

# Depth Partitioning and Coding Mode Selection Statistical Analysis for SHVC

Ibtissem Wali

National Engineering School of Sfax, Tunisia

Amina Kessentini

High School of Computer Science and Multimedia of  
Gabes, Tunisia

Mohamed Ali Ben Ayed

National School of Electronics and Telecommunication of  
Sfax, Tunisia

Nouri Masmoudi

National Engineering School of Sfax, Tunisia

**Abstract**—The Scalable High Efficiency Video Coding (SHVC) has been proposed to improve the coding efficiency. However, this additional extension generally results an important coding complexity. Several studies were performed to overcome the complexity through algorithmic optimizations that led to an encoding time reduction. In fact, mode decision analysis is imperatively important in order to have an idea about the partitioning modes based on two parameters, such as prediction unit size and frame type. This paper presents statistical observations at two levels: coding units (CUs) and prediction units (PUs) selected by the encoder. Analysis was performed for several test sequences with different motion and texture characteristics. The experimental results show that the percentage of choosing coding or prediction unit size and type depends on sequence parameters, frame type, and temporal level.

**Keywords**—Video Coding; HEVC; SHVC; Coding efficiency; Statistical analysis

## I. INTRODUCTION

Nowadays, the demand on digital signal processing applications is more and more increasing. Consequently, digital video applications cover today a very large range of multimedia application: messaging, HD video telephony and video conferencing.etc. In order to overcome this heterogeneity, numerous versions of the same video are stored in the server side to gratify various needs of clients and are delivered using simulcast coding. This leads to increasing the video bit rates and hence maximizing the storage costs. Therefore, the clients' heterogeneity needs motivated some years ago the development of scalable video coding. "Scalability" consists in removing parts of the video bit stream in rank in order to adjust it to various consumers' needs. In fact, using video scalability is equivalent to obtaining different versions of single video sequence and then storing it in one file. Therefore, scalability can serve different users with a single stream and limits the amount of data flowing over the network. Consequently, a new video coding standard was developed with improved compressing tools. SHVC was proposed as a scalable extension of the High Efficiency Video Coding (HEVC) standard [1]. Besides, several solutions were suggested to respond to the proposed scalable extension

[2,3,4]. The scalable extension was developed [5] by the Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11. In fact, HEVC enables a better compression compared to H.264/AVC with maintaining the same video quality [6]. SHVC was proposed based on this standard with several amendments in video compression. Thus, SHVC introduced new inter-layer prediction feature in order to improve the coding efficiency [3]. However, the introduction of this new feature increases the encoder complexity. Therefore, a good understanding of the fundamental aspects of the SHVC is recommended as a first step to know how to intervene in its weak points to ensure its usability. Since, mode decision is a very consuming part [25], an analysis of the encoder decision module is necessary to have an idea about the partitioning modes in terms of size and type.

In this paper, we propose to dissect the SHVC standard and to highlight its contribution using performance analysis. Moreover, we focus on encoder mode decision, and we develop a statistical observation at two levels: CU and PUs. This proposal is performed based on frame type following temporal level.

The remainder of this paper is organized as follows: In section 2, an overview of the scalable extension of HEVC is presented by detailing its major aspects. Section 3 discusses the related works on HEVC. Section 4 illustrates the average time distribution and the video coding performance. Section 5 describes a statistical study and result discussion. Finally, in section 6, we present some concluding remarks and ideas for future work.

## II. OVERVIEW OF THE SCALABLE EXTENSIONS OF HEVC

The introduction of the HEVC and the development of its scalable extension have enhanced considerably video compression [1, 5]. Indeed, SHVC provides several scalability features such as spatial, SNR, bit depth and color gamut scalability [14]. SHVC encoder supports multi-layers coding that consists of HEVC encoders to encode the Enhancement Layers (EL) and HEVC or AVC encoder to encode the Base Layer (BL).

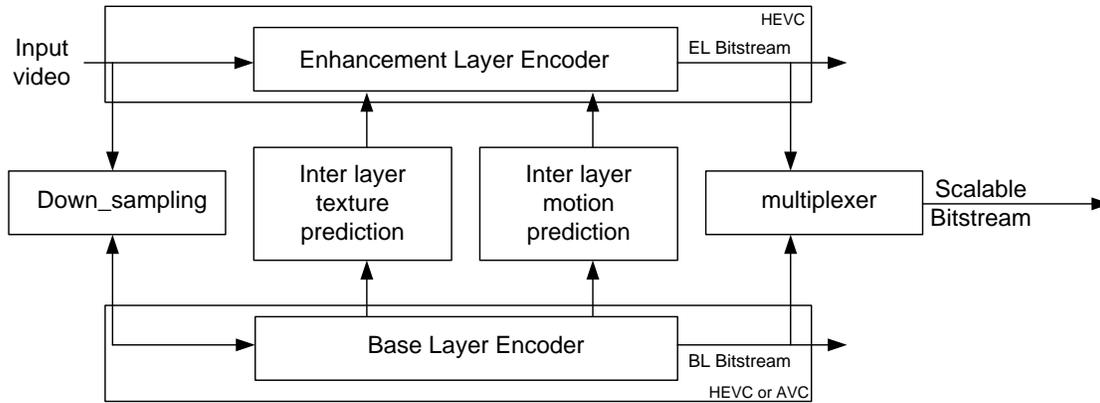


Fig. 1. SHVC structure with two layers

An example of SHVC structure with two spatial layers is shown in Fig 1. The SHVC architecture could support eight layers; one base layer and seven enhancement layers. In Fig 1, we only describe two of these layers; a BL encodes a down sampled version of the original input video, and EL encodes the original one. For coding the EL, the SHVC encoder reuses the coding tools already present in HEVC in addition to the inter-layer prediction added features in order to offer gain in coding efficiency. Consequently, the upsampling of the encoded pictures and the upscaling of their relative motion vectors and texture prediction are used in EL's coding.

The SHVC offers several scalability features enumerated below:

- Temporal scalability: is guaranteed by using the hierarchical B pictures [7].
- Spatial scalability: is ensured by using multi-layer coding in order to supply resolution variety. Besides, SHVC supports resolution up to 4K [8] and 8K [9].
- Quality scalability: This type of scalability is considered as a special case of spatial scalability where the EL and the BL have the same picture size. In SHVC, the Coarse-Grain SNR scalability (CGS) is used. In fact, this type of SNR scalability is the most usual SNR scalability mechanism supported by many standards [10].
- Hybrid codec scalability: layers could be coded whether using the HEVC codec or Non-HEVC codec (e.g.H.264/AVC codec) [11].
- Bit depth scalability: this feature is inherited from HEVC. In fact, this mechanism is ensured where the BL is of lower bit depth (e.g. 8 bit) and the EL is of higher bit depth (e.g. 10 bit) [3].
- Color gamut scalability: The color gamut scalability is applied in the case of a different color space between BL and EL.
- We note that these scalability features may be combined. For example, the combination of bit depth, color gamut, and spatial scalability may be used to fully

enable the migration of the video sequence from HDTV to UHD TV [3].

As an extension, SHVC adds new tools to those provided by HEVC. They are called interlayer predictions tools. This feature is used, if necessary, in order to improve the coding efficiency. These tools are manifested on interlayer texture and interlayer motion prediction.

#### A. Inter-layer texture prediction

Fig.2. illustrates the modality of the new interlayer texture prediction for spatial scalability. In SHVC, the reference layer pictures do not correspond to the pictures in the base layer, but they correspond to the picture known as inter-layer reference (ILR) picture. The ILR picture is generated by the re-sampling process when the BL size differs from the EL size. In order to obtain an ILR picture, the BL is first up-sampled to obtain the BL picture. Then, it is cropped to meet the EL size (when the cropping window process is enabled). Afterward, the ILR picture is inserted into the reference picture lists as inter-layer texture prediction to encode the EL [12].

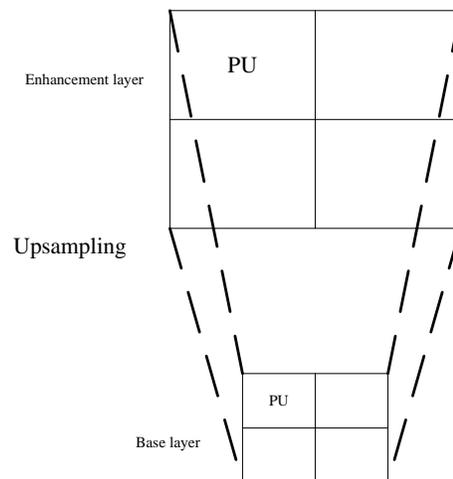


Fig. 2. Re-sampling process of the prediction units

#### B. Inter-layer motion prediction

HEVC includes two motion vector methods: the advanced motion vector prediction (AMVP) and the merge mode. The

motion vectors predictors list (MVPs) is generated by the AMVP from spatial or temporal neighboring. When spatial neighboring is not available, we consider the temporal motion vector prediction (TMVP). In SHVC, the motion field mapping method is introduced to derive the motion field for ILR pictures as depicted in Fig.3.

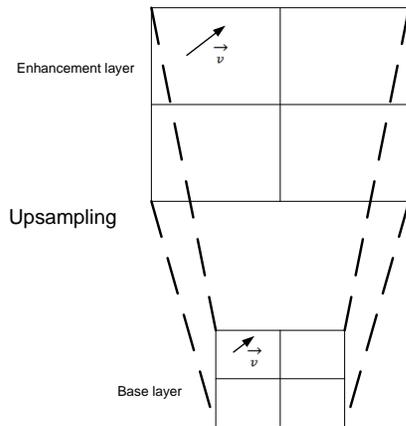


Fig. 3. Motion vector generation

### III. RELATED WORKS

SHVC outperforms previous standards in terms of performance. However the cost of computational complexity appears to be a critical research field. Consequently, several works have deeply explained the new features introduced in SHVC [13,14], such as the new scalable feature CGS and the bit depth. These studies also explained the interlayer procedure which includes the resampling processing, the motion field and the color mapping. To evaluate the performance of SHVC, they used two anchors; the simulcast and the single layer EL. An evaluation of the performance of two coding configurations (random access RA and low delay LD with Bpictures (LDB)) was performed in [13]. If compared to HEVC simulcast for RA configuration, SHVC achieves an average bit rate reductions of 16.5%, 27.0% and 20.9% for 2x, 1.5x, and SNR scalability cases, respectively. For the LDB configuration, SHVC achieves an average bit rate reductions of 10.3%, 21.5%, and 12.5% for 2x, 1.5x, and SNR scalability, respectively. In [14], SHVC attained average bit rate reductions equal to 50.3%, 56.7%, and 51.5%, in the RA configuration, and 63.1%, 62.5%, and 54.1% in the LD with P picture (LDP) configuration, for 2x, 1.5x, and SNR scalability, respectively. The SHVC was compared to simulcast for coarse grain scalability with and without the 3D LUT color mapping tool. In the first case, comparison achieved average bit rate reductions of 18.1% and 29.6%. In the second case, the average bit rate reductions decreased significantly to only 10.2% and 17.2%. In [15], a real time analysis was introduced using a hybrid parallelism. These latter takes advantages from both; the wave front parallelism (WPP) and frame based and provide time repartition of the optimized SHVC decoder for different scalability configurations (using OpenHEVC). The interlayer prediction enabled a gain in bit rate of about 40% for x1.5 resolution in spatial scalability. A hybrid parallelism provided the highest speed up with a good trade-off among decoding time, latency and memory usage [16]. A statistical analysis was

