# Complexity Evaluation in Scalable Video Coding

**Haris Al Qodri Maarif** G0725767@student.iium.edu.my
*Faculty of Engineering*
*Department of Electrical and Computer Engineering*
*International Islamic University Malaysia*
*Kuala Lumpur, 53100, Malaysia*

**Teddy Surya Gunawan** tsgunawan@iium.edu.my
*Faculty of Engineering*
*Department of Electrical and Computer Engineering*
*International Islamic University Malaysia*
*Kuala Lumpur, 53100, Malaysia*

**Akhmad Unggul Priantoro** unggul@iium.edu.my
*Faculty of Engineering*
*Department of Electrical and Computer Engineering*
*International Islamic University Malaysia*
*Kuala Lumpur, 53100, Malaysia*

### Abstract

The scalable video coding is the extension of H.264/AVC. The features in scalable video coding, are the standard features in H.264/AVC and some features which is supporting the scalability of the encoder. Those features add more complexity in SVC encoder. In this paper, complexity evaluation of scalable video coding has been performed. Different scalable configurations were evaluated in which the encoding time and the encoded video quality have been measured. Various scalable configurations with various GOPs, frame rates, QP value, have been implemented and evaluated, which shows the scalability of video coding for various conditions. Based on these results, a low complexity algorithm has been proposed. Results show that the proposed algorithm maintained the image quality (around 0.1 dB differences) while reducing the encoding time (around 30%).

**Keywords:** Scalable Video Coding, JSVM Reference Software, Complexity, Encoding Time, PSNR

## 1. INTRODUCTION

Scalable video coding (SVC) is classified as layered video codec [1] which is the extension of H.264/AVC standard. The extension of H.264/AVC standard in a way that a wide range of spatiotemporal and quality scalability is achieved [11]. SVC-based layered video coding is suitable for different use-cases and different bitstream e.g., supporting heterogeneous devices with a single, scalable bit stream. Such a stream allows for delivering a decode-able and presentable quality of the video depending on the device’s capabilities.

In terms of spatiotemporal and quality, scalability of SVC is referred as a functionality that allows the removal of parts of the bit-stream while achieving a reasonable coding efficiency of the decoded video at reduced temporal, Signal to Noise Ratio (SNR), or spatial resolution [3]. The scalability can be achieved in terms of temporal scalability, spatial scalability and quality scalability. Those three different types of scalability can be combined in order that the single scalable bit stream can support multitude of representations with different spatio–temporal resolutions and bit rates. The efficient scalable video coding provides benefits in many applications [4-6].

## 2. BASIC OF H.264/AVC

SVC was standardized as an extension of H.264/AVC [1]. It reuses some functions that have already been provided at H.264/AVC. Conceptually, the design of SVC covers a *Video Coding Layer* (VCL) and a *Network Abstraction Layer* (NAL), same as H.264/AVC was designed. VCL is representing the code of the source content (input video), the NAL is forming the VCL data in simple form and effective so that the VCL data can be utilized by many systems.

### 2.1. Network Abstraction Layer (NAL)

Data of the encoded video are gathered and organized into Network Abstraction Layer Unit (NALU). NALUs are the packets of data which are containing the integer number of bytes that represent the encoded video. The NALU starts with a one-byte header, which signals the type of containing data, and followed by payload data which represents the encoded video data. A set of consecutive NALU with specific properties is specified as an access unit. One decoded picture is resulted by decoding of an access unit results. A set of consecutive access units with certain properties is referred to as a coded video sequence. A coded video sequence represents an independently decodable part of a NAL unit bit stream. It always starts with an instantaneous decoding refresh (IDR) access unit, which signals that the IDR access unit and all following access units can be decoded without decoding any previous pictures of the bit stream.

For providing quality enhancement layer NALUs that can be truncated at any arbitrary point, the coding order of transform coefficient levels has been modified in a way that the transform coefficient blocks are scanned in several paths and in each path only a few coding symbols for a transform coefficient block are coded.

NALU are classified into VCL NALU and non VCL NALU. VCL NALU is the units which contain encoded slice data partitions, and non-NCL NALU is the units which contain the additional information of the encoded video. The non-VCL NALU provides additional information which can assist the decoding process in the encoder side and also some related process like bit stream manipulation or display. They are parameter sets, which are containing the infrequently changing information for a video sequence, and Supplemental Enhancement Information (SEI).

### 2.2. Video Coding Layer

The Video Coding Layer (VCL) of H.264/AVC is developed based on block-based hybrid video coding approach which is similar to the basic design of the previous video coding standards such as H.261, MPEG-1 Video, H.262 MPEG-2 Video, H.263, or MPEG-4 Visual. In the development of H.264/AVC, the new features are enabled in order to achieve the better performance in compression efficiency relative to any prior video coding standard [7].

In the H.264/AVC, the video frames are partitioned into smaller coding units which is called as macroblocks and slices. [8]. The video frame is partitioned into macroblocks which covers 16x16 luma samples and 8x8 samples of each of the two chroma components. The samples of a macroblock are predicted in terms of spatial or temporal, and the predicted residual signal is represented by using transform coding.

The macroblock are partitioned into the slices in which each of the slices can be parsed independently. The supported basic slices for the H.264/AVC are I-slice, P-slice, and B-slice [8]. I-slice is *intra-picture* predictive coding using spatial prediction from neighboring regions, P-slice is intra-picture predictive coding and inter-picture *predictive* coding with one prediction signal for each predicted region, and B-slice is intra-picture predictive coding, inter-picture predictive coding, and inter-picture *bipredictive* coding with two prediction signals that are combined with a weighted average to form the region prediction.

For I-slices, several directional spatial intra-prediction modes are provided by H.264/AVC. The prediction signal is generated by using neighboring samples of blocks that precede the block to be predicted in coding order. In the luma component, the intra-prediction is either applied to 4x4, 8x8, or 16x16 blocks, whereas for the chroma components, it is always applied on a macroblock basis [8].

