical impact of the QP on the file size of the encoded bitstream. On one hand, Figure 13 attests the direct relationship between the selected quantization parameter and the resulting video quality and file size. On the other hand, Figure 14 represents the visual quality incurred when assigning different QP values to the encoding process of the CREW video sample.

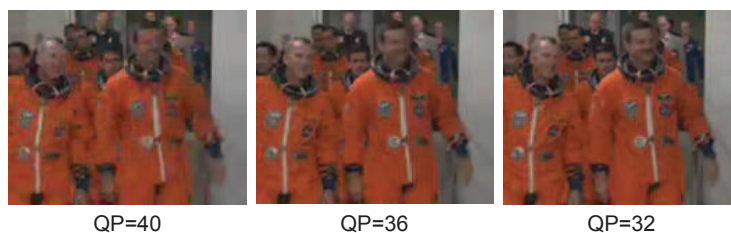


Fig. 14. Quality for different QP-value based H.264/SVC captured pictures (QCIF resolution).

Once the influence of the QP parameter has been explored, a deeper analysis is performed by evaluating the quality scalability intrinsically provided by H.264/SVC. In the following test two SNR scalable layers are incorporated into the encoded stream (lower quality for the inferior layer, QP_L , and better quality for the upper layer, QP_U), since with JSVM an independent QP value can be assigned to each scalable layer. One of the basics of H.264/SVC is the ability to benefit from its inter-layer prediction mechanisms so as to perform efficient scalable encoding. However, there is a close dependency between the selected quality scalabilities and the inter-layer prediction into the resulting video stream, as the experiment results included in Figure 15 clearly show.

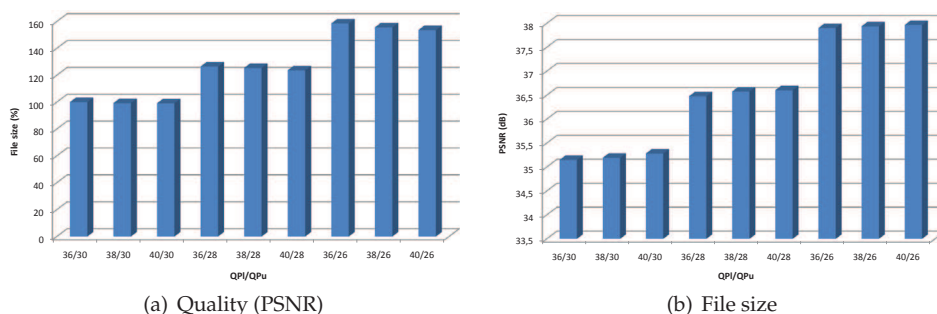


Fig. 15. Evaluation of the dependency between the assigned QP to each SNR scalable layer and the overall quality.

In this example, the quality obtained in the upper layers (defined by QP_U) certainly depends on the quality of the lower layers as specified by QP_L . Referring to Figure 15(a), even if the same QP_U is set, the resulting video quality is slightly different based on the quality of the underlying lower layer. The reason for this phenomenon gravitates on the inter-layer prediction mechanism: since the enhancement layers progressively refine the quality of lower layers, even when the same QP_U is used, the PSNR achieved by the content roughly depends on the quality of lower layers, which is established by the QP_L parameter.

Additional experiments have been carried out to analyze the specific characteristics of H.264/SVC's distinct SNR scalability modes: CGS and MGS. For both experiments, the same configuration for the quantization parameter has been used: $QP_L=39$ for the base layer, and $QP_U=33$ for the enhancement layer. Besides, and in order to simplify the analysis, both modes have been forced to produce the same output bit rate. The results for these experiments are presented in Figure 16, both for video quality and encoding performance metrics. For all evaluated video sequences, the MGS approach produces better quality results, as evidenced in figures 16(a) and 16(b). This interesting result is due to the improved flexibility of MGS's internal prediction algorithm (as more possible references are supported), which contributes to a reduction of matching errors (i.e. residual data). On the other hand, both scalability modes present similar results in terms of codec's performance (encoding execution time).