made for H264/SVC to suggest an improved mode decision for EL. This mode decision depends on the frame type and uses the best mode found in BL. For each type of frame I, P, and B (depending on the temporal level), authors proposed an algorithm that decreases the encoding time reaching 64% [17]. In [18], a HEVC coding quad-tree was early terminated by using residuals statistics at the PUs level. The prediction residuals statistics was computed as the absolute difference between the prediction residuals variances for the two  $N \times 2N$  and for the two  $2N \times N$ . The introduced residual based method allowed reducing the encoding time by an average of about 44% [18]. A statistical analysis of coding units was chosen by the encoder for HEVC in [19]. In this paper, authors showed that a detailed explanation for the use of largest coding units (LCUs) or smallest coding units (SCUs) depends on the image textures. Following the image aspects, they fixed a texture parameter to classify images into textured images and non-textured ones. Since the inter-layer prediction module in SHVC presents the lion's part of the consuming time, several efforts have been made to optimize it. Hence, in [20], three novel methods were developed: copying directly the BL split flags to the EL, disallowing intra-prediction modes for EL, and disallowing modes that split the CUs orthogonally. These approaches were combined to achieve a 15.3% average reduction for the whole encoding time with a bit rate increment of 0.86% and a PSNR decrement equal to 0.12 dB. In [21], a new prediction mode that decreases the bit-rates of the enhancement layer by up to 3.13% of SHVC was introduced. It is based on the fact that in video coding, sometimes the same patterns are repeated in the residual signals. The current SHVC encoder and decoder were modified by implementing a new prediction mode for coding residual information called *Residual* mode. A down sampled residual signal was obtained by using a two dimensional bilinear down sampling operator applied on residual signal. Authors computed afterward the number of required bits to represent this new residual signal which is then up-sampled to have the same size of the predicted signal. Finally, authors calculated the RD-cost and compared it with the original SHVC [21]. In [22], the proposed method is based on the operations determining the CTU structure which is the most time consuming since the encoder has to check all possible CU sizes. In SHVC, the CTU to be encoded depends on: four neighbors, the current CTU in the BL and its corresponding CTU in the previous frame. Consequently, authors used a learning machine approach based on training and testing procedures to translate the tree structure into numbers. Since the CTU bloc can be split into 4 sub blocs of size  $32 \times 32$  and each sub bloc can be divided into 17 possible options, the number of the possible CTU portions has been reduced from  $17^4 + 1 = 83522$  to 18 using a probabilistic approach. The encoding time decreased of about 56.79% and 63.18% for the scalable ratios 1.5 and 2, respectively, when combining all the introduced methods.

When analyzing the previously mentioned works, we note that several studies were carried out to evaluate and highlight the impact of SHVC and compare it to previous standards. We note also that some efforts have been made to propose several algorithms in order to speed up the SHVC encoder. Some algorithms were introduced based on the probabilistic approach and the correlation between BL and EL resolution. These

algorithms can be applied on all types of images. In other words, all algorithms for SHVC standard do not differentiate image type. Indeed, they use the same mode decision. Furthermore, when analyzing previous works for SHVC standard, we notice that there are no statistical observations for different types of image according to the temporal level.

Hence, we propose, in this paper, a brief overview of SHVC, followed by our own evaluation and eventually a detailed statistical analysis. In fact, a statistical observation based on frame type, CU, and PU size is performed.

#### IV. EVALUATION AND RESULTS

The scalable extension of the reference software model SHM v.7.0 [23] was used to generate video bit streams. We considered the common test conditions defined in the SHVC standard [24]. The video sequences, in Table I, were coded in two scalability layers with Random access and low delay P configurations including SNR scalability and spatial scalability for two configurations x1.5 and x2. The obtained quantization parameters values are shown in Table II. Indeed, we have 8 combinations for each configuration for one video sequence. All the simulations were carried on a Windows 7 OS platform with Intel\_core TM i7-4790 @ 3, 6 GHz CPU and 18 Go RAM. In this section, we study the average time distributions and coding efficiency for several test sequences.

TABLE I. VIDEO SEQUENCES CONSIDERED IN THE EXPERIMENTS

Classe	Sequences	Resolution	Frame rate
A	Traffic	1280x800	30
	PeopleOnStreet	2560x1600	30
	Kimono		24
B	ParkScene	960x540	24
	Cactus	1280x720	50
	BasketBallDrive	1920x1080	50
	ve		50
	BQTerrace		60

TABLE II. QUANTIZATION PARAMETER VALUES

Scalability ratio	BL QP	EL delta QP
<b>Spatial 1.5x, 2x</b>	22, 26, 30, 34	0, 2
<b>SNR</b>	26, 30, 34, 38	-6, -4

##### A. Average time distribution

In order to determine most time consuming modules, a profiling for encoder and decoder has been performed [25]. Fig.4 provides the time repartition of the optimized SHVC decoder in the case of spatial x1.5, x2, and SNR scalabilities for both encoder and decoder for Basketballdrive test sequence. For the case of encoder, the inter-prediction and the Rate distortion Cost calculation represent the highest percentages of the encoding time. For the decoder case, the inter-layer prediction and the motion compensation represent more than 60% of the decoding time. In fact, the inter-layer prediction includes the up-sampling and the MVs up-scaling of the interlayer reference picture which are widely used in the decoding process.

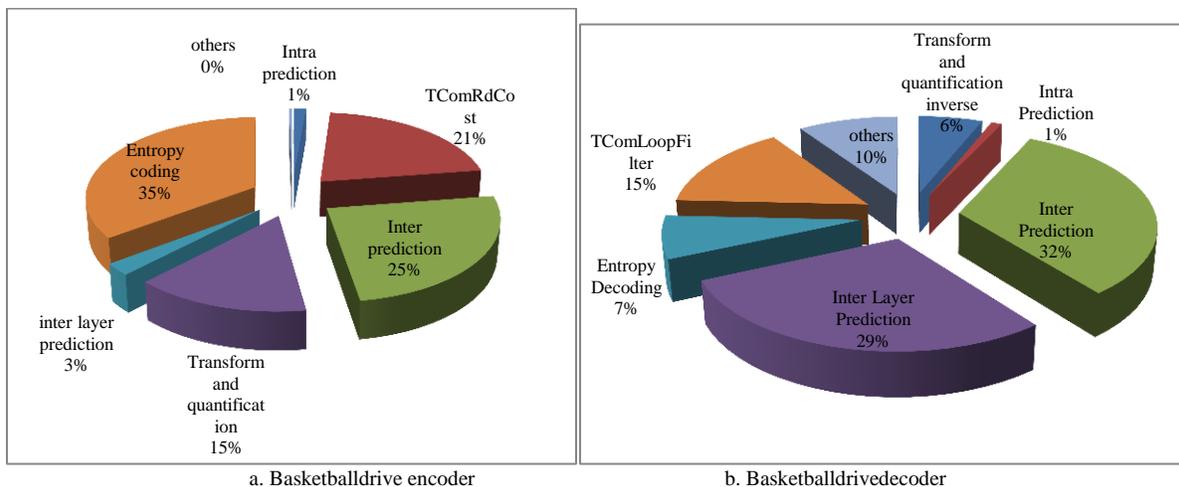


Fig. 4. Average time distribution in both encoder and decoder for Basketballdrive video sequence

##### B. Coding efficiency

In order to highlight the SHVC coding efficiency, we considered the following coding scenarios:

- Single layer coding solutions: For HEVC EL single layer coding, the recent available reference software HM v.16.0 [26] was used.
- Scalable Extension of the High Efficiency Video Coding (SHVC) and its single layer EL were generated with the available software reference SHM v.7.0 [23].
- Simulcasting HEVC solutions: correspond to multiple independent single layer coding and the total sum of EL and BL were used for comparison. The various layers were independently coded with the HEVC standard.

The experiments for spatial 1.5x, 2x, and SNR scalability were performed following the test conditions already fixed in [24]. We provided the average values of BD rate for random access (RA) and low delay (LD) configurations. While BD rate represented the average Bjontegaard[27].

TABLE III. CODING EFFICIENCY OF SHVC COMPARED TO HEVC SIMULCASTS

	Random access			Lowdelay		
	1.5x	2x	SNR	1.5x	2x	SNR
Kimono	-29,6	-13,3	-43,9	-36,3	-31,5	-32,7
BQTerrace	-10,0	-5,0	-45,4	-30,9	-34,2	-37,5
Cactus	-15,4	-9,6	-43,1	-28,3	-30,0	-40,7
BasketBallDrive	-21,1	-13,7	-45,2	-16,2	-22,5	-41,8
Packscene	-15,8	-5,9	-42,6	-20,4	-25,54	-41,4
PeopleOnStreet	N/A	-41,5	-8,2	N/A	-37,2	-43,0
Traffic	N/A	-18	-27,6	N/A	-13,6	-13,6
<b>Average</b>	<b>18,4%</b>	<b>15,25%</b>	<b>36,6%</b>	<b>26,4%</b>	<b>27,2%</b>	<b>35,8%</b>

TABLE III shows the different coding performances provided by SHVC and simulcast using the total EL+BL bitstreams. For the RA configuration, SHVC achieved a BD rate reduction of 18.4%, 15.25%, and 36.6% compared to simulcast for 1.5x, 2x, and SNR scalability, respectively. However, for LD configuration, the bit rate decreases of about 26.4%, 27.2%, and 35.8% for 1.5x, 2x, and SNR scalability, respectively.