In P-slices and B-slices, variable block size motion-compensated prediction with multiple reference pictures [27] is permitted. The macroblock type signals the partitioning of a macroblock into blocks of 16x16, 16x8, 8x16, or 8x8 luma samples. The macroblock also specifies the partition into some submacroblocks. For example, a macroblock type specifies partitioning into four 8x8 blocks, then each of the macroblock can be more partitioned into submacroblocks. The submacroblocks type can be either 8x4, 4x8, or 4x4 blocks.

For P-slices, transmission of one motion vector is applied for each block and the used reference picture can be independently chosen for each 16x16, 16x8, or 8x16 macroblock partition or 8x8 submacroblock. The choosing of macroblock partition is signaled via a reference index parameter, which is an index into a list of reference pictures that is replicated at the decoder [10].

For B-slices, biprediction method is applied by utilizing two distinct reference picture lists, *list 0 and list 1,* and for each 16x16, 16x8, or 8x16 macroblock partition or 8x8 submacroblock. Prediction of list 0 and list 1 are referring to unidirectional prediction by using reference picture of list 0 or list 1, respectively. The bipredictive prediction mode is applied by calculating a weighted sum of a list 0 and list 1 prediction signal. In addition, special modes as so-called *direct modes* in B-slices and *skip modes* in P- and B-slices are provided, in which such data as motion vectors and reference indexes are derived from previously transmitted information [8].

### 2.3. Supported Entropy Coding

Supported method for entropy coding in H.264/AVC are Context-based Adaptive Variable Length Coding (CAVLC) and Context-based Adaptive Binary Arithmetic Coding (CABAC) [10]. Both methods are using context-based adaptivity to improve performance relative to prior standards. CAVLC uses variable-length codes by restricted restricted to the coding of transform coefficient levels due to the adaptivity and CABAC uses arithmetic coding and some sophisticated mechanism for employing statistical dependencies.

## 3. SCALABLE EXTENSION OF H.264/AVC

The most important issue of Scalable Extension of H.264/AVC are coding efficiency and complexity, and all other parts are common types in the H.264/AVC. Since SVC was developed as an extension of H.264/AVC with all of its well-designed core coding tools being inherited, one of the design principles of SVC was that new tools should only be added if it is necessary to efficiently support the required types of scalability.

### 3.1. Temporal Scalability

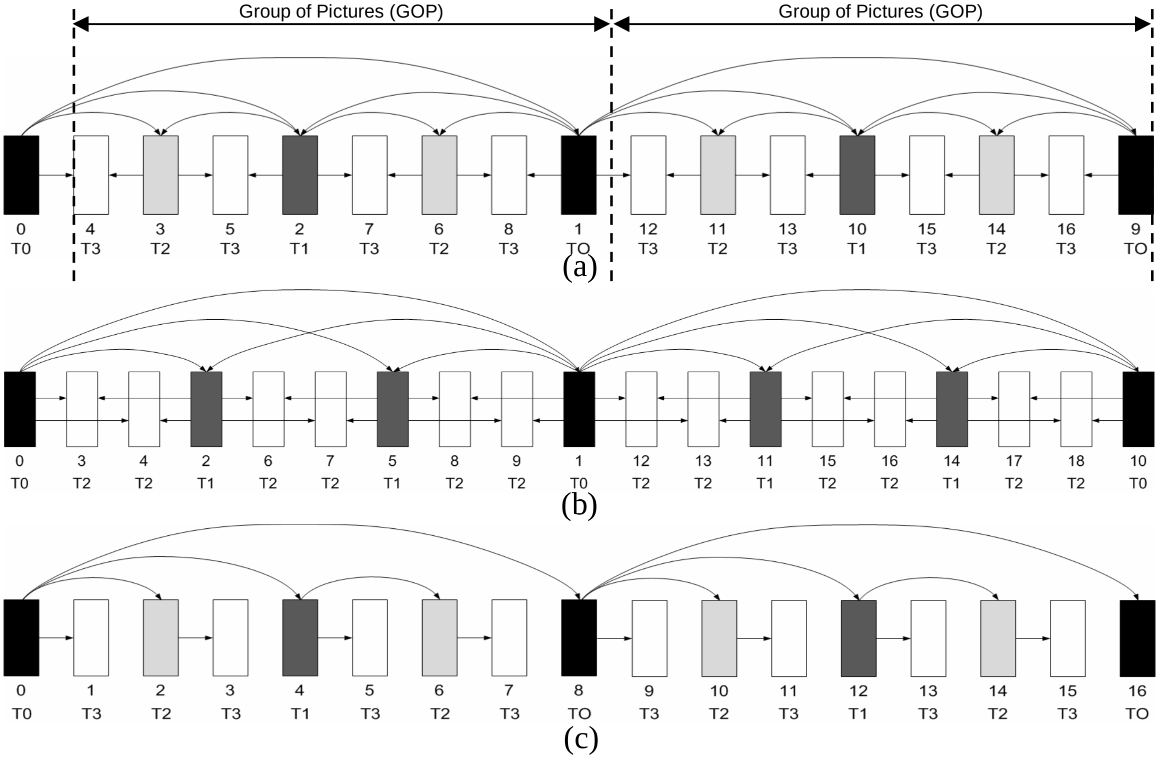
Information in bitstream provides temporal scalability by partitioning set of corresponding access units into a temporal base layer and one or more temporal enhancement layers. As a description for the temporal layer, the temporal layer is identified by a temporal layer identifier which is starting from 0 to *n* (number of enhancement layer). Value 0 is representing the base layer and value 1 to n which increases by 1 from one layer to next layer is representing the enhancement layer. For each natural number, the bit stream which is gained by removing all access units of all temporal layers with a temporal layer identifier is greater than forms another valid bit stream for the given decoder.