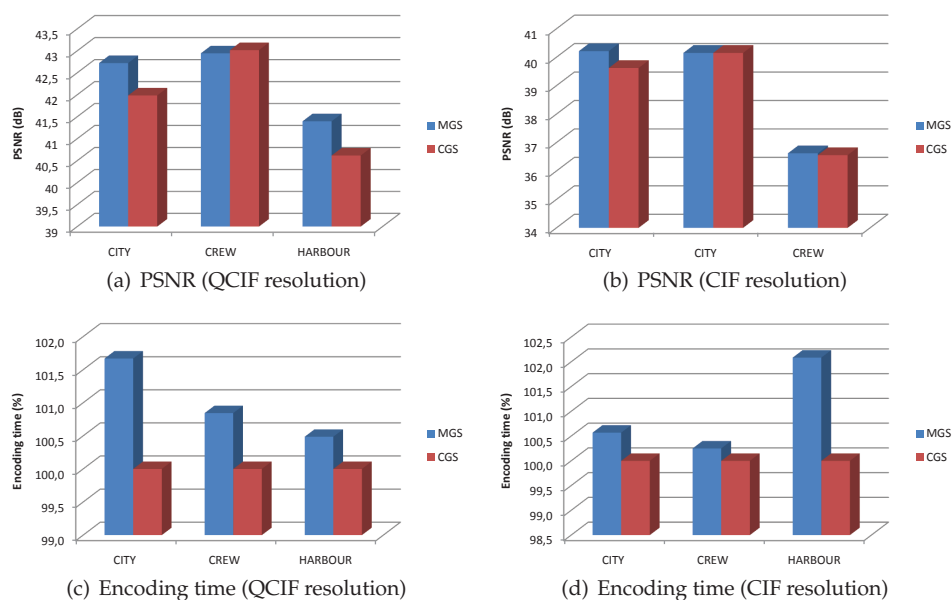


Fig. 16. Comparison between MGS and CGS SNR scalable modes for different resolutions.

3.4 Additional features

Along with its differentiated temporal, quality and spatial scalabilities, the H.264/SVC standard provides several other innovative features, which are subject to practical experimentation through this subsection.

3.4.1 Prediction module

In general, motion estimation techniques stand for those algorithms that allow determining the vectors that describe the correlation between two adjacent frames in a video sequence. In this context, H.264/SVC allows tuning the searching parameters for its motion estimation algorithm: it is possible to decide whether an exhaustive block-searching algorithm or a speed-optimized approach is to be utilized. Furthermore, the search-range of the chosen

block-search function can also be tweaked. However, the exhaustive block-searching function demands a high computational complexity in the encoding process, while its repercussion on the quality and encoding efficiency is not significant. These claims are buttressed by the results of performed experiments given in Table 1. Notice that these results have been generated by encoding QCIF resolution video sequences, since the encoding complexity increases dramatically for higher resolutions. Since video coding quality is comparable for both search-functions (results not shown due to space constraints), it is highly recommended to select the fast-searching algorithm in practical H.264/SVC encoders due to the derived significant reduction in computational load.

A deeper experimental analysis of the searching algorithm is illustrated in Figure 17, where the influence of the search-range parameter is studied for several CIF resolution video sequences. Experimental results verify that the higher the search-range is, the longer the coding time is. No significant impact has been detected in any other metric.

Video sequence	Motion-search algorithm	Search-range	Decoding time (%)
CITY	Fast	Exhaustive	100%
CITY	Exhaustive	Exhaustive	6133,20%
CREW	Fast	Exhaustive	100%
CREW	Exhaustive	Exhaustive	3153,25%
HARBOUR	Fast	Exhaustive	100%
HARBOUR	Exhaustive	Exhaustive	6482,42%

Table 1. Impact of the selected motion-search algorithm in H.264/SVC.

Closely related to the motion compensation, enabling additional 8x8 motion-compensated blocks can notoriously increase the complexity of the encoder. As the experimental results in Figure 18 certify, enabling additional sub-macroblock partitions of 8x8 requires more resources when encoding a given video sequence, whereas it surprisingly has little benefits in the other considered metrics (file size and quality).

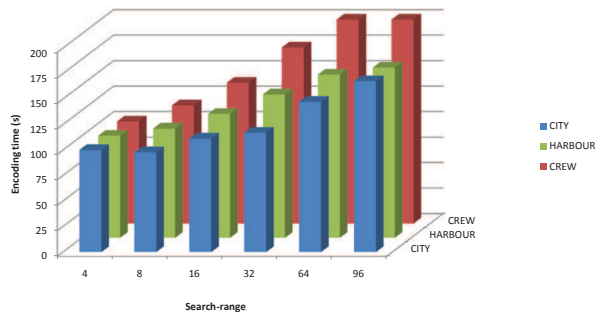


Fig. 17. Search-range parameter impact on H.264/SVC video coding.