TABLE IV. CODING EFFICIENCY OF SHVC COMPARED TO HEVC SINGLE LAYER CODING EL

	Random access			Lowdelay		
	1.5x	2x	SNR	1.5x	2x	SNR
Kimono	12,9	30,8	12,2	20,7	28,2	21,6
BQTerrace	19,7	32,0	9,2	19,3	13,6	23,7
Cactus	27,4	26,2	13,9	32,8	29,2	18,7
BasketBallDrive	19,2	21,7	9,6	35,4	25,4	26,3
Packscene	25,5	28,5	14,7	27,1	18,7	17,1
PeopleOnStreet	N/A	11,5	59,5	N/A	9,3	N/A
Traffic	N/A	13,0	42,1	N/A	18,5	N/A
<b>Average</b>	<b>20,9%</b>	<b>23,41%</b>	<b>23%</b>	<b>27,1%</b>	<b>20,44%</b>	<b>25,5%</b>

In TABLE IV, SHVC is compared to HEVC simulcast coding for only EL resolution. In such a case, the BD rate increases by 20.9%, 23.41%, and 23% for RA configuration and 27.1%, 20.44%, and 25.5% for LD configuration for 1.5x,

2x, and SNR scalability, respectively. The increase of BD rate is explained by the fact that when broadcasting a scalable bitstream, a lower resolution one may be simply extracted and decoded. Thus, encoding two resolutions is more costly in terms of bit rate if compared to encoding a single resolution.

TABLE V. CODING EFFICIENCY OF SHVC EL COMPARED TO HEVC SINGLE LAYER CODING EL

	Random access			Lowdelay		
	1.5x	2x	SNR	1.5x	2x	SNR
Kimono	48,2	-32,28	-40,3	-61,8	-26,4	-32,7
BQTerrace	-21,0	-8,27	-23,2	-21,0	-15,4	-9,7
Cactus	-34,6	-14,5	-32,8	-32,4	-13,9	-25,6
BasketBallDrive	-41,2	-19,7	-37,0	-34,8	-16,7	-1,2
Packscene	-34,7	-10,0	-34,5	-34,1	-19,4	-29,7
PeopleOnStreet	N/A	-59,1	-55,4	N/A	-32,58	-27,9
Traffic	N/A	-16,0	-27,2	N/A	-21,9	-21,9
<b>Average</b>	<b>16,7%</b>	<b>22,85%</b>	<b>35,8%</b>	<b>36,8%</b>	<b>20,92%</b>	<b>21,3%</b>

TABLE V shows the BD rate results when comparing SHVC versus HEVC for EL only. We note that the BD rate decreases by 16.7%, 22.85%, and 35.8% for RA configuration and 36.8%, 20.92%, and 21.3% for LD configuration for 1.5x, 2x, and SNR scalability, respectively.

When comparing SHVC extension to different scenarios as described above, we note that even when bit rate performance depends on sequence features, the scalable extension is more efficient and preferment compared to other scenarios due to the contribution of the new inter layer prediction coding tool.

V. STATISTICAL OBSERVATION

Since the importance of the SHVC standard was highlighted through the BD rate comparison, a further analysis was necessary in order to achieve good SHVC understanding. Consequently, a statistical analysis of SHVC was performed. This section is organized as follow: the first part gives the statistics at CUs level for BL and EL, while the second part gives the statistics at PUs level for BL and EL. Analysis was performed to all test sequences with 4xGOPs of frames/sequence (The GOPs are fixed in the common test conditions).

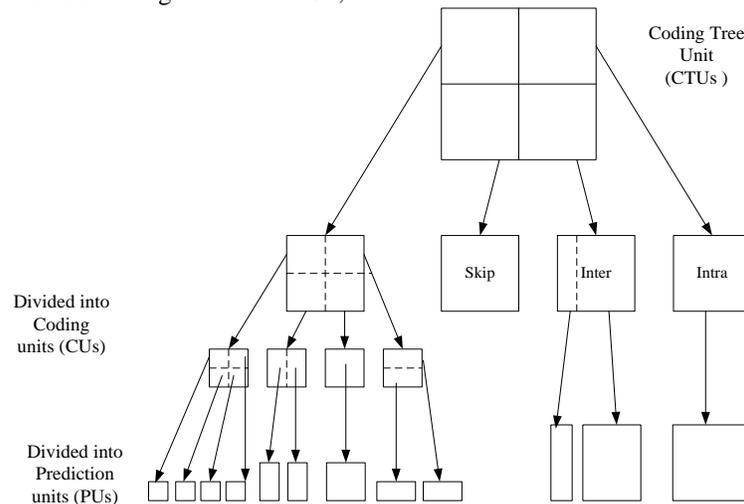


Fig. 5. Coding tree unit is subdivided into CUs along the associated coding quad tree

In addition, a statistical analysis of SHVC for Random access configuration was carried out for different pairs of QPs {34, 36} and {22, 24} for A and B test sequences classes. In fact, the coding layer in HEVC was based on the coding Tree Unit (CTU) using the quad tree structure. As shown in Fig.5. the CTU can be split into  $M \times M$ ,  $M/2 \times M/2$ ,  $M/3 \times M/3$ , and  $M/4 \times M/4$  with  $M$  equal to 64 [28], depending on their respective distortion rates. Furthermore, each CU can be split into prediction units (PUs) {symetric:  $M \times M$ ,  $M \times M/2$ ,  $M/2 \times M$ ,  $M/2 \times M/2$ , and asymeric:  $M \times M/4$ ,  $M \times 3M/4$ ,  $M/4 \times M$ , and  $3M/4 \times M$  } then into transform units (TUs) [29]. We focus in this section on statistics at the CUs and PUs levels as revealed in Fig.5.

A. Statistics at CU level:

At the CUs level, we extracted the percentages of the prediction in Intra, Inter, and Skip modes for BL and EL for each type of frames (I, P, and B). Since the SHVC adopts temporal scalability, a hierarchical prediction structure was

used. In fact, the B pictures may refer to several temporal levels. Thus, the B0 picture corresponds to the first temporal level, while B1, B2, and B3 referred to the second, the third, and the fourth levels, respectively. Analysis was performed for all test sequences for the two studied classes. However, we present analysis details for BasketballDrive video sequence for two different QPs values.

Fig.6. shows the distribution percentages for intra, inter, and skip modes for each type of frames. For BL, all I frames are coded as intra. P frames in EL are considered as I frames in BL. However, they were predicted not only in intra, but also in inter modes. This is explained by the fact that, in addition to the intra, inter and skip modes, we have the new inter-layer prediction modes introduced in SHVC.

For BL as well as for EL, the percentage of intra-prediction modes for B frames decreases and gets close to zero for the last temporal level T3. We may notice also that the results are almost similar for both high and low QPs.

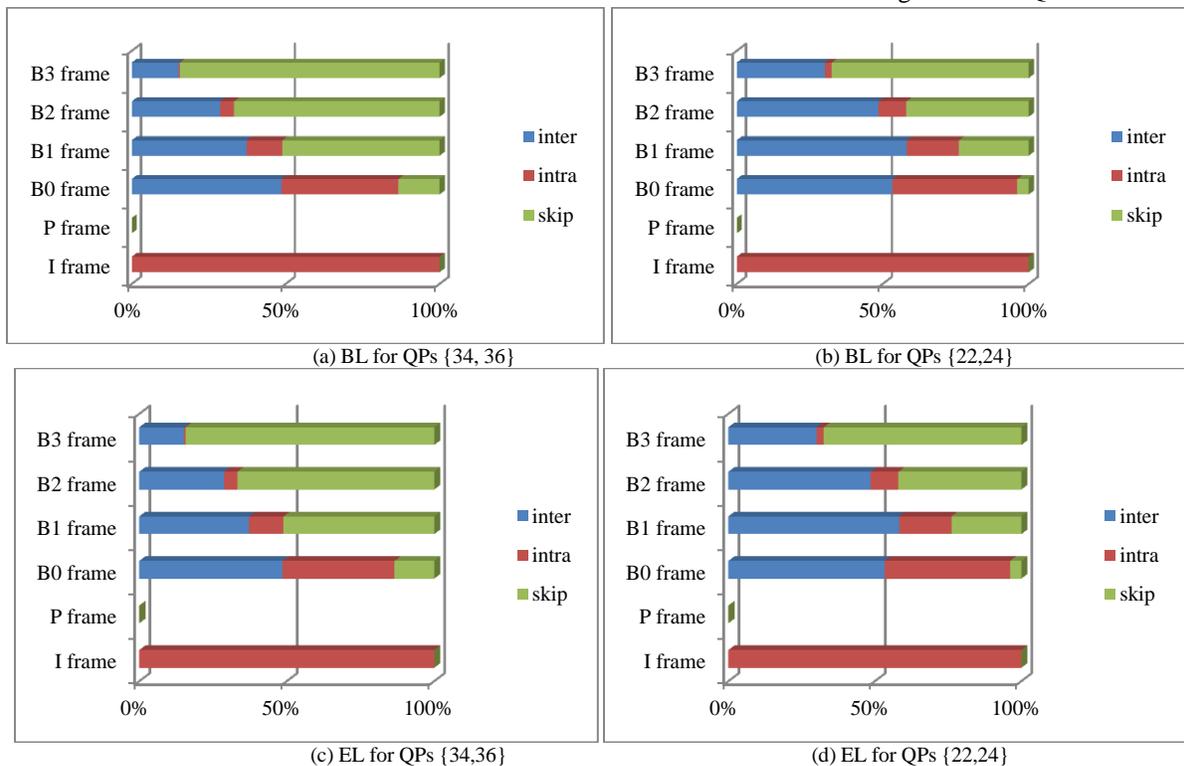


Fig. 6. Percentages of Intra, Inter, and Skip mode predictions at the CUs level. (a) BL for QPs{34,36}.(b) BL for QPs{22,24}.(c) EL for QPs{34,36}.(d) EL for QPs{22,24}

For further analysis, we focused on each depth distribution for each prediction mode; intra, inter, and skip modes for each frame type. This analysis was performed for each layer (BL and EL) and for different QPs pairs.

1) BL:

a) Inter-prediction mode: As depicted in Fig.7, we note that depth 0, where CU size is equal to 64x64, is usually the most selected for B1, B2, and B3 frames. For B0 frames, the most significant depth is depth 1 where CU size is equal to 32x32.

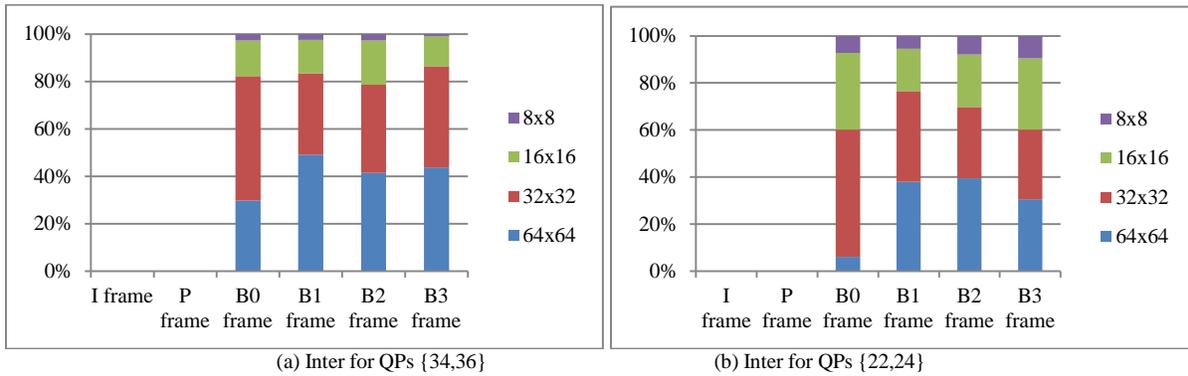


Fig. 7. Percentages of CU's size for inter-modes prediction for BL. (a) Inter for QPs{34,36}.(b) Inter for QPs{22,24}

b) *Skip mode*: Fig.8 shows that skip distribution depends obviously on QPs values. For QPs {34, 36}, depth 0 (64x64) is usually the best mode for B1, B2, and B3 frame. Whereas, depth 1 is generally selected for B0 frame. For

smaller QPs values {22,24}, we note the importance of depths 0, 1, and 2 for B1, B2, and B3. However, for B0, depth 0 is less selected compared to the other 3 depths.