Enabling the temporal scalability in hybrid video codec can be applied by restricting motion-compensated prediction to reference pictures with a temporal layer identifier that is less than or equal to the temporal layer identifier of the predicted picture. The previous video coding standards such as MPEG-1 [11], H.262 MPEG-2 Video [3], H.263 [4], and MPEG-4 Visual [5] are also supporting temporal scalability. Specifically, in H.264/AVC the flexibility for temporal scalability was increased [6] because of its reference picture memory control.

### *3.1.1 Hierarchical Prediction Structures*

The concept of hierarchical prediction structures for enabling the temporal scalability is achieved by combining multiple reference pictures. It means that the construction of the reference picture lists can be done by using more than one reference picture, as the concept of H.264/AVC, and the pictures with the same temporal level as the picture to be predicted can be included to the reference picture lists. The prediction structure for base layer and enhancement layers are applied differently for each layer. For the base layer, the prediction is only based on the previous picture on the particular layer, while for the enhancement layer, the prediction is based on the two surrounding pictures of a lower temporal layer. A picture of the temporal base layer and all temporal refinement pictures between the base layer picture and the previous base layer picture build a group of pictures (GOP). The hierarchical prediction structures for enabling temporal scalability can be realized with dyadic and non-dyadic case.

The hierarchical prediction structures with dyadic temporal enhancement for enabling the temporal scalability temporal enhancement layer are based on the concept of hierarchical B-pictures [22, 13]. Fig. 1(a) illustrates the case of dyadic temporal enhancement layer. As described in the figure, the encoding process of the enhancement layer is coded as B-pictures. In this case the reference picture lists 0 is restricted for the temporally preceding picture and lists 1 is restricted for the succeeding picture, with a temporal layer identifier less than the temporal layer identifier of the predicted picture. Each set of temporal layers $\{T_o,\ldots,T_k\}$ can be decoded independently of all layers with a temporal layer identifier *T* > *k*.



(a)

(b)

(c)

**FIGURE 1:** Hierarchical Prediction Structures for Enabling Temporal Scalability

Prediction structures are not only restricted to the dyadic case but also to the non-dyadic case. For example, Fig. 1(b) shows a nondyadic hierarchical prediction structure in the enhancement layer. In this example, the hierarchical prediction structure provides 2 independently decodable subsequences with 1/9th and 1/3rd of the full frame rate.

The coding delay or structural delay between encoding and decoding of hierarchical prediction structures can be adjusted by deactivating the motion compensated prediction. For example, Fig. 1(c) describes a hierarchical prediction structure with controlled coding delay, in which the motion-compensated prediction is not employed the encoding process.

#### *3.1.2. Coding Efficiency of Hierarchical Prediction Structures*

The coding efficiency of hierarchical prediction structures is based on how the value of the quantization parameters of the encoder is chosen for different layer of scalable extension of H.264/AVC. Theoretically, all pictures in the temporal base layer are going to be used as references pictures for temporal enhancement layer. Therefore, the temporal base layer should be encoded with the highest fidelity. The value of quantization parameter for each subsequent hierarchy temporal later should be in the larger value as the quality of the enhancement layer is only influencing fewer pictures in the next subsequent hierarchy temporal enhancement layer.

In order to obtain high quality encoded video, the quantization parameter value can be calculated by computationally expensive rate-distortion analysis. This process adds additional complexity in the encoder, hence increasing the computational time. To overcome a complex operation problem, some strategies can be employed to reduce the complexity of the operations. For example, as mentioned in [21], one strategy is chosen based on quantization parameter to overcome this condition. In more details, the strategy is based on quantization parameter value of temporal base layer $QP_o$ and the quantization parameter value of enhancement layer $QP_t$ is defined by $QP_t = QP_o + 3 + T$. However, the strategy is giving fluctuation result in relatively large peak signal-to-noise ratio (PSNR) inside a group of pictures (GOP). Subjectively, the reconstructed video appears to be temporally smooth without annoying temporal “pumping” artifacts.

The coding efficiency of hierarchical prediction structure can be further enhanced by changing the size of Group of Pictures (GOP) and the encoding/decoding delay (low delay and high delay). The quality of encoded video from the encoder is positioned at acceptable level video quality. The trends are valid for any video sequences (e.g. IPPP and IBBP) with different frame rate and video resolution [10].

As a conclusion, providing temporal scalability in encoding process does not provide any negative effects on coding efficiency. Some small losses in coding efficiency may be noticed when low delay application is required. For the high delay encoding, some effects can be tolerated and the usage of hierarchical prediction structures are not only provide temporal scalability, but also significantly improves coding efficiency.

### 3.2. Spatial Scalability

The conventional approach multilayer coding is used in SVC for supporting spatial scalable coding. The multilayer coding is the coding method which is used by previous video coding standard, such as approach of multilayer coding, which is also used in H.262 MPEG-2 Video, H.263, and MPEG-4 Visual. Each layer in multilayer coding is corresponding to a supported spatial resolution and it is referred as spatial layer or *dependency identifier D*. For base layer D is 0 and for the next layer, D is started from 1 increase by 1 for next spatial layer.

The pictures in different spatial layer are encoded by its layer prediction information and motion parameters, or simply called as single layer coding. The activation of inter-layer prediction, which is utilizing the information from the lower layer, is done as a mechanism in order to improve rate-distortion efficiency of the enhancement layer. It will ensure that the complexity operation of motion-compensated prediction and deblocking are inly applicable in the target layer (output picture).

In order to restrict the memory requirements and decoder complexity, SVC specifies that the same coding order is used for all supported spatial layers. The representations with different spatial resolutions for a given time instant form an access unit and have to be transmitted successively in increasing order of their corresponding spatial layer identifiers.

### *3.2.1. Inter-Layer Prediction*

The inter-layer prediction is the mechanism in spatial scalability which utilize the information from the lower layer signal in order to increase rate-distortion efficiency of enhancement layer. The utilizing of lower layer information has been done previously by prior video coding standards, such as H.262 MPEG-2 Video, H.263, and MPEG-4 Visual. The inter-layer prediction method is assigning the reconstructed samples of the lower layer signal. The prediction signal can be produced by three methods. Those methods are motion-compensated prediction inside the enhancement layer, upsampling the reconstructed lower layer signal, and averaging such an upsampled signal with a temporal prediction signal.

