H.264/SVC Performance and Encoder Bit-stream Analysis

Amina Kessentini, Imen Werda, Amine Samet, Mohamed Ali Ben Ayed, Nouri Masmoudi
Laboratory of Electronics and Information Technologies (LETI)
University of Sfax, National school of Engineering
B.P.W, 3038 Sfax, TUNISIA

ABSTRACT
With the introduction of diverse variety of display transmission and resolutions channel capacities, the Joint Video Team (JVT) has developed the H.264/SVC as an extension of H.264/AVC. In fact, it provides a single compressed bit-stream with several scalability levels. Such a dataflow needs to be analyzed. Consequently, this paper is the first that decorticates and investigates the H264/SVC bit-stream in order to highlight its contribution from one hand and to analyze deeply the different sub bit-stream modules in terms of size and importance on the other hand. Results of a first analysis shows that multicast coding using H264/SVC standard provides an average bit rate reduction of 18% compared to simulcast. Second analysis demonstrates the importance of inter layer prediction. Then a third study illustrates two best combinations for two network bandwidth limitation. Finally, analysis of different subfields that constitute H264/SVC bit stream shows the importance of the residual module which can form up to 72% of the total data output. Results also illustrate the significance of the inter-layer prediction. In fact, base layer information takes the lion’s share of bit consumption mainly for B frame.

General Terms
Signal processing, video compression.

Keywords
Scalable Video Coding, bit-stream, complexity, inter-layer prediction.

1. INTRODUCTION
Currently, the real-time video transmission over the Internet is more and more used in several applications. This heterogeneity causes problems of bit-stream adaptation to ensure a good quality of customer service. One solution to this problem is to deliver the same video with different characteristics such as resolution, video quality and frame rate. Thus, the user can choose the video according to his needs. Despite the efficiency of this solution, it causes numerous problems for instance: storage, video management at server level and redundancy packages from the video. Certainly, this redundancy is due to the fact that two versions of the same video have several packages in common. To overcome these problems, scalable video coding (SVC) standard was adopted. In fact, SVC is the recent development of the Joint Video Team (JVT) of ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Pictures Experts Group (MPEG) [1]. It can be considered as an extension of the H.264/AVC [2, 3, 4]. It is designed to offer temporal, quality (SNR), and spatial scalability. The resulting SVC bit stream is scalable thus; an explicit part can be deserted to get a sub-stream which may be decoded in order to provide an output video sequence with reduced frame-rate or for targeted spatial resolution or quality in terms of PSNR. In this paper, a survey of the complexity of this standard was developed in order to highlight the complexity source of this standard. No H264/SVC bit stream analysis was performed in literature to emphasize the size and the importance of the different modules that constitute the sub-streams that insure the scalability. In fact, using the video scalability, different versions of a single video sequence is stored in one file. Therefore, it can serve different users with a single stream and limits the amount of data flowing over the network. In fact, many features or tools are introduced in order to lead a better coding efficiency which causes a major increase of the video encoder [5, 6]. The remainder of this paper is structured as follows. In section II, an overview of the scalable video coding extension is introduced in order to underline the SVC features. Then a modular bit stream description is detailed in section III. This is followed by SVC performance analysis in section IV. Then, a bit stream analysis is detailed in section V. Finally, conclusion and perspectives are presented in section VI.