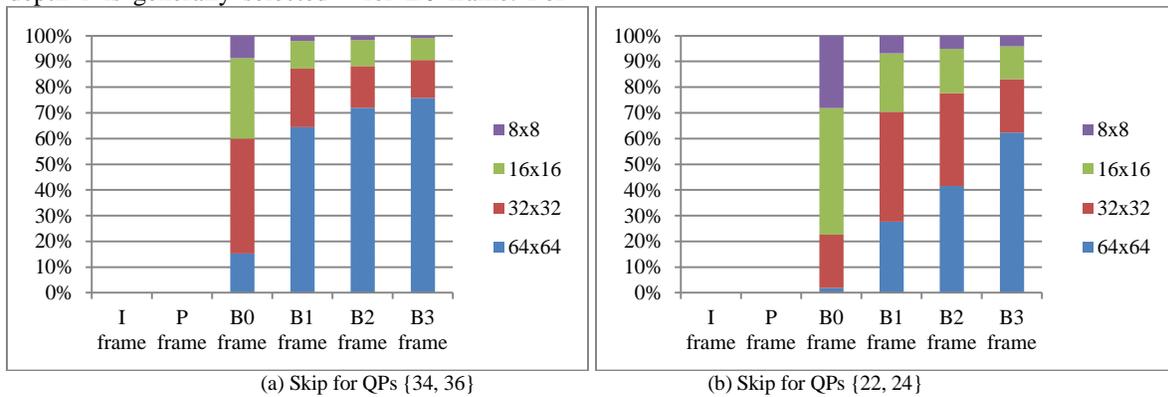


Fig. 8. Percentages of CU's size used for skip modes prediction for BL. (a) Skip for QPs {34, 36}. (b) Skip for QPs {22, 24}

c) *Intra-prediction mode*: When analyzing results presented in Fig.9, we note that depth 0 is selected only with high QPs values. The CU sizes are proportional to the QPs

values. In fact, as QPs values decrease, smaller CUs sizes are selected.

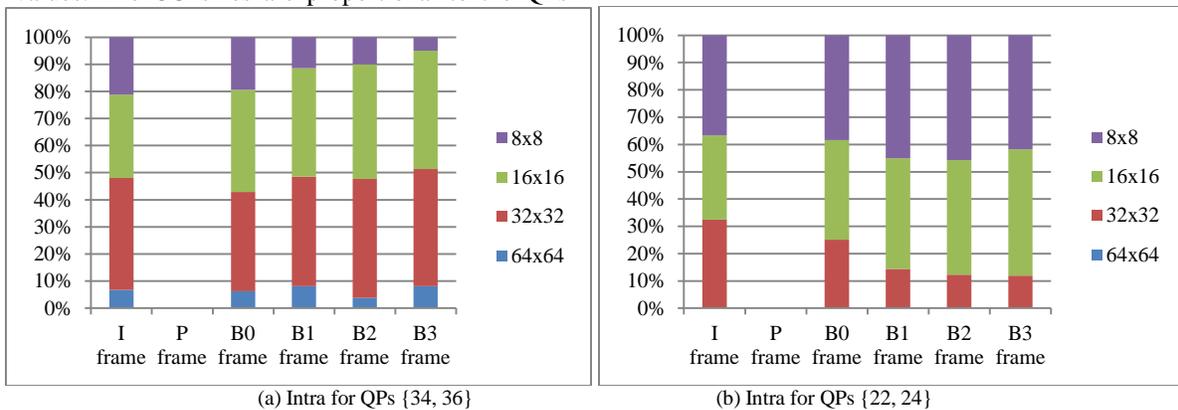


Fig. 9. Percentages of CU's size used for intra-modes prediction for BL. (a) Intra for QPs {34, 36}. (b) Intra for QPs {22, 24}

2) *EL*:

For EL, we note the appearance of the P frames in the inter-prediction modes. We observe also the existence of P frames, instead of I frames, in the intra-prediction; this is explained by the fact that P frames are considered here as I frames. They are

not only predicted in intra-mode, but also in inter-layer prediction.

a) *Inter-prediction mode*: as shown in Fig.10, for QPs pair {34,36} depth 0 is the most selected for B1, B2, and B3 frames. Contrarily to depth 1 which was the most selected for P and B0 frames.

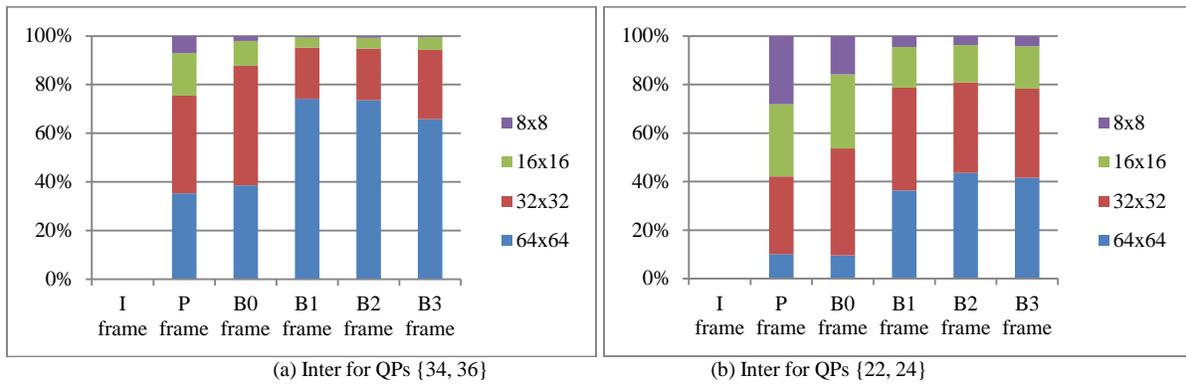


Fig. 10. Percentages of CU's size used for inter-modes prediction for the EL. (a) Inter for QPs {34, 36}. (b) Inter for QPs {22, 24}

While decreasing the QPs values, depth1 will be more selected for B1, B2, and B3. However, for B0 and P frames, depth2, and depth3 are more used.

b) *Skip mode*: the distribution of this mode is illustrated in Fig.11. We notice that this mode depends constantly on the QPs values. For QPs {34, 36}, depth 0 presents the higher

percentages for B1, B2, and B3 frames. However, for B0 frames, depth 0 and depth 1 percentages are almost equal. Depth 0 usually goes along lower QPs values. Depth 1 becomes more important with higher temporal level. Therefore, for B0, we note the appearance of depth 2 and depth 3.

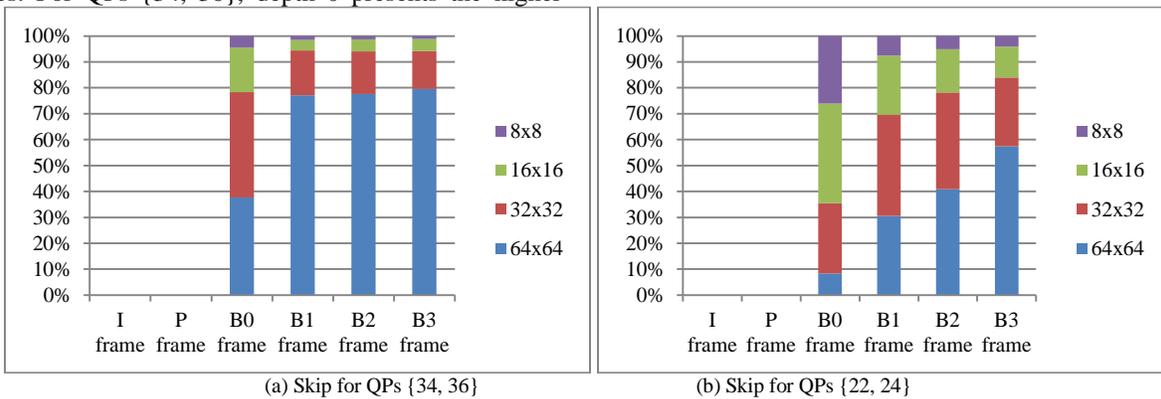


Fig. 11. Percentages of CU's size used for Skip modes prediction for the EL. (a) Skip for QPs {34, 36}. (b) Skip for QPs {22, 24}

c) *Intra-prediction mode*: we note, in Fig.12, that depth 0 is absent with low QPs values and widely used for higher

QP values. Moreover, we remark the importance of depth3 for lower QPs values.

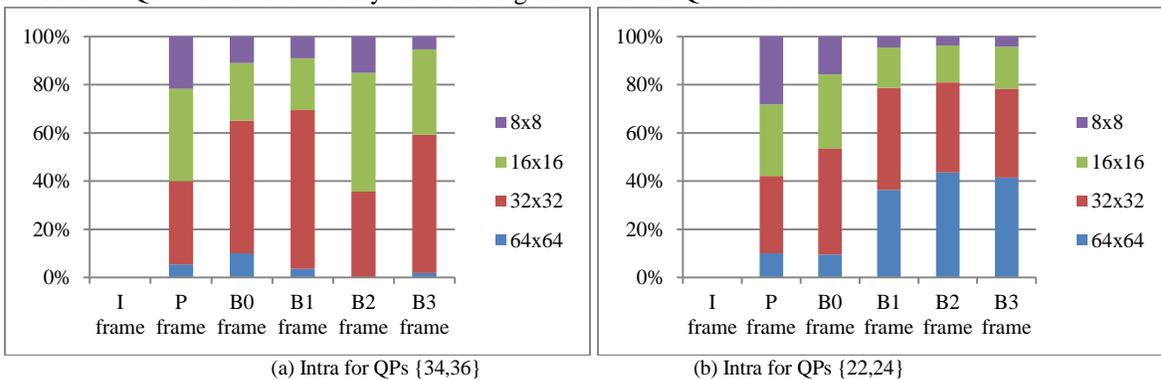


Fig. 12. Percentages of CU's size used for Intra-modes prediction for the EL. (a) Intra for QPs {34, 36}. (b) Intra for QPs {22, 24}

### 3) Sequences comparison:

In the previous sub-section, the sequence test Basketballdrive is considered as a sequence with high motion. It uses larger blocs of CUs than low motion sequences, such as BQterrace. In Fig.13, we compare BQterrace and Basketballdrive. We observe that, for Skip mode prediction,

the CU's sizes for B3 images are used with 92.90% compared to 63.92% used in BQterrace. The use of larger CUs is noted also for the other types of frames. The use of larger bloc sizes is due to the fact that the video sequence presents a high motion. Contrarily to slow motion video sequences, where the motion is slow consequently smaller CUs are used.