Consequently, regarding motion estimation mechanisms in H.264/SVC it is highly recommended to use fast-searching algorithms, small search-ranges, and no additional 8x8 block compensation if the target application requires minimizing the encoder complexity.

3.4.2 Deblocking filter

Within this subsection, the benefits of applying distinct deblocking filter approaches in H.264/SVC video coding have been analyzed. Deblocking filters are exploited in block-coding

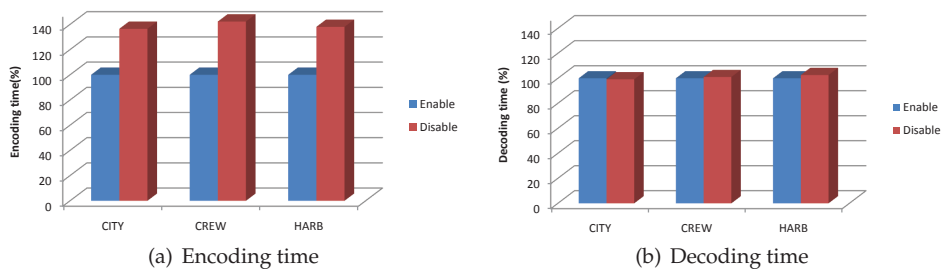


Fig. 18. Impact of enabling additional 8x8 sub-macroblock partitions.

techniques by applying them to blocks within frames, which lead to an improved prediction as they smooth potentially sharp edges between macroblocks. The H.264/SVC deblocking filter operates within the motion-compensated prediction loop, embodying an enhanced quality for the end user (Schwarz et al., 2007).

In these experiments the in-loop deblocking filter and the inter-layer deblocking filter included in the H.264/SVC standard are evaluated. To this end, the following cases have been considered in the JSVM reference software: 1) no filter is applied (LF_0); 2) filter is applied to all block edges (LF_1); 3) two stage filtering where slice boundaries are filtered in the second stage (LF_2); and, finally, 4) two-stage deblocking filtering is applied to the luma component (its frame boundaries are filtered in a second stage), but chroma is not filtered (LF_3). The assessment of the benefits and drawbacks of each of the aforementioned filtering cases has been done, on top of the metrics used heretofore (i.e. encoding/decoding time, encoding efficiency and PSNR), by resorting to the MSU Blocking Metric (*MSU Video Quality Measurement Tool*, 2010). The MSU Blocking Metric measures the frame-to-frame blocking effect in a given video sequence, by detecting object edges with heuristic methods. A higher value of the MSU Blocking Metric corresponds to a better video quality.

The experiments for the analysis of the in-loop deblocking filter have been performed over different video sequences and configurations combining temporal, spatial and SNR scalable layers. Table 2 shows experiment results for one single spatial layer (QCIF resolution) and two quality layers (a similar behavior has been obtained for other combinations). From these extensive tests an interesting conclusion can be extracted: the performance of the in-loop deblocking filter heavily depends on the specific video sequence and the combination of scalable layers. On one hand, the quality obtained when applying each of the tested filtering techniques diverges substantially and hinges, not only on the dynamics and features of the original video sequence, but also on the specific combination of scalabilities in the H.264/SVC encoding process. On the other hand, the coding and decoding complexity of these filters shows a clear dependency on each input video sequence.

Video Sequence	LF_0	LF_1	LF_2	LF_3
CITY	1222159	1175891	1175891	1174807
CREW	1051660	1356914	1356914	1362196
HARB	1208833	1252369	1252369	1251459

Table 2. Impact of selected in-loop deblocking filtering techniques in the performance of H.264/SVC (in terms of average MSU Blocking Metric).

Similarly, the inter-layer deblocking filter has been evaluated over the above mentioned scenarios. The same analysis and procedure has been done and, again, the obtained results have not been conclusive. In this case, the benefit of applying different techniques is not significant and, for the same H.264/SVC encoding configuration, results are tightly coupled to the characteristics of the processed video sequence.

Therefore, the best filtering technique can not be determined beforehand and, for each multimedia application or scenario, a deep analysis needs to be done in order to select the appropriate deblocking filtering technique.