2. SCALABLE VIDEO CODING EXTENSION OVERVIEW
Application areas today, range from video telephony and video conferencing over mobile TV to DVD, and HD DVD. For these applications a variety of video transmission and storage systems may be employed. Scalable video coding is considered as a suitable solution to the problems caused by the characteristics of modern video transmission systems. With the introduction of the H.264/AVC video coding standard [2, 3, 4] several amendment in video compression have been established. Then, the scalable video coding SVC standard is recently developed by the JVT [1, 7, 8, 9]. The term “scalability” passes on to the removal of parts of the video bit stream in rank to adjust it to the various needs of end users. Thus, a video bit stream is called scalable while parts of the stream can be unconcerned in a way that the resulting sub-stream forms an additional valid bit stream for some target decoder. The H.264/SVC encoder is shown in Figure 1. We may observe that video source is undergone using a down sampling, thus the base layer encoder, Layer 0, takes a lower resolution video sequence as input and encodes it with ordinary H.264/AVC standard. For the enhancement layer, Layer1, the input is in a higher resolution and be coded as an ordinary H.264/AVC moreover interlayer predictions provide additional coding choices such as motion vectors, intra
prediction, and residual information from base layer. The SVC extension of H.264/AVC presents three types of scalability [1,7,8,9]: Temporal, Quality, and Spatial scalability, as described in the following section.

2.1 Temporal Scalability
Temporal scalability is offered using the hierarchical B pictures. The most popular one is the dyadic prediction structure as illustrated in Figure 2.

Fig 2. Dyadic Hierarchical Prediction Structure

The Hierarchical Group of Picture (GOP) starts with an Instantaneous Decoding Refresh (IDR) picture and ends with either intra (I) or Predicted (P) picture. In general, the GOP is composed of 8 pictures. Numbers directly underneath the picture identify the coding order and the symbols Tx identify the temporal layer while x presents the corresponding temporal layer identifier. The enhancement layer pictures are typically coded as B pictures. The B1 picture of level 1 with T1 as temporal layer identifier uses only the surrounding I or P references for prediction. The Bi pictures of level i > 1 with temporal layer identifier Ti may use the surrounding or the key pictures I or P in addition to the Bj pictures such that level j is less than i with Tj for prediction. In addition to the dyadic case presented in Figure 2, there are many other methods to the hierarchical prediction structure such as non-dyadic hierarchical prediction structure, depicted in Figure 3 (a). As a result of this structure we may have two independently subsequence with 1/9 th and 1/3 rd of the full frame rate. Furthermore, a second structure is illustrated in Figure 3 (b), where we may witness no motion compensated prediction for the future picture [1, 10].

2.2 Spatial Scalability
SVC supports multilayer coding in order to supply resolution variety illustrated in Figure 4.

Fig 3. Two Non Dyadic Hierarchical Prediction Structure

Each layer in SVC encoding corresponds to a spatial resolution. To be able to identify them, a dependency identifier is submitted to each layer. The base layer gets 0 as the dependency identifier and it is enhanced by 1 from one spatial layer to the next one. Dn corresponds to the maximum number of spatial layer which can be no more than 7 layers. For each spatial layer and in order to reduce redundancy between spatial layers, inter layer prediction is applied. In fact, it exploits spatial layer’s information to predict the next layer’s inter residual data, motion vector (MV) and intra texture [9, 11].

2.2.1 Inter-layer intra prediction
As an amendment of the SVC extension, the inter-layer intra prediction (ILIP) is defined in order to predict the enhancement layer macroblock (MB) from the previous layer intra MB. The ILIP may be used when the enhancement layer MB is coded with base mode flag equal to 1 and the co-located 8x8 sub-macroblock in its reference layer is intra coded [1]. So, the corresponding reconstructed intra signal of the reference layer is up-scaled using 4-tap FIR filters horizontally and vertically. We notes that the adjacent block of previous layer are not intra coded and a single-loop decoding is provided. Consequently, the inter-coded reconstruction in the reference layer is avoided [12].

2.2.2 Inter-layer Residual prediction
In order to operate the residual information coded in a lower resolution, an inter-layer prediction is employed. Therefore, whatever the type of a MB, an additional flag is added to each
MB syntax to notify the usage of inter-layer residual prediction (ILRP). If this flag is true the residual signal of a reference layer is up sampled with the aim of being used as a prediction for the residual signal of the enhancement layer MB [1, 11]. Consequently, the difference between the residual data of enhancement layer and the reconstructed residual of previous layer is coded.