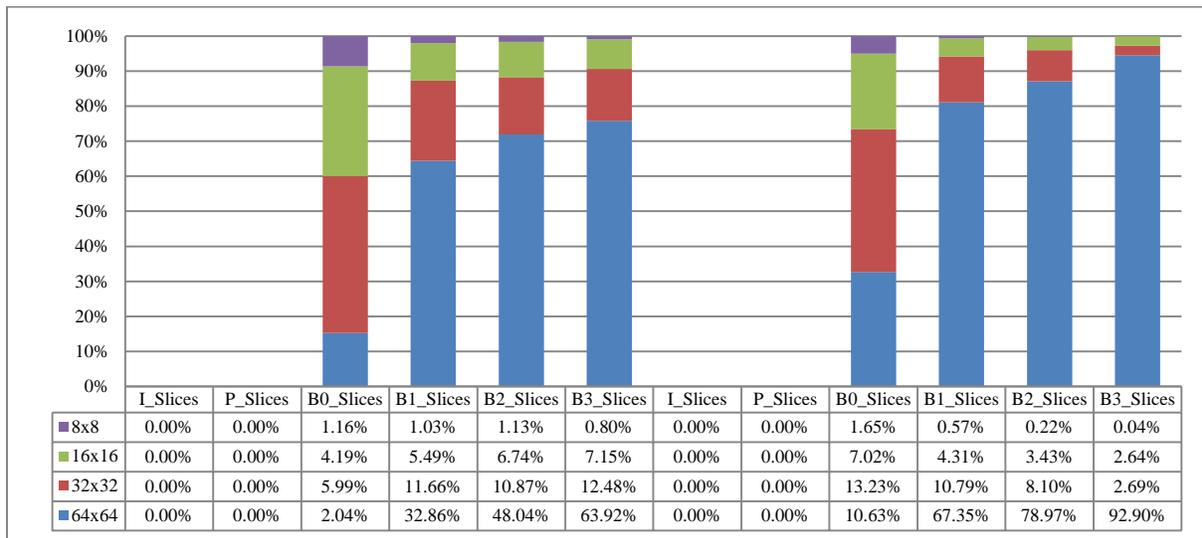


Fig. 13. Comparison of Skip mode percentages provide by BQterrace and Basketballdrive

Another comparison was performed between textured and non-textured test sequences. In Fig.14, we compare PeopleOnstreet and traffic sequences. The observation significantly shows that the choice of the CU's size depends on the characteristics of the image including the texture parameter which considerably affects the frequency of using the CU's sizes. In fact, PeopleOnStreet is considered as a textured video

sequence that required small CU's size to code the spatial details on the scene. Consequently, we notice that 12.95% of bloc size of 8x8 was used for coding B2 images. However, for traffic sequence, where smooth areas dominate the scene, we notice that only 2.12% of small CU sizes were used for B2 images and almost zero for B3 images.

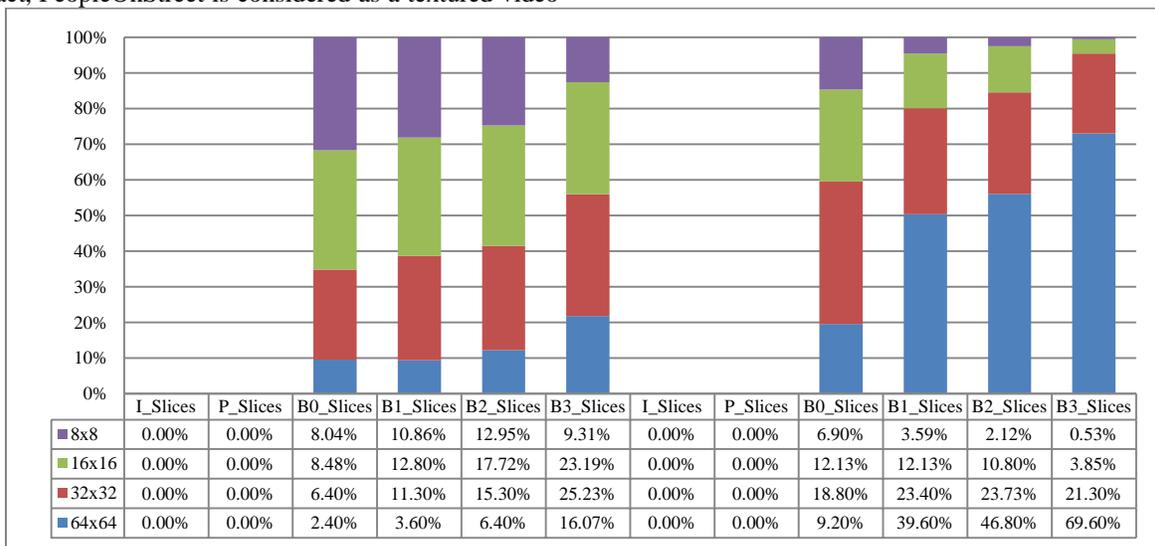


Fig. 14. Skip mode percentages comparison between peopleonstreet and Traffic

A comparison between BL and EL resolutions is presented in Fig 15. Results are nearly the same for the two resolutions. It will be wise to compare the results for the two scalability ratios

1.5x and 2x to ensure that future proposed optimization method will be optimum for all resolutions. Statistics, presented in Fig.16, show similar percentages for both resolutions.

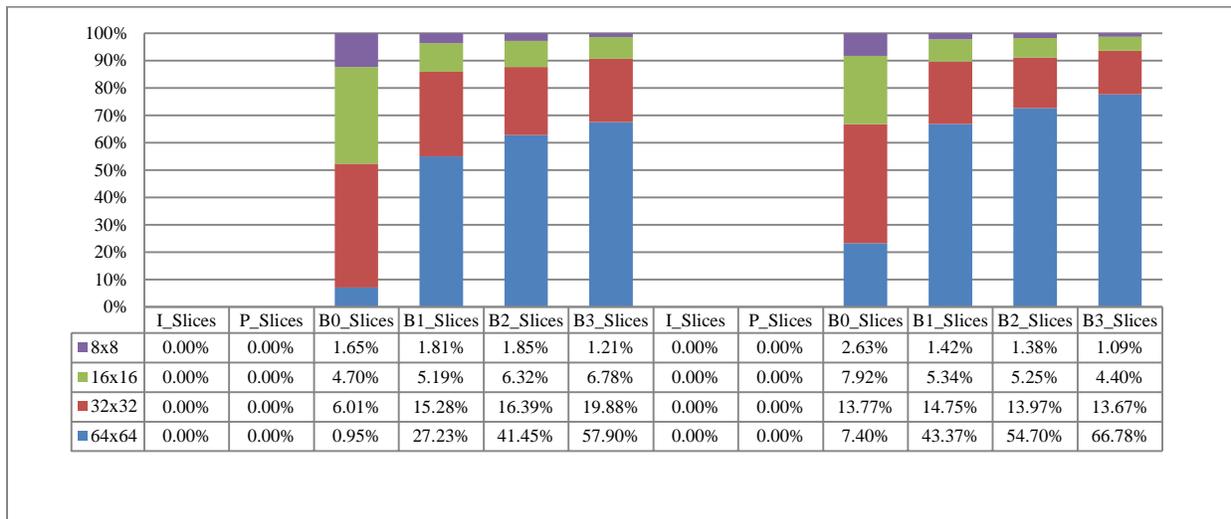


Fig. 15. Comparison between BL and EL video sequences statistics for Basketball drive video

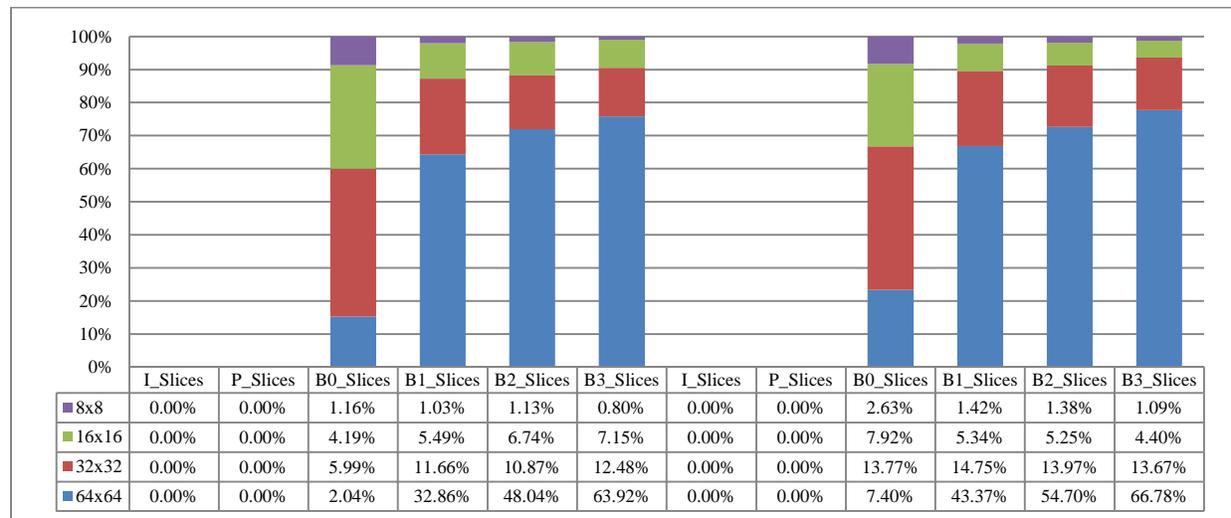


Fig. 16. Comparison between two scalability ratios 1.5x and 2x for the same BL Basketball drive video

When analyzing statistics for all test sequences videos supported in the common test conditions for two QPs pairs {34,36} and {22,24} and for two scalability ratios 1.5 and 2, we notice that smaller CUs sizes are selected more than larger ones for lower QPs values. While maximizing the QPs values, the selection of larger CUs sizes increases, as it is the case in Skip mode with an average of 73.26% for QPs{34,36} of 64x64 CUs size for B3 images compared to 34.43% for QPs{22,24}. The observation significantly shows that the frequency of the CUs size choice depends also on the video sequence characteristics. Consequently, the choice of the CUs size relies on the characteristics of the image including the texture parameter which can affects remarkably the CU's sizes selection frequency. It is also important to cite that non-motion videos use larger CUs sizes than those with higher motion. From Fig.15, It is obvious that the prediction modes in EL follow those in BL in terms of percentages. Those out coming observations will be the milestones for introducing optimized algorithms leading to an efficient SHVC implementation.