Inter-layer prediction is not always using information from reconstructed lower layer samples which represents the complete lower layer information, but also the information is taken from other lower layer information, such as temporal prediction signal. The inter-layer predictor has to compete with the temporal predictor, especially for some special cases such as slow motion video and high spatial detail. The information from temporal predictor is giving the better predicted data than the data from lower layer. For giving better result and higher efficiency in spatial scalability, two additional inter-layer prediction concepts [15] have been added in SVC: *prediction of macroblock modes and associated motion parameters* and *prediction of the residual signal*.

In order to gain the better coding efficiency and high quality encoded video, the new mechanism is implemented to reach the intended goal. SVC is applying switchable mechanism which allows switching between intra and inter-layer motion prediction by receiving local signal characteristics. Inter-layer prediction can only work on a spatial layer identifier *D* less than the spatial layer identifier of the layer to be predicted. The layer employing inter-layer prediction is referred as *reference layer*, and it is signaled in the slice header of the enhancement layer slices. Since the SVC inter-layer prediction concepts include techniques for motion as well as residual prediction, an encoder should align the temporal prediction structures of all spatial layers.

The interlayer motion prediction is a mechanism which is utilizing the lower layer information to predict the motion of the next picture in a video sequence. In order to activate inter-layer motion prediction by employing motion data from lower layer in spatial scalability, the new macroblock type is introduced in SVC and it is referred as reference layer skip mode. When the reference layer macroblock is inter-coded, the enhancement layer macroblock is also inter-coded. In that case, the partitioning data of the enhancement layer macroblock together with the associated reference indexes and motion vectors are derived from the corresponding data of the co-located 8x8 block in the reference layer by so-called *inter-layer motion prediction*.

The reference layer skip mode specifies prediction data from reference layer and encoded residual signal. The macroblock partitioning is determined by upsampling and re-aligning the partitioning of reference layer region that is covering the same area on the predicted picture. As an example, the following is the example of dyadic spatial scalability without cropping, each enhancement layer macroblock corresponds to an 8x8 submacroblock in the reference layer and the enhancement layer macroblock partitioning is obtained by scaling the partitioning of 8x8 base layer block by a factor of 2 in both vertical and horizontal.

The Inter-layer residual prediction is able to be utilized in any inter-coded macroblocks. It is signaled by a flag in which is added newly SVC macroblock which is transmitted on a macroblock basis. The transmitted flag is 1 or addressed as true, the residual signal of reference layer is upsampled by using a bilinear filter which is applying block basis transform in order to restrict filtering in across transform block boundaries. The upsampled signal is used as information for predicting the residual signal of the current macroblock.

When an enhancement layer macroblock is coded with This mechanism is the prediction when the macroblock in enhancement layer is encoded with *base mode flag* equal to 1 or by using reference layer skip mode. Generated prediction signal is gained by upsampling the reconstructed intra signal of the reference layer.

To prevent complete decoding of the lower layers which can decrease coding efficiency, the inter-intra prediction is restricted to macroblocksk in enhancement layer. The constrained intraprediction has to be applied in the reference layer which does not have inter-predicted samples as the data for intra prediction. By this condition, the supported layer can be decoded by a single loop decoding [16, 17] which is avoiding the inter-coded macroblocks in the reference layer.

### *3.2.2. Generalized Spatial Scalability*

Spatial scalability standard for scalable video coding is similar to the previous version of video coding standard, such as H.262 MPEG-2 Video and MPEG-4 Visual. The spatial scalability is supported in spatial scalability with arbitrary resolution ratios (decreasing from one layer to another layer). The resolution between base layer and enhancement layer can be decreased as long as the ratio of the picture resolution is not changed. It means that neither the horizontal nor the vertical resolution can decrease from one layer to another layer.

The design of spatial scalability in SVC is supporting the possibility for the enhancement layer to represent only the selected area in the base layer or reference layer. Another design is the possibility for the enhancement layer to have additional content beyond the reference layer. This possibility is called as a cropping picture. It can be combined and modified on a picture-by-picture basis.

Furthermore, another SVC design for spatial scalable coding is also including the interlaced sources. All the basic inter-layer prediction concepts are maintained for spatial scalable video coding with arbitrary resolution ratios and cropping as well as for the spatial scalable coding of interlaced sources. But other extensions such as the derivation process for motion parameters as well as the design of appropriate upsampling filters for residual and intra-blocks needed to be generalized.

### *3.2.3. Complexity Considerations*

Inter-layer intra prediction has a possibility to be applied only at the enhancement layer in the encoder side. The limitations are able to increase the coding efficiency [15]. Furthermore, the constraining of inter-layer prediction in the enhancement layer can significantly decrease the decoder complexity [16, 18]. This condition is called as *constrained inter-layer prediction* which has an intention to avoid the computationally complex and memory access intensive operations of motion compensation and deblocking for inter-coded macroblocks in the reference layer.

By these conditions, the enhancement layer can be decoded with *single motion compensation loop*. Referring to the complexity in the decoder side, the SVC has the smaller complexity compared to single-layer coding which all require multiple motion compensation loops at the decoder side. Additionally, it should be mentioned that each quality or spatial enhancement layer NAL unit can be parsed independently of the lower layer NAL units, which provides further opportunities for reducing the complexity of decoder implementations [19].

### *3.2.4. Coding Efficiency*

The evaluation utilizes fixed bitrate for base layer and varied bitrate for enhancement layer as well as the GOP size of 16 pictures and IPPP [10]. Also the unconstrained inter-layer prediction and decoding with multiple compensation loops was applied as an additional simulation. The first access unit was intra-coded and CABAC was used as entropy coding method.