3.4.3 Pre-processing filter

To conclude with this practical section, this set of experiments evaluate the practical impact of including an additional pre-processing filter supported by the H.264/SVC standard: the so-called Motion-Compensated Temporal Filtering. This filter has been suggested as an additional solution to improve data similarity between consecutive layers by mainly helping temporal decomposition. Basically, the MCTF scheme consists of a 2-tap filter based on Haar or 5/3 wavelet transforms (Schafer et al., 2005), which must be applied over the original input video, i.e. before any encoder processing.

Within the JSVM reference platform, this filter is an independent software module (labeled as "MCTFPreProcessorStatic"). It receives as input a raw video sequence (in YUV format), generating a filtered output file. In order to integrate this MCTF module into the encoding process, the original video sequences are first filtered and then fed to the JSVM encoder, which is preconfigured to work with the new filtered files. For this experiment, the output bit rate has been adjusted to the same value in order to simplify the comparison.

Results in Figures 19(a) and 19(b) present the obtained video quality with and without MCTF pre-processing filter. It is doubtlessly proven that the filter produces a small improvement in video quality. In order to further quantify the impact of the inclusion of the MCTF filter in the encoding procedure, the filtering time – the delay caused by the "MCTFPreProcessorStatic" – is added to the JSVM encoding time. The comparative results are presented in Figures 19(c) and 19(d) for CIF and 4CIF resolutions, respectively. It is clearly observed therein how enabling MCTF significantly deteriorates the global performance, increasing the total execution time in more than 300% in all cases.

4. Recommended configurations for practical integration

The experimental results shown in the previous section highlight the practical influence of several H.264/SVC configuration parameters in the performance of the codec. Therefore, the correct setting of these parameters is critical in order to customize practical scalable solutions. Due to the inherent complexity of the H.264/SVC specification, a plethora of variables must be taken into account so as to tailor each configuration to the particular demands and requisites (objective or subjective) of the scalable application at hand. Even if each particular scenario might present specific requirements, the tradeoff between two opposing metrics must be met in most practical applications: to maximize the video quality (disregarding any computational complexity and processing requirements of the codec), or to minimize the encoding complexity with the minimum associated reduction in quality.

On one hand, and based on the results of previous sections, for those applications where quality is more relevant than computational performance (e.g. video storing), the following recommendations have been concluded: an extensive use of B frames (in order to reduce the

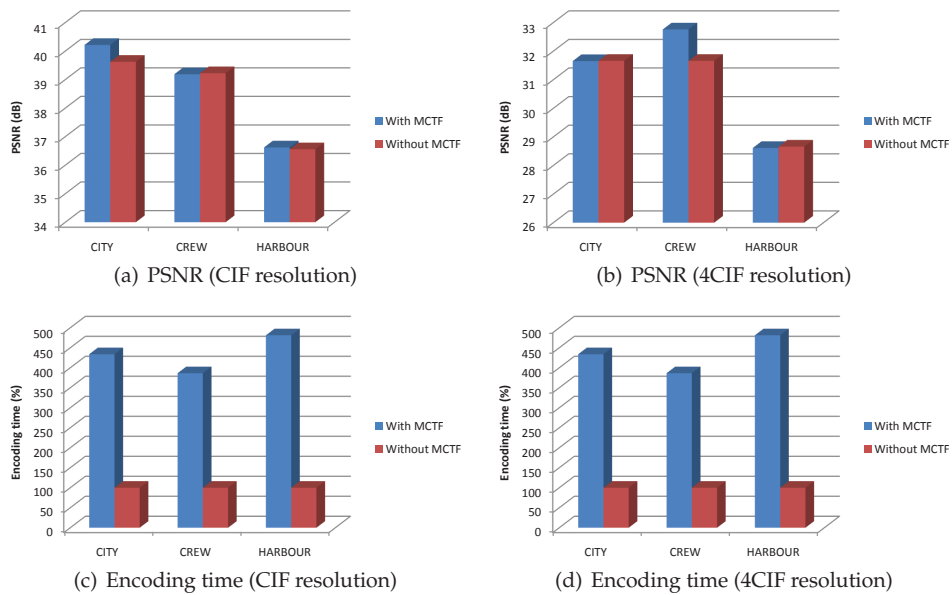


Fig. 19. Impact of enabling MCTF pre-processing filter.

bit rate increment due to the quality requirements), the selection of a high search-area size for inter-layer prediction, the adoption of the MGS mode for the SNR scalability and, finally, setting a sufficiently small quantization parameter. On the other hand, for high-performance scalable applications (e.g. IPTV-based solutions), other configuration schemes are more suitable: small GOP values, I and P frame-based GOP structures, high QP values, the use of fast-searching algorithms, disable additional 8x8 motion-compensated blocks and, when possible, the avoidance of non-dyadic spatial scalability ratios. Moreover, and as a general rule for both cases, the inclusion of the MCTF pre-processing filter is deemed unnecessary, since no quality or performance improvement has been obtained in our experiments. The responsibility for selecting advanced techniques as deblocking filters is left on the application, as their performance strongly depends on the specifically processed video sequence.