### B. Statistics at the PU's level:

As a second statistical analysis, we focus on the PU's level for all test sequences and for two QP's pairs {22,24} and {34,36}. Statistical results are presented in details only for Basketballdrive due to the similarity of the results. Obviously, the interlayer prediction is present only in the enhancement layer but does appear only at the PU level. Analysis is made for all types of frames, all depths, both EL and BL resolution, and for 1.5x and 2x resolutions.

#### 1) BL:

At PU level, the prediction mode is either inter or intra as presented in Fig.17. We note that I images are totally predicted in intra. Then, the percentages of intra-prediction decrease in the B images. In the BL, there is no interlayer prediction, which explains the absence of P images. For QPs {22\_24}, the percentage of intra-prediction is more important than prediction for higher QP {34\_36}. For further analysis, we focus on each depth distribution for each prediction mode; intra and inter mode for each frame type.

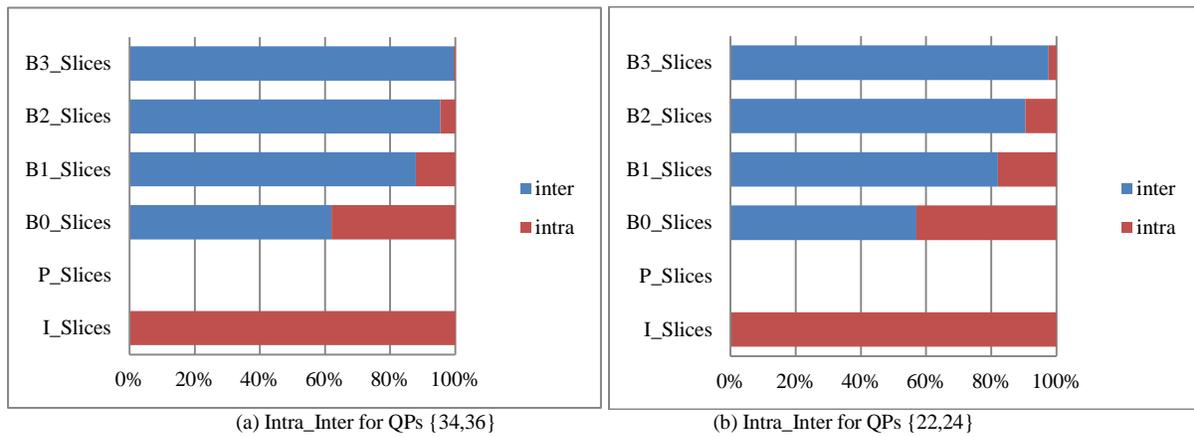


Fig. 17. Percentages of depth used for Inter and Intra modes prediction at the PUs level for BL. (a) QPs {34, 36}. (b) QPs {22, 24}

a) *Inter-prediction*: From, Fig.18, we note that the percentage of MxM bloc size is the highest for B images and increases considerably with the increment of the temporal

level. We observe that the asymmetric blocs are less used for high as well as low QPs values.

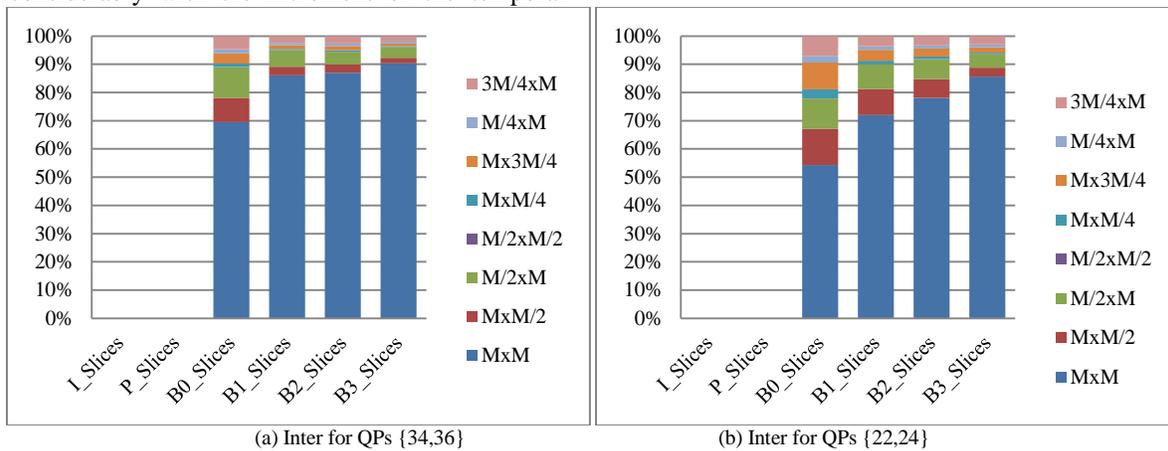


Fig. 18. Percentages of depth used for Inter modes prediction at the PUs level in the BL. (a) QPs {34, 36}. (b) QPs {22, 24}

b) *Intra-prediction*: MxM bloc sizes are more used compared to M/2xM/2 bloc sizes. The use of M/2xM/2 bloc sizes increase when QPs values are low as depicted in Fig.19.

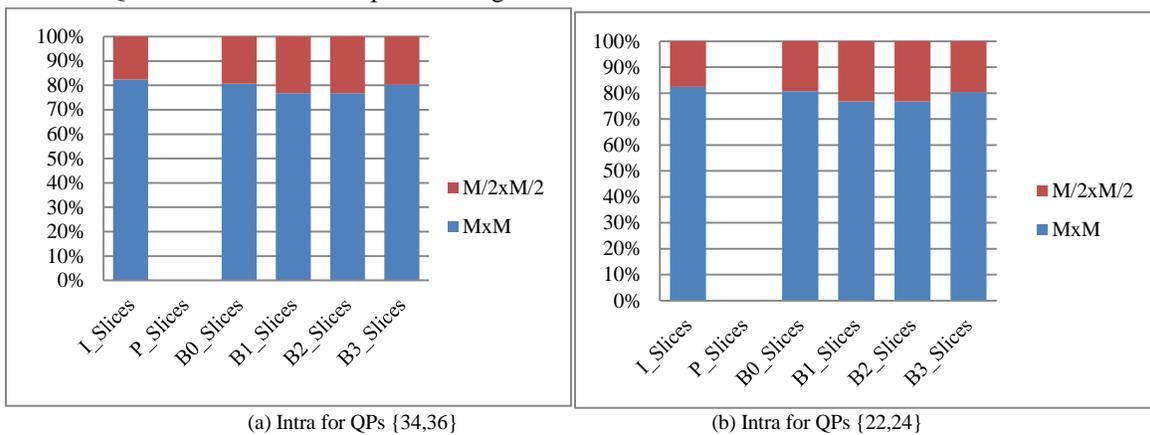


Fig. 19. Percentages of depth used for Intra modes prediction at the PUs level in the BL.(a) QPs {34, 36}.(b) QPs {22, 24}

2) *EL*:

In this sub-section, we focus on EL distribution. As depicted in Fig.20, we notice the presence of P images caused

by the use of interlayer prediction. We note also the important percentages of inter-layer mode. For example, for B1, B2, and B3 frames, prediction mode is almost inter-layer. Besides, the

partition percentages of inter and intra-prediction for EL and for each frame are almost the same for those of BL. The unique

difference is that the intra percentages are higher for low QPs. We investigate now each prediction mode.

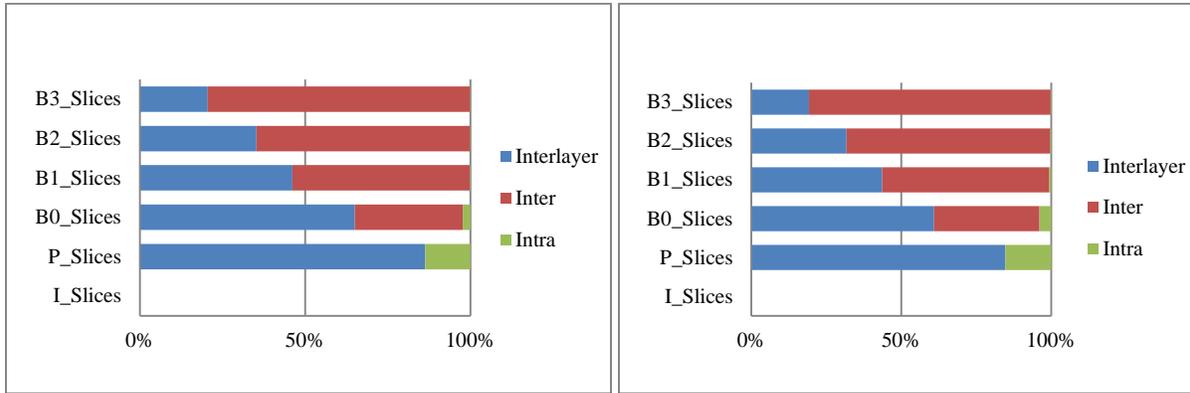


Fig. 20. Percentages of depth used for Intra and Inter modes prediction at the PUs level in the EL. (a) QPs {34, 36}. (b) QPs {22, 24}

a) *Inter-layer prediction:* We note that the P images are totally predicted in inter-layer prediction as noted in Fig.21.

Asymmetric partitions are less used than symmetric ones and the MxM size are the most used for all images.