The simulation shows that the effectiveness of a tool or combination of tools strongly depends on the sequence characteristics and the prediction structures [10]. The overall performance of scalable video coding compared to single-layer coding reduces when moving from a GOP size of 16 pictures to IPPP coding. To increase the coding efficiency, multiple loop decoding can further be applied with some significant increase in decoder complexity. The rate-distortion performance was not giving some enhancement for multiloop decoding which is using only inter-layer intra-prediction. However, it should be noted that the hierarchical prediction structures which not only improve the overall coding efficiency but also the effectiveness of the inter-layer prediction mechanisms, are not supported in these prior video coding standards.

### 3.3. Quality Scalability

Quality scalability is considered as a special case of spatial scalability. The case for quality scalability is lies on the identical picture sizes for base and enhancement layer in scalable video coding. The quality scalability can be defined into two quality scalability, *course grain scalability* (CGS) and *medium grain scalability* (MGS). The quality is called as CGS when the identical picture size for base and enhancement layer are supported by spatial scalable coding, and the variation of CGS approach, which allows a switching between different layers in any access units, is referred as MGS.

For CGS, the interlayer prediction as for spatial scalability is applied. Since base and enhancement layers in CGS are identical, the upsampling operation and the inter-layer deblocking are not involved in encoding process. The inter-layer intra and residual prediction are applied in transform domain. The refinement of texture information is gained by requantizing the residual texture signal in the enhancement layer with smaller quantization step size. Furthermore, the multilayer concept for CGS only allows a few selected bit rates to be supported in a scalable bit stream, since the number of supported rate points is identical to the number of layers. Finally, the multilayer concept for quality scalable coding becomes less efficient, when the relative rate difference between successive CGS layers gets smaller [10].

To increase the flexibility of bit stream adaptation and error robustness as well as improving the coding efficiency for bit streams that have to provide a variety of bit rates the MGS concept is introduced. MGS is the variation of the CGS approach which allows switching between all layers (base layer and enhancement layer). In the MGS concept, any enhancement layer NAL unit can be excluded from bit stream, and thus packet-based quality scalable coding is provided. SVC standard has the possibility to distribute the enhancement layer transform coefficients into several slices. The transform coefficients are signaled in the slice headers, and the slice data only include transform coefficient levels for scan indexes inside the signaled range. Furthermore, the information for a quality refinement picture can be distributed over several NAL units corresponding to different quality refinement layers.

## 4. JSVM REFERENCE SOFTWARE

The JSVM (Joint Scalable Video Model) reference software is the reference software for H.264/SVC or Scalable Video Coding standard. The software is used as the tool to evaluate the performance of scalable video coding standard and implement the proposed algorithm for scalable video coding. It is supporting the single layer coding and multiple layer coding.

The reference software is the joining project between Joint Video Team (JVT) and ITU-Video Coding Experts Group (VCEG) which is an on going standard [19]. Since the scalable video coding standard is still under development, the JSVM Reference Sofware is also under development and changes frequently. The JSVM Reference Sofware is an open source code and written in C++ code. Since the JSVM is the reference software, the source code of the software is provided and can be accessed easily from the CVS server. The CVS server was setting up by Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen.

To build the JSVM software by using Microsoft Visual Studio, it needs file with .sln extension. The .sln extension is the workspace file which is collecting all the information of the software. In order to build the software, the .sln file should match with the C++ compiler version. The .sln files are located in folder *JSVM/H264Extension/build/windows.* The folder is containing workspace file *H264AVCVideoEncDec.sln*, *H264AVCVideoEncDec_vc8.sln* and *H264AVCVideoEncDec_vc9.sln,* which is valid to Microsoft Visual Studio .NET 2003 (VC7), Microsoft Visual Studio .NET 2005/2006 (VC8), and Microsoft Visual Studio .NET 2007/2008 (VC9), respectively*.* In order to build the software, the .sln files should be chosen and opened in appropriate version of Microsoft Visual Studio .NET.

To build the JSVM reference software by Linux with gcc compiler needs the makefiles which act like as an workspace file in windows. In order to build the software, the gcc compiler should match with the version of software. In our project, the gcc compiler in ubuntu 8.04 was used to compile JSVM reference software. The makefile is located in the folder *JSVM/H264Extension/build/linux* and the corresponding sub-folders.

After building process is finish by using c compiler in windows, the binaries and libraries files are located in the folders *bin* and *lib*, respectively, For the 64 bits software, the binaries and libraries files are located in the folders *bin64* and *lib64*. In each corresponding folders, there are two different versions for each binary and library, which is with and without “d” in the end of the file name. The files with end of “d” represent binaries or libraries that have been built in debug mode, while the files without end of “d” dot represent binaries or libraries that have been built in release mode.

### 4.1. PARAMETER SETTINGS

The JSVM reference software requires some configurations to perform specific encoding process. To set up the parameter in JSVM reference software, the configuration files is required in both encoding and decoding process. The configuration file is using .cfg extension. The .cfg extension files will be read by the JSVM reference software when the software is running. The configuration files are stored in folder *JSVM/bin.* By default, the JSVM reference software will read encoder.cfg (for encoding process) and decoder.cfg (for decoding process).

The parameter in configuration files should be defined properly in order to achieve the simulation objectives. The parameters in configuration files are both dependent and independent each other. There are two types of configuration files for encoding process, main configuration files and layer configuration files. The main configuration file is the parameter setting for the whole scalable video coding system and the layer configuration file is the parameter setting for particular layer. The number of layer configuration files is depending on the parameter setting in main configuration file.

In configuration files, all the setting up parameters should be configured properly in order to meet the objective in encoder. In both configuration files, main and layer configuration, some specific parameter must be defined and configured. The parameter in main and layer configuration files are different, the specific parameters and its value are explained in [20], JSVM reference manual. Generally, the main configuration file configures the parameters for input and output file, number of frame rate, encoder mode, number of enhancement layer, and GOP (Group of Pictures) size. On the other hand, the layer configuration file configures the video input for the respective layer, video size, and the coding process in the respective layer.