2.2.3 Inter-layer motion prediction
As an amendment of the SVC, a new MB type is introduced signaled by a syntax element called base mode flag. When the corresponding 8x8 block in the reference layer lies inside an inter-coded MB and the base mode flag is equal to 1, so the MB is then predicted using the inter-intra prediction. Otherwise, when the reference MB is inter-coded, the enhancement MB is inter coded. Thus, the motion vectors, the partitioning data and the reference indexes are derived from lower resolution using inter-layer motion prediction (ILMP).

The MB partitioning is obtained by up-sampling the corresponding partitioning of the co-located 8x8 block in the reference layer. The SVC allows the usage of scaled motion vectors of the co-located 8x8 block in the reference layer. In order to indicate whether inter-layer motion vector predictor is used, a flag so called motion prediction flag for each used reference picture list is transmitted. When motion prediction flag is equal to 1, the reference indexes for the MB partition are coded using the reference layer. So, the co-located reference indexes for MB partition are used and the motion vector of the co-located blocks is scaled form the motion vector of the enhancement layer MB. Otherwise, when motion prediction flag is equal to 0, the reference indexes for the corresponding reference picture list are coded in the enhancement layer and the motion vector prediction is employed as specified in H.264/AVC [1,11,12].

2.3 Quality Scalability
The H.264/SVC supports three quality scalability modes such as Coarse-Grain Scalability (CGS), Fine-Grain Scalability (FGS) and Medium-Grain Scalability [13].

2.3.1 Coarse-Grain Scalability (CGS)
This type of scalability may be considered as a special case of spatial scalability where the enhancement and the base layer will have the same picture size. So, the same inter-layer prediction method for spatial scalable used, but without up sampling operations and the inter-layer residual and intra prediction are achieved in the transform domain. For CGS mode, the residual texture signal is re-quantized using a quantization step smaller than the quantization step of the preceding layer [14].

2.3.2 Medium-Grain Scalability (MGS)
This mode uses similar techniques of CGS. However, MGS try to solve a several problems encountered for CGS such as the lack of flexibility for bit stream adaptation and the limited number of rate points. So, the flexibility is improved by allowing the removal of these quality levels at any point in the bit stream. MGS can be used to allow up to 128 quality extraction points [13].

2.3.3 Fine-Grain Scalability (FGS)
With the aim of supporting the FGS, a Progressive Refinement (PR) slices have been inserted. Each PR slice represents a refinement of the residual signal that corresponds to a bisection of the quantization step size (QP increase of 6) [14]. These signals are represented in a way that only a single inverse transform has to be performed for each transform block at the decoder side. The ordering of transform coefficient levels in PR slices allows the corresponding PR Network abstraction layer (NAL) units to be truncated at any arbitrary byte-aligned point, so that the quality of the SNR base layer can be refined in a finer-granular way [14]. The FGS in MPEG-4 present a low performance because the Motion Compensated Prediction (MCP) uses always the SNR base layer. For the SVC design, as detailed in Figure 5, the MCP uses the highest quality reference of temporal refinement pictures.

This difference improves the coding efficiency without increasing the complexity when hierarchical prediction structures are used [14].

3. BIT-STREAM DESCRIPTION
The SVC is an extension of the H.264/AVC. In fact it uses the H264/AVC basics. While describing the bit-stream we will depict the H.264/AVC features for understanding the concepts of extending H.264/AVC towards SVC. The H264/AVC design covers two additional layers compared with previous standards, a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL) as mentioned above [1,15,16]. The VCL guarantee a good efficiency in compression through coding the source content. The NAL arranges these data and offers header information in a way that enables simple and effective customization of the use of VCL data for broad variety of systems.

3.1 Network Abstraction Layer (NAL)
The coded video data are composed by NAL units, as depicted in Figure 6. Each one contains an integer numbers of bytes.