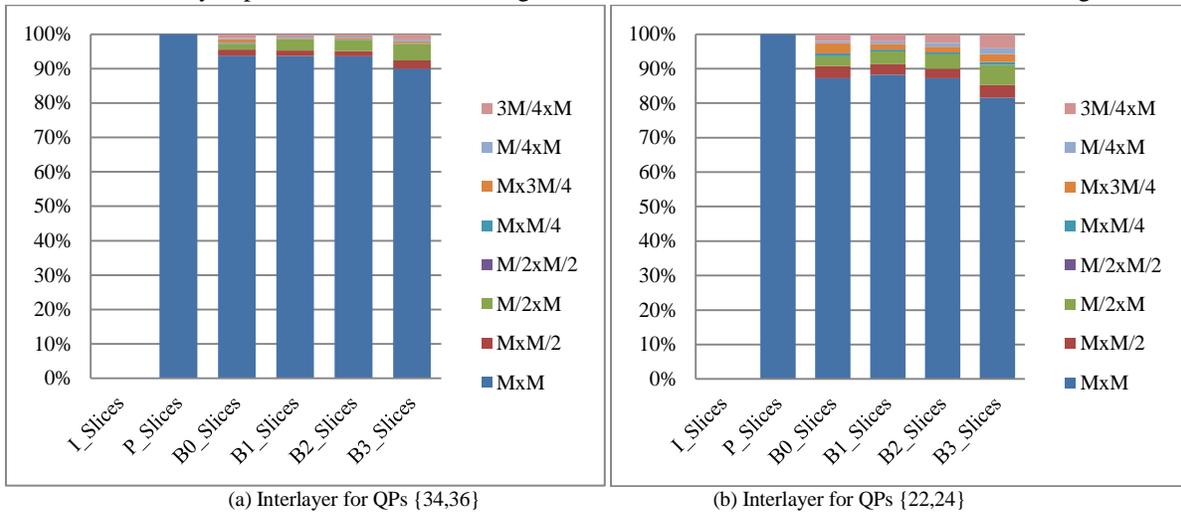


Fig. 21. Percentages of depth used for Inter layer modes prediction at the PUs level in the EL. (a) QPs {34, 36}. (b) QPs {22, 24}

b) *Inter-prediction:* as seen in Fig.22 MxMsize is usually more used in all B frames. Asymmetric blocs are less

used. However, for lower QPs, asymmetric blocs are more applied for B0 images.

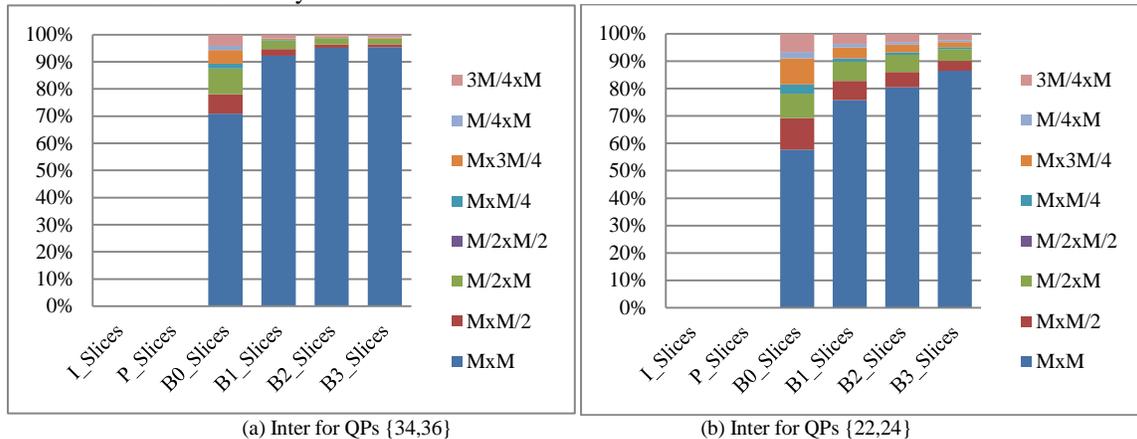


Fig. 22. Percentages of depth used for Inter-modes prediction at the PUs level in the EL. (a) QPs {34, 36}. (b) QPs {22, 24}

c) *Intra-prediction*: Statistical results are shown in Fig. 23; intra-prediction is more used in lower QPs pair. We notice

that MxM sizes are usually the most used ones. Nevertheless, for lower QPs values, the M/2xM/2 bloc sizes are more used.

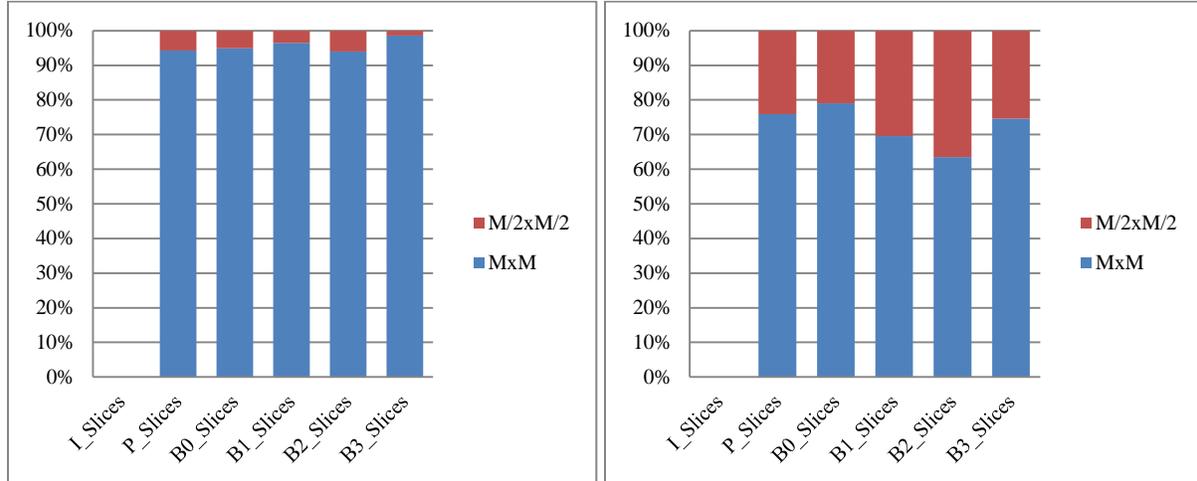


Fig. 23. Percentages of depth used for Intra modes prediction at the PUs level in the EL. (a) QPs {34, 36}. (b) QPs {22, 24}

3) *Sequences comparison*:

When comparing two different sequences for inter-layer prediction, as detailed in Fig. 24, we observe that symmetric bloc sizes, especially the MxM size, are the most used. Despite the excessive use of MxM bloc sizes, the difference in terms of

percentages between high motion and low motion videos is very clear. We note that high motion video, such as basketballdrive, uses 60.97% of MxM blocs for B0 images compared to 26.65% in BQterrace.

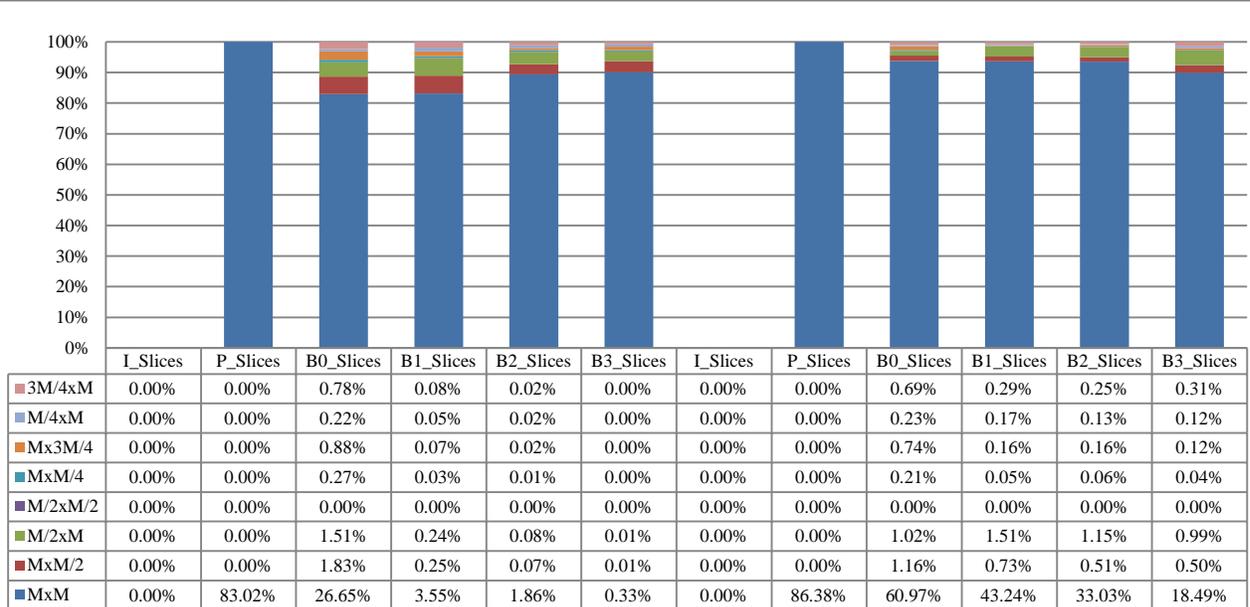


Fig. 24. Interlayer mode percentages comparison between BQterrace and Basketballdrive

As depicted in Fig. 25, the comparison of two 2x video sequences shows that the use of smaller bloc sizes increases for textured videos. Results show clearly that percentages of asymmetric partitioning are higher for PeopleOnStreet 2.01%

for B1 images compared to 0.33% in Traffic. It is also obvious that percentage for Inter-layer prediction depends on sequence characteristics. Consequently, experimental results reveal that non-textured sequences use more inter-layer prediction.

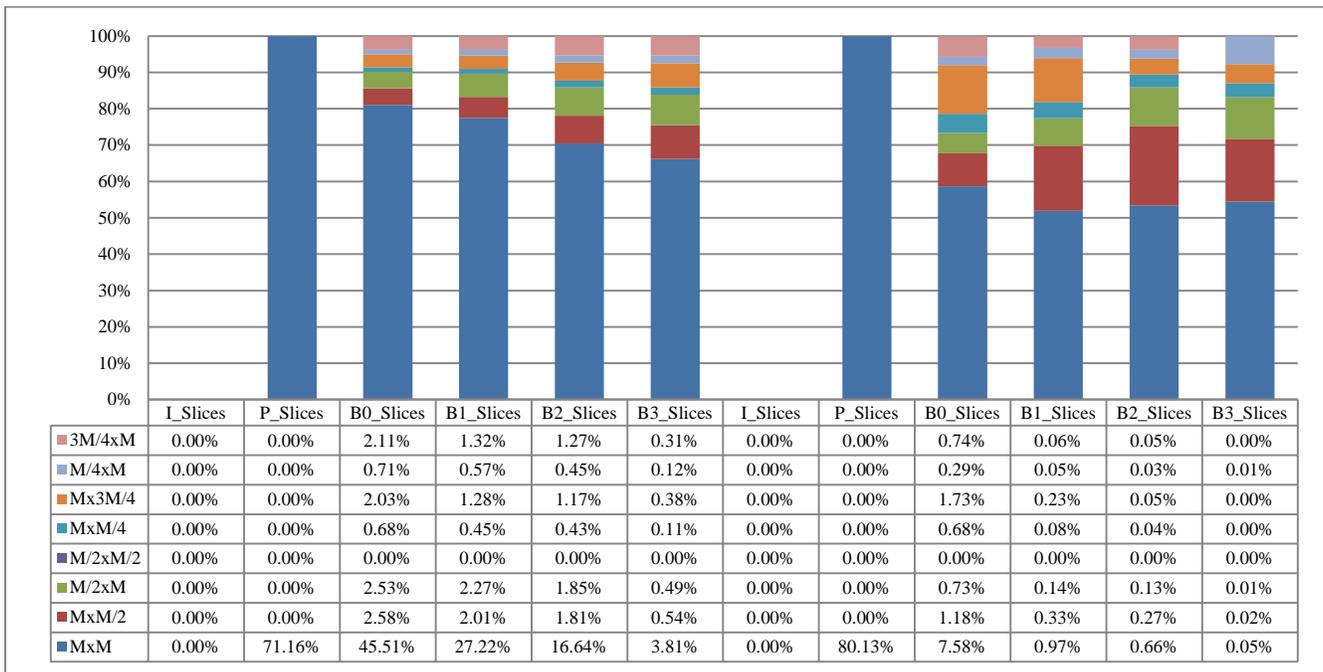


Fig. 25. Interlayer mode percentages comparison between PeopleOnStreet and Traffic

Fig.26 shows the moderate use of asymmetrical blocs for the same sequence with the 2x ratio compared to the 1.5x. In general, the results obtained, in the case of two different

scalability ratios for the same video sequence, are nearly the same as it was seen at the CU's level statistics.