In order to illustrate this advice, two experimental scenarios have been defined: a high-quality and a high-performance demanding scalable application. In both experiments, a conventional reference configuration is compared to the proposed advanced approaches. This hereafter coined basic-reference configuration consists of the following configured parameters: GOP size equal to 8 in a "IBBP" frame pattern, ILP with fast-search mode, search-area equal to 48, CGS mode for SNR scalability, $QP_U=32$ for the upper quality layer, and $QP_L=38$ the lowest quality layer.

4.1 High-quality configuration

For this quality-demanding scenario, a hybrid scalable configuration with temporal (4 layers) and SNR (2 layers) scalability has been designed. This high-quality configuration is designed so as to provide a quality improvement with respect to the basic-reference configuration. The key parameters modified for the proposed high-quality configuration are the use of

only B frames, an expanded search-area of 92 and MGS mode for providing SNR scalability. Specifically, the QP values determined for this high-quality configuration are $QP_U=25$ and $QP_L=30$. Please recall that these parameters are just particular examples of the general guidelines provided in this chapter, and might need further tweaking in other real scenarios. The practical results obtained from the evaluation of the two suggested configurations (basic-reference and high-quality) for the three video sequences at CIF resolution are shown in Figure 20. Note that, for the sake of fairness in the comparison, the output bit rate of all configurations has been adjusted to the same value (1 Mbps) in order to evaluate only variations in quality and performance. First, it is important to observe the quality improvement obtained in Figure 20(a) when using the suggested high-quality configuration, with gains up to 2.5 dB in some cases. However, a considerable impact in the global computational performance is obtained for this last configuration (Figure 20(b)): the encoding time increases more than five times in some cases.

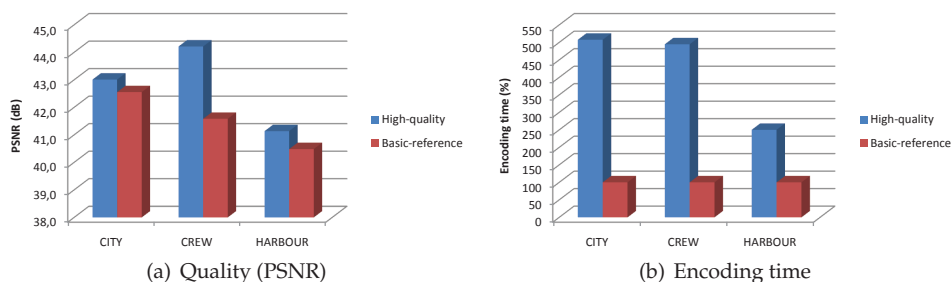


Fig. 20. Comparative between basic-reference and high-quality configurations.

4.2 High-performance configuration

For real-time performance-demanding applications such as widespread video conference systems or video-surveillance systems, the time spent in encoding a video sequence is critical. In such cases, the computational performance of the codec is considered decisive as long as the quality of the video stream does not degrade dramatically. For these applications a high-performance configuration – aimed at achieving fast execution – is proposed with the following parameters: GOP size equal to 4 with "I PPP" structure (one I and three P frames per GOP without including B frames), fast search-mode ILP with search-area reduced to 16, and quantization steps adjusted to $QP_U=36$ and $QP_L=38$. Here again, these specific values are a consequence of the general design guidelines provided throughout this chapter.

When comparing both the basic-reference and the high-performance configurations in terms of quality (Figure 21(a)), observe that the degradation in PSNR varies depending on the encoded video sequence, i.e. the PSNR for the CREW video sequence is almost equal with both configurations, whereas the PSNR for CITY and HARBOUR video sequences decreases approximately down to 1 and 2 dB respectively. However, this drawback finds its counterpart at the noticeable computational performance improvement shown in Figure 21(b), where it is concluded that the encoding time for the high-performance configuration is at least two times faster than the basic-reference solution for all the evaluated video sequences.