The video data format

The first byte of each NAL unit is the header, which indicates the NAL unit type. Then, the next 3 bytes, as amendment of the SVC, presents the header extension. The remaining bytes represent the payload data [6].

3.1.1 NAL unit Header
The syntax of NAL unit header with SVC extension is depicted in Figure 7 [17].
changing for a video sequence. The SEI messages offer additional information for the decoding process. In fact, they are not necessary for decoding video sequence’s samples. According NAL unit header, SVC or AVC NAL unit is coded. A succession of AVC and SVC NAL unit form the bit-stream which corresponds to base layer coded picture following to enhancement layer coded picture. We note that additional flags are incorporated in RBSP to specify the type of NAL unit that means SVC NAL unit or AVC one. In this paper we focus on additional flag for SVC NAL unit and we note the most significant one:

- **Base_mode_flag** specifies that the macroblock prediction mode and the corresponding motion data are inferred from base layer.
- **Motion_prediction_flag** indicates that inter-layer motion prediction is used or not.
- **T_coeff_level_prediction_flag** specifies that inter-layer intra prediction is used or not.
- **Residual_prediction_flag** specifies that the residual signal of the current macroblock is predicted using base layer or not.

### 3.2 Video Coding Layer (VCL)

The VCL details a well-organized representation of the coded video data [15]. The H.264/AVC VCL design is very similar to that of proceeding video coding standard such as H.261 and H.263. However, H264/AVC contains new features that enable it to achieve an important improvement in compression efficiency relative to any prior video coding [1]. The major difference of H.264/AVC to previous standard is the main flexibility and adaptability as detailed in [3]. According to the video coding strategy, the picture is partitioned into macroblocks and slices. For the H.264/AVC, macroblock size is smaller than that in previous standard. In fact, it can achieve 4x4 block size. In order to remove the redundancy of data samples, two types of prediction are used such as temporal and spatial. The resulting prediction signal is represented using transform coding. The macroblocks of a picture are then arranged into slices which can be parsed independently of other slices. The H.264/AVC supports three slice coding types: I-slice, P-slice, and B-slice. The major feature of the H.264/AVC is the entropy coding using two methods context-based adaptive variable length coding (CAVLC) and context-based adaptive binary arithmetic coding (CABAC).

### 4. SVC PERFORMANCE EVALUATION

The SVC performance evaluation is divided into two parts. The first one consists of presenting some coding efficiency experiments and the second one presents a comparison of several bit rate combination. Our experiments were made using the Joint Scalable Video Model (JSVM 9.17) software [18].

#### 4.1 Experimental conditions

The SVC. Simulcast and single layer solutions were generated with the JSVM software 9.17. Results are presented according to the SVC test conditions such as Closed loop encoding. Single Loop decoding and Application of hierarchical B-frames which are fixed in [19]. In our Experiments we use video sequences that are common in the video coding community: Parkjoy, Crowdrun and Sunflower sequences. We use the high-definition (HD) video content with two spatial resolutions: 720p and 1080p. We note that the second resolution can be derived from cropped the first one using the down-sampling operation.
implemented in JVT software. In order to achieve the target bit rate no rate control algorithm was used; the quantization parameter was varied to adjust the output Peak Signal-to-Noise Ratio (PSNR). Furthermore, many conditions are fixed such as a GOP size will be equal to 8 (this size was fixed refer to [1]) and an intra period fixed to 16 pictures. Simulation conditions are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Test Conditions For Encoding Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frames Per Second</td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>Intra Period</td>
</tr>
<tr>
<td>Search Range</td>
</tr>
<tr>
<td>Search type</td>
</tr>
<tr>
<td>Number of frames</td>
</tr>
</tbody>
</table>

We note that image sequences are encoded with a hierarchical B-frame structure with an intra frame present every 16 frames. Moreover, two spatial layers are enabled with QP configurations as (QP0, QP1) = (40, 44), (35, 39), (32, 36), and (28, 32). We note that ΔQP = - 4 as recommended in [20].