The executable file in JSVM reference software is part of the reference software used as a tool to run the encoding process and evaluate the output from the encoder. Mostly used tools are executable files for encoding, decoding, and PSNR calculation. *H264AVCEncoderLibTestStatic* is the executable files for encoding and generating the scalable video coding (SVC) bit stream.

*H264AVCEncoderLibTestStaticd* is the executable files for decoding and reconstructing the encoded video sequences. *PSNRStatic* is a tool to measure the PSNR between the encoded and decoded video sequences.

## 5. COMPLEXITY ANALYSIS IN SCALABLE VIDEO CODING

Complexity analysis is SVC encoder can be derived by two analysis approaches, i.e. time complexity and storage capacity. Time complexity is calculated by numbers of operation required to encode video by using a specific algorithm, so that some cleaver algorithm which is using some specific mode or features to encode has faster encoding time than full mode algorithm. On the other hand, space complexity is analyzed by approximated buffer size space approach while implementing the algorithm. The time complexity of SVC encoder will be studied and analyzed in this section.

The time complexity analysis is analyzed by two basic steps. In the first step, the number of cycles needed to execute a particular sub function in an algorithm is calculated. Then, the calculated cycles in a sub function is multiplied by the frequency in which sub function was used. Finally, time complexity is combination of all sub functions executed in the algorithm [23].

As discussed in section 3, the scalable video coding has more features that the H.264 single layer. Those extra features bring more complexity into SVC encoder than AVC encoder. The complexity in scalable video coding is because of the scalability itself in SVC encoder. Temporal, spatial and quality scalability are the component of the complexity in SVC encoder.

**TABLE 1.** Scalable H.264/AVC test streams with 3 different GOP sizes

| GOP | Freq (Hz) | QP High | | | QP Low | | |
|---|---|---|---|---|---|---|---|
| | | Bitrate (kbps) | Min BitRate (kbps) | Avg PSNR (db) | Bitrate (kbps) | Min BitRate (kbps) | Avg PSNR (db) |
| 4 | 7.5 | 96.71 | 89.6 | 38.1 | 390.23 | 342.89 | 42.8 |
| | 15 | 121.34 | 111.9 | 38.0 | 470.93 | 418.46 | 42.5 |
| | 30 | 144.86 | 131.35 | 37.9 | 536.35 | 479.81 | 42.3 |
| 8 | 3.75 | 77.9 | 74.69 | 38.4 | 268.82 | 250.4 | 43.2 |
| | 7.5 | 95.8 | 91.58 | 38.2 | 331.01 | 311.42 | 42.8 |
| | 15 | 120.93 | 114.53 | 38.2 | 410.09 | 387.16 | 42.5 |
| | 30 | 144.26 | 133.78 | 38.1 | 476.01 | 448.97 | 42.3 |
| 16 | 1.88 | 57.5 | 55.34 | 39.3 | 194.47 | 183.18 | 44.5 |
| | 3.75 | 76.73 | 73.96 | 38.7 | 259.09 | 246.67 | 43.5 |
| | 7.5 | 95.92 | 92.09 | 38.5 | 324.24 | 310.56 | 43.0 |
| | 15 | 121.38 | 115.44 | 38.4 | 405.82 | 389.33 | 42.7 |
| | 30 | 144.85 | 134.84 | 38.3 | 472.36 | 451.7 | 42.5 |

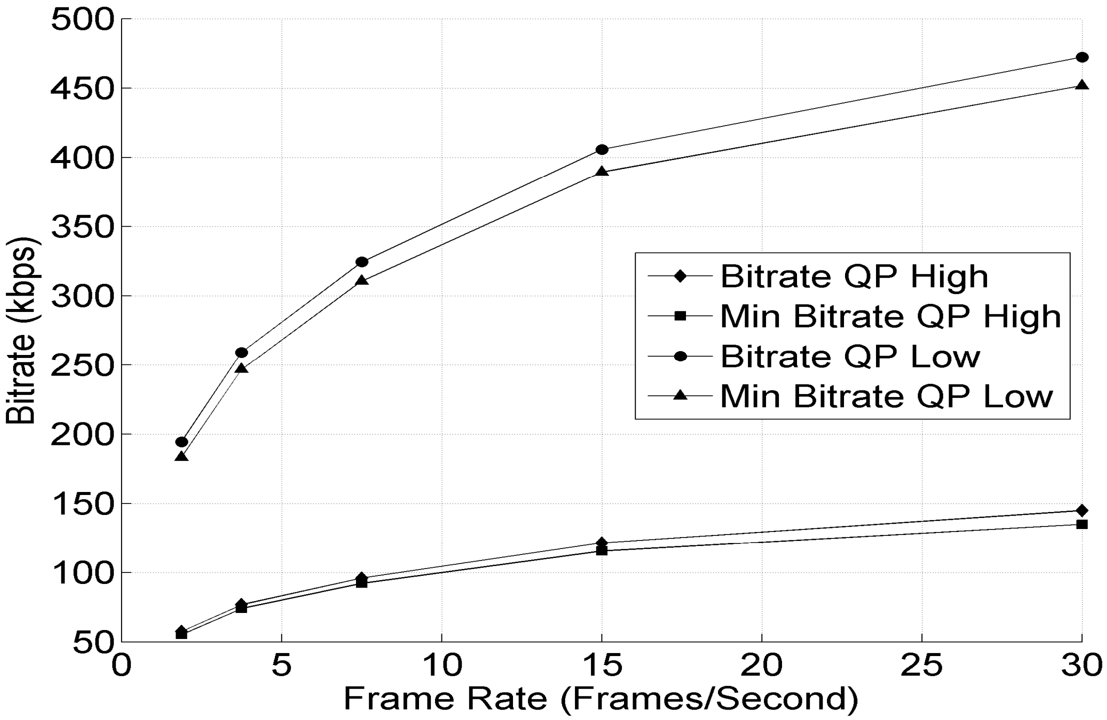


**FIGURE 2:** Bitrate of each frequency for GOP16

In this section, the simulation performance of encoding process and bitstrem analysis using JSVM reference software will be disscussed. Two different simulations are going to be part of analysis. The first simulation shows the temporal, spatial, and quality scalablity in term of streamed over wireless networks, which is showing the capability of SVC encoder to provide various spatial and temporal resolution for transmitting the stream video to different types of receiver and bandwidth as well as network condotion.. The second simulation provides a brief analysis about encoding time comparison between high complexy encoder and low complexity encoder. The improvement showed in the second simulation is the improvement in terms of encoding time that has been reduced and it is showing the maintaned encoded video quality.