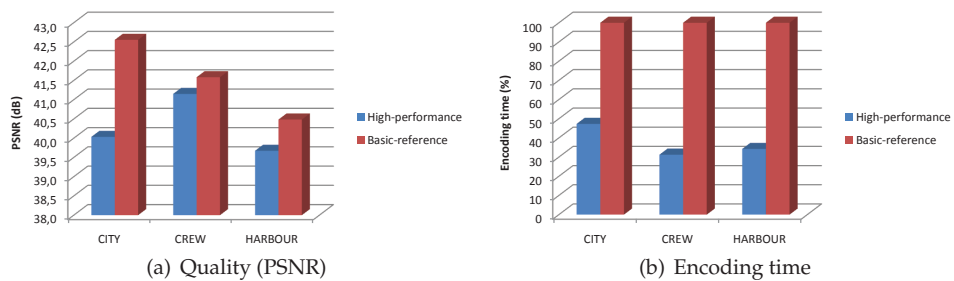


Fig. 21. Comparative between the basic-reference and the high-performance configurations.

5. Conclusion

The goal of this tutorial has been to provide an overview of the advances of the H.264/SVC video standard, focusing on both its features and on an experimental analysis of its configuration parameters. H.264/SVC's superiority over other non-scalable approaches is mainly due to its three different scalabilities (temporal, spatial and SNR), which allow for an improved encoding flexibility and efficiency. By combining different scalabilities into a single bitstream it is possible to achieve, in comparison to previous scalable solutions, similar compression ratios with much lower encoding complexity.

After a brief introduction to this scalable standard, the encoding architecture of H.264/SVC and its most important characteristics have been presented in Section 2. The goal of this section has been to discern the most relevant parameters of the H.264/SVC codification, so as to pave the way for later evaluation of their empirical impact on video quality, coding efficiency and performance while considering, at the same time, its scalability levels.

Next, Section 3 has elaborated on the practical performance of H.264/SVC. Several among the numerous parameters to be configured in this standard are highly influential to the overall coding performance. The imprint of the GOP structure has been proven to be crucial in all the considered metrics, not only because it determines the temporal scalability features of the video stream, but also due to its GOP size, the frame type contained therein and their arrangement. Regarding spatial scalability, H.264/SVC's rescaling algorithms have been examined for both the dyadic and the non-dyadic resolution ratios. Finally, as a result of the experiments done on the quantization parameter and the analysis of the supported SNR scalability modes (i.e. CGS and MGS), interesting concluding remarks have been drawn regarding the H.264/SVC's SNR scalability.

Leveraging the insights of all the performed experiments, Section 4 collects the most important conclusions for practical applications of H.264/SVC video coding. From the experiments contained in this chapter, a tradeoff between video quality and coding complexity has been identified. Therefore, for each scenario, the configuration of the H.264/SVC video coding needs to be adjusted, following the guidelines provided in this last section.

All in all, this chapter intends to be an useful wherewithal to help the reader understanding the H.264/SVC standard, as well as a practical design guide for researchers and practitioners for future scalable video applications.

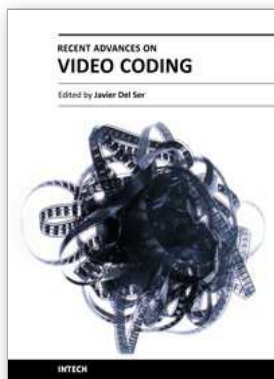
6. Acknowledgements

The authors would like to thank several funding resources. On the one hand, TECNALIA's work was supported in part by the Spanish Ministry of Science and Innovation through the CENIT (ref. CEN20071036) and the Torres-Quevedo (refs. PTQ-09-01-00739, PTQ-09-02-01814 and PTQ-09-01-00740) funding programs, while the work of UFRGS was supported by the FINEP (Projects and Studies Financing) program.