4.2 Coding efficiency

Figure 9 shows the rate distortion for Parkjoy and Crowdrun sequences. These figures illustrate the rate distortion for three scenarios: SVC, single layer and simulcast. The simulcast and single layer results were obtained using the same implementation and encoding algorithms while reconfigured to encode a single layer. This figure offers approaching into the SVC scenario’s performance. In order to highlight this performance a variation in bit rate and peak signal-to-noise ratio (PSNR) measurements for all sequences test is presented in Table 2. When comparing the SVC extension to single layer and simulcast scenarios, we note that even if (quality and bit rate) performance is sequence dependent, it provides an average bit rate reduction of 18.02% compared to simulcast.

Furthermore, the SVC offers an average of 8.33% increase in bit rate compared to single layer. As noted above, when broadcasting a SVC bit stream, a lower resolution one may be simply extracted and decoded. One of the major features is the inter-layer prediction and we note that when analyzing Figure 9 and Table 2 a general inter-layer prediction performance is illustrated. In order to highlight the inter-layer prediction performance, we eliminate the inter-layer process and we focus on the bit-stream and PSNR variation.

Table 3 illustrates the impact of the whole inter-layer prediction on both video-quality and coding efficiency performances for three video sequences at HD resolution. In order to evaluate the performance of the three inter-layer prediction modes, a second analysis is conducted. In fact, we disable one of the inter-layer predictions modes and we keep the other two unchanged while preserving the same PSNR output.

Table 4 illustrates results in terms of bit rate. According to this analysis, when disabling one type of the inter-layer prediction we note an increase in terms of bit rate. In fact, for a targeted PSNR output, the encoder has to consume more bits in order to compensate the role of the disabled module. This fact is more obvious especially for a large QP value. This observation is clear for both modes ILIP and ILRP. However, the impact of the ILMP is not as significant as the others modes. Our observation
is also valid for CIF and QCIF sequences as demonstrated in [21, 22].

<table>
<thead>
<tr>
<th>QP (28,32)</th>
<th>NO ILIP</th>
<th>NO ILRP</th>
<th>NO ILMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>QP (35,39)</td>
<td>NO ILIP</td>
<td>NO ILRP</td>
<td>NO ILMP</td>
</tr>
</tbody>
</table>

### 4.3 Bit rate combination

In a second part of our SVC performance evaluation, we search to extract the best combination to feat better the network bandwidth limitation and to improve the trade-off base layer-enhancement layer. Thus, two combinations are studied such as 12 Mb/s and 8 Mb/s. We varied the quantization parameter in order to obtain several base and enhancement layer combinations of two bit rate threshold respectively 12 Mb/s and 8 Mb/s. Figure 10 illustrates the 12 Mb/s combination for two test sequences Parkjoy and Crowdrun.

![Fig 10. 12 Mb/s combination for two sequences](image)

When comparing the different 12 Mb/s combination, to achieve a good 1080p video quality without losing the 720p resolution, the combination (8, 4) is the appropriate combination for these two test sequences. Figure 11 shows the 8 Mb/s combination for two sequences Parkjoy and Crowdrun. The most sophisticated combination is (4, 4). This result is used for the rest of the bit-stream analysis.

![Fig 11. 8 Mb/s combination for two sequences.](image)

### 5. SVC Video Stream Analysis

As a continuation of the results obtained in the last section, we decide to analyze the coded bit-stream that SVC provides, in order to extract the most critical part in terms of bits consumption. So in the rest of this section a statistical analysis of SVC bit-stream is performed. This section is organized in two subsections: the first one presents the base layer analysis and the second one reserved for the enhancement one. In this section results for Parkjoy sequence are illustrated. We note that similar results are obtained for the other sequence. Our analysis is developed for 97 frames distributed as illustrated in Table 5.