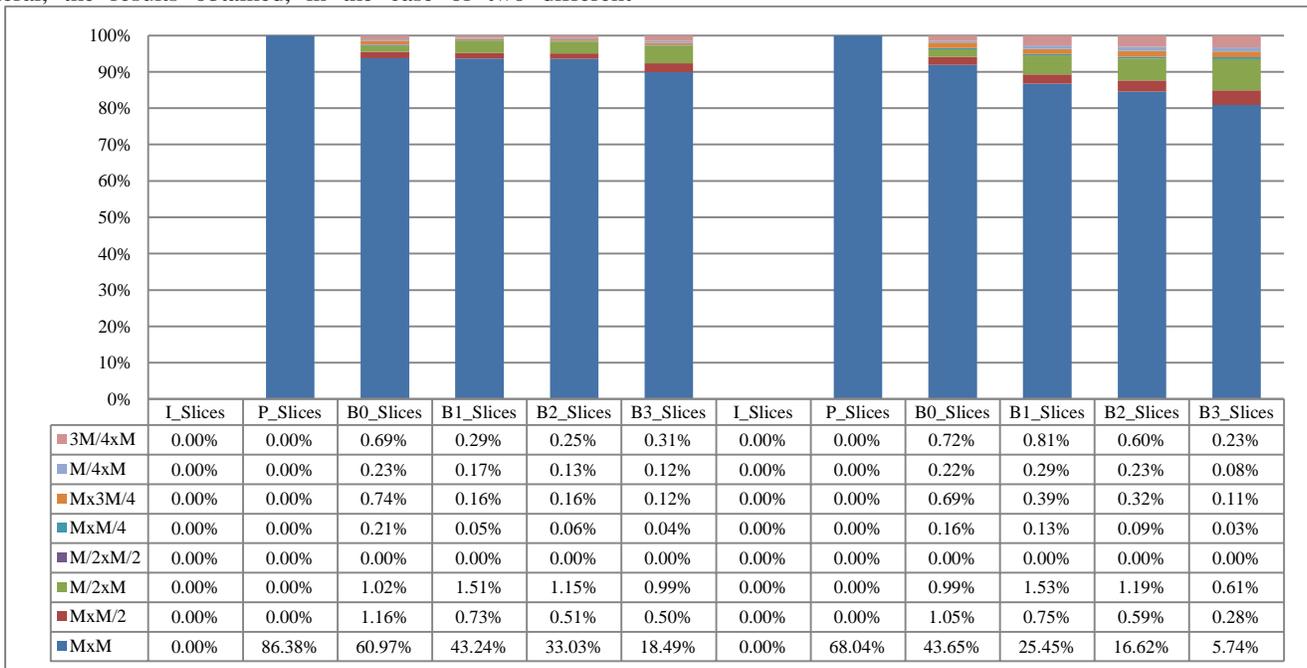


Fig. 26. Comparison between Two scalability ratios 1.5x and 2x for the same EL Interlayer Basketball drive

## VI. CONCLUSION

This paper presents a statistical analysis for partitioning units used to encode videos with HEVC scalable extension SHVC. Statistical analysis was carried out for all video test sequences as well as for different resolution ratios, especially for each frame and prediction type. The obtained results allowed us to conclude that the choice of the partitioning units depends on either the frame type or the characteristics of the video sequence: texture, resolution and motion. Consequently, smallest coding unit and prediction unit were used for coding texture sequence as an example. Furthermore, it was obvious that the distribution mode between BL and EL is strongly correlated. In fact, this observation is an important step for proposing efficient algorithms in order to speed up the encoding process with minor PSNR loss.

### REFERENCES

- [1] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1648–1667, December 2012.
- [2] Zhongbo Shi, Xiaoyan Sun, Feng Wu: "Spatially Scalable Video Coding For HEVC". *IEEE Trans. Circuits Syst. Video Techn.* 22(12): 1813-1826 (2012)
- [3] H. Lee et al, "Scalable Extension of HEVC for Flexible High-Quality Digital Video Content Services", *ETRI Journal*, vol. 35, no. 6, pp. 990-1000, Dec. 2013.
- [4] P. Helle, H. Lakshman, M. Siekmann, J. Stegemann, T. Hinz, H. Schwarz, D. Marpe, and T. Wiegand, "A Scalable Video Coding Extension of HEVC," in *IEEE Conference on Data Compression*, March 2013, pp. 201–210.
- [5] Gary J.Sullivan, Jill M.Boyce, Ying Chen, Jens-Rainer Ohm, C.AndrewSegall, Anthony Vetro "Standardized Extensions of High Efficiency Video Coding (HEVC)" *IEEE journal of selected topics in signal processing*, vol. 7, no 6, December 2013
- [6] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding standards including High Efficiency Video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1969–1684, December 2012.
- [7] E. S. Ryu, Y. Ryu, H. J. Roh, J. Kim and B. G. Lee, "Towards robust UHD video streaming systems using scalable high efficiency video coding," *Information and Communication Technology Convergence (ICTC), 2015 International Conference on*, Jeju, 2015, pp. 1356-1361.
- [8] S.-F. Tsai et al, "A 1062 Mpixels/s 8192x4320p high efficiency video coding (H.265) encoder chip", 2013 Symposium on VLSIC, pp. 188-189, 2013.
- [9] R. Takada et al, "s", *IEEE International Conference on Consumer Electronics (ICCE 2015)*, Jan. 2015.
- [10] K. Rapaka, J. Chen and M. Karczewicz, "Efficient key picture and single loop decoding scheme for SHVC," *Visual Communications and Image Processing (VCIP), 2013*, Kuching, 2013, pp. 1-6.
- [11] JVT Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264-ISO/IEC 14496-10 AVC), Mar. 2003, JVT-G050, available on [http://ip.hhi.de/imagecom\\_G1/assets/pdfs/JVT-G050.pdf](http://ip.hhi.de/imagecom_G1/assets/pdfs/JVT-G050.pdf)
- [12] L. Chen, M. M. Hannuksela and H. Li, "Disparity-compensated inter-layer motion prediction using standardized HEVC extensions," *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*,
- [13] Yan Ye; Andrivon,P., "The scalable Extensions of HEVC for Ultra-high\_definition video delivery" in *Multimedia, IEEE*, vol.21 , no.3.,58-64, July-Sept.2014
- [14] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramanian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20-34, Jan. 2016.
- [15] Hamidouche, W.; Raulet, M.; Deforges, O., "Real time SHVC decoder: Implementation and complexity analysis," in *Image Processing (ICIP), 2014 IEEE International Conference on* , vol., no., pp.2125-2129, 27-30 Oct. 2014
- [16] Hamidouche, W.; Raulet, M.; Deforges, O., "Parallel SHVC decoder: Implementation and analysis," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on* , vol., no., pp.1-6, 14-18 July 2014
- [17] Amina Kessentini, Amine Samet, Mohamed Ali Ben Ayed, Nouri Masmoudi " Fast Mode Decision Algorithm for H.264/SVC Enhancement Layer", Springer-Verlag 2013, J Real-Time Processing, DOI10.1007/s11554-013-0362-1
- [18] H. L. Tan, C. C. Ko and S. Rahardja, "Fast Coding Quad-Tree Decisions Using Prediction Residuals Statistics for High Efficiency Video Coding (HEVC)," in *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 128-133, March 2016.
- [19] F. Belghith, H. Kibeya, M. A. Ben Ayed and N. Masmoudi, "Statistical analysis and parametrization of HEVC encoded videos," *Information Technology and Computer Applications Congress (WCITCA), 2015 World Congress on*, Hammamet, 2015, pp. 1-5.
- [20] R. Bailleul, J. De Cock and R. Van De Walle, "Fast mode decision for SNR scalability in SHVC digest of technical papers," *2014 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2014, pp. 193-194.
- [21] H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos and J. Slevinsky, "A new mode for coding residual in scalable HEVC (SHVC)," *2015 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2015, pp. 372-373.
- [22] H. R. Tohidypour, M. T. Pourazad and P. Nasiopoulos, "Probabilistic Approach for Predicting the Size of Coding Units in the Quad-Tree Structure of the Quality and Spatial Scalable HEVC," in *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 182-195, Feb. 2016.
- [23] ISO/IEC-JTC1/SC29/WG11 and ITU-T SG 16 WP 3, "Scalable HEVC (SHVC) Test Model 7(SHM 7)," in *ISO/IEC JTC 1/SC 29/WG11 (MPEG) Doc. N14705 or ITU-T SG 16 Doc. JCTVC-R1007\_v1*. Sapporo, Japan, July 2014
- [24] V. Seregin and Y. He, "Common SHM test conditions and software reference configurations," in document *JCTVC-O1009*. Geneva, Switzerland, November 2013.
- [25] I. Wali, A. Kessentini, M. Ali Ben Ayed and N. Masmoudi, "Scalable extension of the high efficiency video coding SHEVC performance study," *Computer Networks and Information Security (WSCNIS), 2015 World Symposium on*, Hammamet, 2015, pp. 1-4.
- [26] "High efficiency video coding HEVC test model 16.0 (HM.16.0)" [https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/HM-16.0/](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.0/)
- [27] Bjontegaard, G.: Calculation of average PSNR differences between RD-curves Doc. VCEG-M33, Austin (2001)
- [28] I. Wali, A. Kessentini, M. Ali Ben Ayed and N. Masmoudi, "Statistical analysis of SHVC encoded video, " *Second International Image Processing, Application and Systems Conference, (IPAS'16), 2016-World Conference on*, Hammamet, 2015, pp. 1-4.
- [29] SZE, V., BUDAGAVI, M., & SULLIVAN, G. J. (2014). High efficiency video coding (HEVC): algorithms and architectures. <http://public.eblib.com/choice/publicfullrecord.aspx?p=1802624>