7. References

- Ffmpeg project* (2010). <http://www.ffmpeg.org/>. Version 0.6.1; accessed online on February-09-2010.
- H.263 ITU-T Rec.* (2000). Video coding for low bit rate communication.
- H.264/AVC* (2010). Information technology - Coding of audio-visual objects - Part 10: Advanced video coding, ISO/IEC 14496-10:2010.
- H.264/SVC* (2010). Ammendment G of Information technology - Coding of audio-visual objects - Part 10: Advanced video coding, ISO/IEC 14496-10:2010.
- Huang, H.-S., Peng, W.-H. & Chiang, T. (2007). Advances in the scalable amendment of h.264/avc, *IEEE Communications Magazine* 45(1): 68.
- Husemann, R., Roesler, V. & Susin, A. (2009). Introduction of a zonal search strategy for svc inter-layer prediction module, *VLSI-SOC 2009*, Florianopolis, Brazil.
- JSVM reference software* (2010). http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm. Version 9.19.4; accessed online on February-09-2010.
- JVT reference software* (2010). <http://iphone.hhi.de/suehring/tml/download/>. Version 17.2; accessed online on February-09-2010.
- Marpe, D., Wiegand, T. & Hertz, H. (2006). The h.264/mpeg4 advanced video coding standard and its aplications, *IEEE Communications Magazine* 44(8): 134–143.
- MPEG-2 Video* (2000). Information technology – Generic coding of moving pictures and associated audio information: Video, ISO/IEC 13818-2:2000.
- MPEG-4 Visual* (2004). Information technology - Coding of audio-visual objects - Part 2: Visual, ISO/IEC 14496-2:2004.
- MSU Video Quality Measurement Tool* (2010). http://compression.ru/video/quality_measure/video_measurement_tool_en.html. Accessed online on February-09-2010.
- Ohm, J.-R. (2005). Advances in scalable video coding, *Proceedings of the IEEE* 86(1): 42–56.
- Rieckl, J. (2008). *Scalable video for peer-to-peer streaming*, Master's thesis, University of Wien.
- Schafer, R., Schwarz, H., Marpe, D., Schierl, T. & Wiegand, T. (2005). Mctf and scalability extension of h.264/avc and its application to video transmission, storage and surveillance, *Proceedings of the SPIE*, pp. 343–354.
- Schwarz, H., Marpe, D. & Wiegand, T. (2006). Overview of the scalable h.264/mpeg4-avc extension, *Proceedings of IEEE International Conference on Image Processing*, pp. 161–164.
- Schwarz, H., Marpe, D. & Wiegand, T. (2007). Overview of the scalable video coding extension of the H.264/AVC standard, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1103–1120.

- Segall, A. & Sullivan, G. (2007). Spatial scalability within the h.264/avc scalable video coding extension, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1121–1135.
- Unanue, I., Del Ser, J., Sanchez, P. & Casasempere, J. (2009). H.264/svc rate-resiliency tradeoff in faulty communications through 802.16e railway networks, *Ultra Modern Telecommunications and Workshops*, 2009. ICUMT '09. International Conference on, pp. 1–6.
- Wien, M., Cazoulat, R., Graffunder, A., Hutter, A. & Amon, P. (2007). Real-time system for adaptive video streaming based on svc, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1227–1237.
- Wien, M., Schwarz, H. & Oelbaum, T. (2007). Performance analysis of svc, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1194.
- YUV video repository (2010). <http://www.tnt.uni-hannover.de/>. Accessed online on February-09-2010.



Recent Advances on Video Coding

Edited by Dr. Javier Del Ser Lorente

ISBN 978-953-307-181-7

Hard cover, 398 pages

Publisher InTech

Published online 24, June, 2011

Published in print edition June, 2011

This book is intended to attract the attention of practitioners and researchers from industry and academia interested in challenging paradigms of multimedia video coding, with an emphasis on recent technical developments, cross-disciplinary tools and implementations. Given its instructional purpose, the book also overviews recently published video coding standards such as H.264/AVC and SVC from a simulational standpoint. Novel rate control schemes and cross-disciplinary tools for the optimization of diverse aspects related to video coding are also addressed in detail, along with implementation architectures specially tailored for video processing and encoding. The book concludes by exposing new advances in semantic video coding. In summary: this book serves as a technically sounding start point for early-stage researchers and developers willing to join leading-edge research on video coding, processing and multimedia transmission.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Iraide Unanue, Inigo Urteaga, Ronaldo Husemann, Javier Del Ser, Valter Roesler, Aitor Rodriguez and Pedro Sanchez (2011). A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and Performance, Recent Advances on Video Coding, Dr. Javier Del Ser Lorente (Ed.), ISBN: 978-953-307-181-7, InTech, Available from: <http://www.intechopen.com/books/recent-advances-on-video-coding/a-tutorial-on-h-264-svc-scalable-video-coding-and-its-tradeoff-between-quality-coding-efficiency-and>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821