<table>
<thead>
<tr>
<th>Frame type</th>
<th>I</th>
<th>P</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame number</td>
<td>7</td>
<td>6</td>
<td>12</td>
<td>24</td>
<td>48</td>
</tr>
</tbody>
</table>

#### 5.1 Base layer analysis

Among this subsection we focus on base layer partition. Table 6 illustrates the base layer bit-stream partition. This table presents the average in terms of bits of I frame, P frame and B frame with three temporal levels.

<table>
<thead>
<tr>
<th>Frame type</th>
<th>I</th>
<th>P</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of bits consumption (%)</td>
<td>55.59</td>
<td>19.1</td>
<td>16.8</td>
<td>6.58</td>
<td>1.93</td>
</tr>
</tbody>
</table>

When analyzing this percentage, we find that I frame is the greediest picture in terms of bits. In fact, a further analysis is then performed to peel the bit-stream of base layer. Thus, we focus on each picture partition in percentages of prediction information such as inter, intra, skip information and the
residual one. Table 7 illustrates bit-stream details for different frame type. According to this analysis, the residual information takes the lion’s share of the bits consumption. We note that coding the residual information for I frame can reach 76% of the total I frame bits. For all frame’s type such as P and B frame with different temporal level the percentage of residual information is more than 50%. In fact, the encoder consume more bits for coding residual information when comparing with others type information such as inter information which can yield for large case 27.3%.

Table 7. Percentage details partition for each frame type

<table>
<thead>
<tr>
<th></th>
<th>Intra</th>
<th>Inter</th>
<th>Skip</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>P</td>
<td>11.44</td>
<td>24.93</td>
<td>0.17</td>
<td>63.46</td>
</tr>
<tr>
<td>B1</td>
<td>1.03</td>
<td>18.35</td>
<td>8.58</td>
<td>72.04</td>
</tr>
<tr>
<td>B2</td>
<td>0.28</td>
<td>19.02</td>
<td>13.77</td>
<td>66.93</td>
</tr>
<tr>
<td>B3</td>
<td>0.12</td>
<td>27.3</td>
<td>13.32</td>
<td>59.26</td>
</tr>
</tbody>
</table>

5.2 Enhancement layer analysis

We focus among this sub-section on enhancement layer. Table 8 shows the bit stream enhancement partition. It presents the average in terms of bits of I frame, P frame and B frame with three temporal levels.

Table 8. Enhancement layer partition

<table>
<thead>
<tr>
<th>Frame type</th>
<th>I</th>
<th>P</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of bits consumption (%)</td>
<td>48.18</td>
<td>24.92</td>
<td>18.06</td>
<td>6.78</td>
<td>2.06</td>
</tr>
</tbody>
</table>

When analyzing this table, we find that I frame is the greediest picture in terms of bits. To highlight the SVC contribution, our idea is to dissect the enhancement layer’s bit-stream in two levels. We depict on the first hand the enhancement layer partition for each frame in terms of information type: intra, inter and residual information. On the other hand we detail the analysis in terms of inter layer information. These analyses have been made on three test sequences: Parkjoy, Crowdrun and Sunflower with two spatial layers and three temporal layers. Table 9 illustrates the enhancement frame partition respectively I frame, P frame and B frame with 3 temporal levels. When analyzing this table, it is obvious that whatever the frame type, the residual information takes the major part (in term of bits) of the bit-stream. We focus now on inter-layer prediction impact. Thus, table 10 presents percentage enhancement layer partition for each frame in terms of information type: intra, intraBL, inter, interBL, residual, residualBL, and skip information. Where intra, inter and residual information are enhancement layer information, as well as intraBL, interBL, and residualBL information are inherited from base layer. Where analyzing this table, we note that almost intra information is inherited from the base layer, even the residual information which can reach more than 60% of bits from the base layer using the residual inter-layer for some cases such as I frame and B2 frame. For P frame, it is obvious that residual information is shared between enhancement and base layer. We note that the skip information present a weak part comparing with other type which depict a good inter prediction. As presented in Table 10 a major part of bit-stream is reserved for encoding residual information mainly residual base layer information. According this analysis, the huge residual inter layer importance is performed.