For the first simulation, the JSVM reference software version 9.15 was used. In the encoding process, Foreman video test sequence was used for the evaluations. One base layer and two enhancement layers are employed for encoding process as well as two quantization parameter (QP) values. For video input, YUV Foreman sequence was used in CIF format with frame rate of 30 frames per second. The GOP size of 4, 8, and 16 were used which also automatically define the number of B, I, and P frames. The quantization parameters (QP) used were 28 and 38 as an optimal value [1] for high (QP low) and low (QP high) encoded video quality, respectively.

Temporal scalability provides the encoder capability to encode the video into different video frequencies. These are having a tendecy of the streamed video to be more adapt to network conditions. As shown in Table 1, the output from the scalable video coding has different frame rate which can be selected based on current network condition. From the table, it can be seen that the smaller the value of GOPs, the smaller bitrates. This is because of the number of encoded frames become less and the bits used to encode the input video also become less.

The varieties of frequency are depicted in Table 1. The variations provide the information about temporal scalability of SVC encoder and which bitstream can be streamed over the wireless network. Once network is in a very good condition the encoded video with the highest frequency as well as best video quality will be transmitted through network, and vice versa.

Figure 2 shows the variety of bitrate of each frequency for GOP 16. The variety of frequencies show the support of encoding process in different temporal-resolutions for different network conditions. The frame rate (frequency) supported is different for different GOPs value. The higher the GOP value, the more diverge the temporal resolution. Moreover, it also shows the scalability of the encoded video for each GOPs value.

Table 1 and Figure 2 show the large variety of bitrate of the picture in the GOP16. In the GOP16, the temporal resolution can be in five different frequencies or frame rates. It is ranging from 30 Hz until 1.875 Hz, so GOP16 has more temporal scalability compare to other GOPs. The encoded video in 30 Hz frequency will be transmitted when the network is in the best condition down to 1.875 Hz when the network is in the worst condition.

**TABLE 2:** Scalable Extension H.264/AVC for Low Complexity and High Complexity

| Video | BDBR (%) | BDPSNR (dB) | Time Saving (%) |
|---|---|---|---|
| News | -5.521 | 0.118 | 30.22 |
| Foreman | -4.238 | 0.1 | 29.65 |

In second simulation, the analysis of the complexity of SVC was presented in term of encoding time. The time comparison between encoder with the high complexity and low complexity are showed. The encoder with high complexity showed the longer encoding time than the encoder with low complexity. Not only the encoding time, but also the quality itself will be compared between high complexity and low complexity.

The JSVM reference software version 9.15 was also used to do the second simulation process. The streaming video with YUV format, i.e. Foreman and News video sequences, were employed as the input tested video sequence to observe the output from video encoder. The frame rate of video sequences was 30 frames per second. Two layers were used, i.e. base layer and enhancement layer in different spatial resolution. Video in QCIF (177x144) resolution was used as base layer, and CIF (356x288) resolution was used as enhancement layer. The GOP 16 was used and quantization parameters (QP) used was 38.

Two encoding schemes, i.e. high complexity and low complexity algorithm, were implemented and evaluated. Performance evaluation of the encoded video is based on subjective survey and objective evaluation. Subjective survey is based on the personal opinion and objective evaluation is based on the calculation of BDBR, BDPSNR, and Time Saving. BDBR is value of different bitrate, BDPSNR is the different value of PSNR and the Time Saving shows the computation time between the high complexity and low complexity [21].

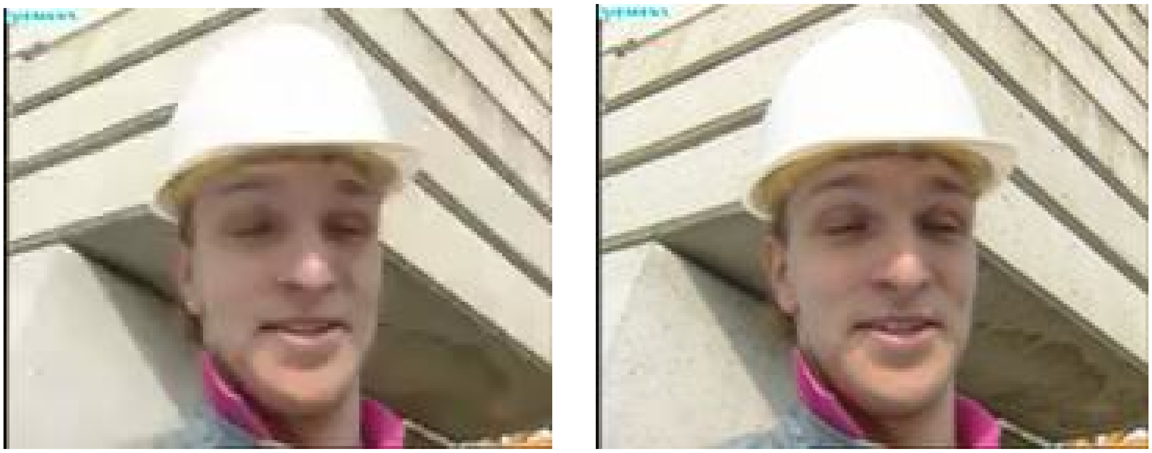


**FIGURE 3:** RECONSTRUCTED IMAGE

For objective evaluation, values for BDBR, BDPSNR, and Time Saving are shown in Table 2. The table directly shows comparison between high complexity and low complexity schemes. It can be seen that the low complexity scheme provides higher time saving for encoding time up to 30 % with the negligible different PSNR of 0.100 – 0.118 dB, and 4% – 5% bit rate decreases. The statistic represent the low complexity can be used for encoding process while maintain high video quality.