Table 9. Percentage partition for each frame type

<table>
<thead>
<tr>
<th></th>
<th>Intra</th>
<th>Inter</th>
<th>Skip</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>7.61</td>
<td>0</td>
<td>92.39</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>13.2</td>
<td>35.07</td>
<td>50.21</td>
<td>1.52</td>
</tr>
<tr>
<td>B1</td>
<td>0.96</td>
<td>17.3</td>
<td>79.18</td>
<td>2.56</td>
</tr>
<tr>
<td>B2</td>
<td>0.33</td>
<td>16.67</td>
<td>77.35</td>
<td>5.65</td>
</tr>
<tr>
<td>B3</td>
<td>0.11</td>
<td>22.63</td>
<td>64.95</td>
<td>12.31</td>
</tr>
</tbody>
</table>

We focus now on inter-layer prediction impact. Thus, Table 10 presents percentage enhancement layer partition for each frame in terms of information type: intra, intraBL, inter, interBL, residual, residualBL, and skip information. Where intra, inter and residual information are enhancement layer information, as well as intraBL, interBL, and residualBL information are inherited from base layer. When analyzing this table, we note that almost intra information is inherited from the base layer, even the residual information which can reach more than 60% of bits from the base layer using the residual inter-layer for some cases such as I frame and B2 frame. For P frame, it is obvious that residual information is shared between enhancement and base layer. We note that the skip information present a weak part comparing with other type which depict a good inter prediction. As presented in Table 10 a major part of bit-stream is reserved for encoding residual information mainly residual base layer information. According this analysis, the huge residual inter layer importance is performed.

6. CONCLUSION

In this paper we propose to dissect the SVC standard and to highlight its contribution. First of all, SVC performance evaluation is performed: this analysis shows that SVC provides an average bit rate reduction of 18.02% compared to simulcast and offers an average of 8.33% increase in bit rate compared to single layer while simultaneously supporting two resolutions. Secondly, detailed analysis shows the importance of the inter-layer intra and residual prediction compared to inter layer motion prediction. In fact, when disabling inter-layer-intra prediction, bit rate will increase by up to 40% for some sequences. Then, we search a best combination between the lower and higher resolution to reach a good coding efficiency in order to feaut better the network bandwidth limitation and to improve the trade-off base layer- enhancement layer. This third analysis illustrates that the best combination for 12 Mo bandwidth is (8, 4) in fact we specify 8Mo for coding the base layer and the rest 4 Mo are fixed for the enhancement layer. As well as the best combination for 8Mo bandwidth is (4, 4). Finally, the last study is performed on SVC bit-stream with the best 12 Mb/s combination. This analysis shows that the largest percentages of bits are observed for the residual information. Furthermore, when observing frame details we note that the additional residual prediction from the base layer percentage can yield even to 65% of bits especially for B frames and mainly for the temporal layer three.
7. REFERENCES


Table 10. Percentage details partition for each frame type

<table>
<thead>
<tr>
<th>Intra information</th>
<th>Inter information</th>
<th>Residual information</th>
<th>Skip information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra</td>
<td>IntraBL</td>
<td>Inter</td>
<td>InterBL</td>
</tr>
<tr>
<td>I</td>
<td>0.3</td>
<td>7.94</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>0.46</td>
<td>12.74</td>
<td>14.02</td>
</tr>
<tr>
<td>B1</td>
<td>0.16</td>
<td>0.8</td>
<td>8.85</td>
</tr>
<tr>
<td>B2</td>
<td>0.07</td>
<td>0.26</td>
<td>11.04</td>
</tr>
<tr>
<td>B3</td>
<td>0.07</td>
<td>0.04</td>
<td>15.51</td>
</tr>
</tbody>
</table>