For subjective evaluation, as can be seen in Figure 3, high quality encoded video is still achieved while the encoding time is significantly faster. There are some reduced qualities in the low complexity scheme but it is negligible as shown in Table 2. As mentioned in [10], the quality encoded video by SVC will become poorer when the bandwidth is low.

The main improvement from our simulation is achieving the time saving for encoding process while the encoded video quality is maintained in an acceptable quality. From the two experiments, our results are in line with other research which is studying about reducing the complexity of SVC encoder. For example, Goh [15] state, the fast mode decision by using correlation of neighbor macroblock had reduced the complexity which was represented by the time saving achievement with the negligible loss. On the other hand, Nguyen [17], also implemented the low complexity encoding by downsampling the enhancement layer, as the result, the complexity also reduced and the time saving was achieved as well as the video quality was maintained.

## 6. CONCLUSION

The complexity analysis of SVC has been presented, the spatio-temporal scalability and encoding time have been evaluated. At the first experiment, the spatio-temporal scalability shows capability of SVC to provide various spatio-temporal resolution used for transmitting the encoded video. At the second experiment, the encoding time represenst the complexity of the system, more complexity, more time needed for encoding process. In the second experiment, the encoding time has been shown for evaluation purposes. The complexity analysis of SVC encoder is mainly analyzing the encoding time. The high complexity scheme performs longer encoding time, the low complexity scheme. The high complexity represent all features in SVC encoder was used for encoding, and the low complexity represents optimized use of the features in SVC encoder. On the other hand, the video quality in low complexity scheme has an acceptable quality and neglibigle PSNR difference between the reconstructed and original video (BDPSNR around 0.1dB), subjective and objective evaluation showed that reconstructed video quality was maintained and negligible reduced quality of reconstructed video was detected.

## 7. REFERENCES

[1] H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "Combined scalability support for the scalable extensions of H.264/AVC", in *International Conference on Multimedia and Expo*, 2005.

[2] G. Liebl, M. Wagner, J. Pandel, and W. Weng, "An RTP payload format for erasure-resilient transmission of progressive multimedia streams", *in IETF,* October, 2004.

[3] *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s—Part 2: Video*, ISO/IEC 11172-2 (MPEG-1 Video), ISO/IEC JTC 1, Mar. 1993.

[4] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ITU-T and ISO/IEC JTC 1, Nov. 1994.

[5] *Video Coding for Low Bit Rate communication*, ITU-T Rec. H.263, ITU-T, Version 1 : Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

[6] *Coding of audio-visual objects—Part 2: Visual*, ISO/IEC 14492-2 (MPEG-4 Visual), ISO/IEC JTC 1, Version 1: Apr. 1999, Version 2: Feb. 2000, Version 3: May 2004.

[7] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Version 1: May 2003, Version 2: May

2004, Version 3: Mar. 2005, Version 4: Sept. 2005, Version 5 and Version 6: June 2006, Version 7: Apr. 2007, Version 8 (including SVC extension): Consented in July 2007.

[8] B.J. Kim, Z. Xiong, and W. A. Pearlman, “Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D SPIHT),” in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. l0, no. 8, pp. 1374-1387, December, 2000.

[9] H. Schwarz, D. Marpe, and T. Wiegand, “Further Results on Constrained Inter-layer Prediction”, Joint Video Team, Doc. JVT-O074, April, 2005.

[10] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, September, 2007.

[11] *Video Codec for Audiovisual Services at p x 64 kbit/s*, ITU-T Rec. H.261, ITU-T, Version 1: Nov. 1990, Version 2: Mar. 1993.

[12] H. Schwarz, T. Hinz, H. Kirchhoffer, D.Marpe, and T.Wiegand, “Technical Description of the HHI Proposal for SVC CE1”, ISO/IEC JTC 1/SC 29/WG 11, Doc. M11244, October, 2004.

[13] H. Schwarz, D. Marpe, and T.Wiegand, “Independent Parsing of Spatial and CGS Layers”, Joint Video Team, Doc. JVT-S069, Mar. 2006.

[14] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model (JSVM) 2.0 Reference Encoding Algorithm Description," ISO/IEC JTC1/SC29/WG1 1, Doc. N7084, Buzan, Korea, April. 2005.

[15] G. Goh, J. Kang, M. Cho, and K. Chung, “Fast Mode Decision for Scalable Video Coding Based on Neighboring Macroblock Analysis”, in *Proceedings of the 2009 ACM Symposium on Applied Computing*, pp. 1845-1846, 2009

[16] C-W. Chiou, C-M. Tsai, and C-W. Lin, “Fast Mode Decision Algorithm for Adaptive GOP Structure in the Scalable Extension of H.264/AVC”, in *IEEE International Symposium on Circuits and Systems,* pp. 3459-3462, May, 2007

[17] C. An, and T.Q. Nguyen, “Low Complexity Scalable Video Coding”, in *Proceedings of ACSSC,* 2006

[18] H. Schwarz, D. Marpe, and T.Wiegand, “Hierarchical B Pictures”, Joint Video Team, Doc. JVT-P014, July, 2005.

[19] JSVM Website

[20] JSVM 9.16 Overview

[21] T. Schierl, H. Schwarz, D. Marpe, and T. Wiegand, "Wireless Broadcasting using the scalable extension of H.264/AVC”, in *Proceedings of Int. Conference on Multimedia and Expo*, pp. 884-887, 2005.

[22] G. V. d. Auwera, P. T. David, and M. Reisslein, "Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG–4 advanced video coding standard and scalable video coding extension," in *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 698-718, September, 2008

[23] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, “H.264/AVC Baseline Profile Decoder Complexity Analysis”, *in IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 704-716, July, 2003.