SindhUniv. Res. Jour. (Sci. Ser.) Vol.49(1) 107-112 (2017)

SI NDH UNIVERSITYRESEARCH JOURNAL (SCIENCESERIES)

Scalable Video Coding: Performance Analysis and Trade-offs

N. MINALLAH[++], S. IKRAM, N. AZIZ, A. KHALIL

Department of Computer Systems Engineering, University of Engineering and Technology Peshawar, Pakistan

Received 03rd August 2016 and Revised 12th February 2017

Abstract-With the current advancement trend in the field of computing technologies, the research and development in the area of video coding is critical. It is due to this reason that various coding techniques have been developed and different video coding standards are available in the market. A compression efficient video encoding with in-built robustness towards transmission errors is the current focus of recent advancements in video encoding technology. Different diverse video coding standards were developed in the past each with considerable progress in the evolution of video coding technologies. Currently, H.264/SVC standard is one of the attractive choice with considerable compression efficiency and support for the scalable features. This paper provides the performance analysis of the different coding techniques and scalability features of H.264/SVC standard using JSVM software. Three different types of video sequences with high, low and medium level of variations in the video contents and different dynamisms are considered. Different performance metrics such as video decoding time, bit rate and PSNR are considered while employing different types of H.264/SVC scalabilities and techniques.

Keywords: Scalable Video Coding, Bit Rate, PSNR, H.264/SVC, PSNR, JSVM, MPEG-2, MPEG-4

1. INTRODUCTION

With the technology advancement, the users' demands compel the researchers to provide better video quality along with efficient network requirements (Unanue, *et al.*,2011) (Schierl, *et al.*, 2007). Frame Error rate, Peak signal to Noise ratio, Bit Error rate and resolution are the main factors that affect the network requirements. These factors of the received video determine the quality of video transmission. For better and efficient transmission, the received video requirements must be fulfilled by transmitted video (Sanz-Rodríguez, *et al.*, 2015) (Moon, *et al.*, 2005) (Hewage, 2009). Designing a video encoder that performs adaptive video streaming has been always a difficult task (Kessentini, *et al.*, 2011) (Ben, *et al.*, 2013), because different systems have different infrastructure and working mechanism. Each system has its own bit error rate, frame error rate, computational complexity, compression efficiency and video quality.

The complexity of encoder is increased by encoding a video for different devices with different settings. Encoder also becomes more complicated when bandwidth, network requirements and devices are highly specific.

A system designed to generate a video stream which must be compatible to decoding capacity and available bandwidth of the network along with these combinations of different settings. Such a system also requires multiple times encoding and decoding a video and results in an expensive technique. In real time, the main concept to achieve a high quality video is to use a high efficient codec so that a video is encoded once and in the same way a received video should also be decoded once. The devices having low performance capability will transmits a video stream with less bandwidth and low resolution over the network. Such transmission requires a small percentage of encoded video with low bandwidth without complex processing of the video over the network. At the receiver side, a small video stream is received that results in easier decoding to generate the transmitted stream with low resolution.

In this method, the encoder can easily encode a video with low or high channel bandwidth as required by the receiving display device. This adaptive phenomena of the encoder which involves encoding a video according to available network and display device lead to scalable video codec concept ("EE Times." [Online]).

The basic concept on which the scalable video codec is based is to update the conventional method by transmitting a video stream with multi stream video incorporated in it as different modules also called Layers (Huang, *et al.*, 2007). An efficient video transmission system is developed by using scalable video codec based on layered approach.

3. Scalable Video Codec

This section will provide a brief overview of the considered scalable video codec i.e. H.264/SVC and the scalability features that it incorporates (Ambalgi, 2011).

H.264/SVC (Scalable Video Codec) Standard

The H.264/SVC standard provides the robust and efficient video transmission over heterogeneous networks with various characteristics (Rieck and Rupp,

[++]Corresponding author's e-mail: NasruMinallahn.minallah@uetpeshawar.edu.pk

2008). H.264/SVC standard is much better than the antecedent video codec's with scalability features (Huang, *et al.*,2007)H.264/SVC standard is also the most famous and advanced scalable video codec as compared to previous standards due to its improved efficiency feature that produces scalable bit streams (Ohm,2007)(Richardson, (2011). This new version provides better encoder capability to encode video with higher resolutions and higher frame rate ("wiki." [Online]),(Magli,and Casas, 2014).

Scalability Schemes in H.264/SVC Standard

The H.264/SVC standard mainly works on three different types of scalability methods that are as follows:

1) Temporal Scalability:

In Temporal scalability, the video is encoded such that different frame rates are incorporated in the encoded stream as sub streams. The sub streams employ hierarchical prediction structure for inter-frame predictive coding (Rieckand Rupp, 2008). This is achieved by adopting three different types of frames structures. These three different types of structures are referred to as I-frame (Intra), P-frame (Predictive) and B-frame (Bi-Predictive). In Intra (I) frame, compression techniques are used only in the existing picture without any reference to any other picture. In P (Predictive) frames, for prediction technique a preceding reference picture is used. In Bi-Predictive (B) frames, prior and posterior pictures are used as referenced picture and inter picture and bi predictive coding is employed (Unanue, *et al.,* 2011) (Schierl, *et al.*, 2007) Sanz-Rodríguez, *et al.*, 2015) (Moon, *et al.*, 2005) (Hewage, 2009).

Additionally, the arrangement of the frames within any encoded video is devised by GOP (Group of Pictures) structure. While encoding the H.264/SVC video stream, GOP structure mainly works on the GOP size parameter that is present in the main configuration file of the encoder. In Temporal scalability, this GOP size is responsible for classifying the different frames i.e. I, P and B frames into various temporal layers (Hewage, 2009).

2) Spatial Scalability:

In this technique multiple layers are used in the bit stream, which mainly work on the concept of refining the spatial resolution (Rieck and Rupp, 2008).

In Spatial scalability, main encoded bit stream consists of many sub streams with different resolutions. The two scenarios in Spatial scalability are as follows (Schafer and Sikora, 1995):

1) Dyadic Case: In this case, the successive layers have a ratio of2:1.

2) Non-Dyadic Case: In this case, any random resolution can be selected. It is also known as Extended Spatial Scalability.

3) SNR/Quality Scalability:

Signal to Noise Ratio (SNR) scalability is also called Quality scalability. Quantization of the video is the main target in this type of scalability. Before encoding, the coefficients of transform are quantized and the level of quantization is achieved by manipulating the quantization parameter. Due to small quantization step size more levels of quantization are produced which result in better quality of the video. The three types of SNR scalabilities are as follows:

i) Coarse Gain Scalability (CGS): This technique reuses the concept of Spatial scalability. In this approach, the video is subdivided in various quality layers like Spatial scalability technique. Due to the same resolution of consecutive layers, this technique is taken as a special case of Spatial scalability. The motion estimation techniques are performed at each quality layer independently. The quality layer are allowed to switch between themselves only at specific points (i.e. at I frames).

ii) Medium Grain Scalability (MGS): In this scalability type, both quality layers i.e. base layer and enhancement layer can be referenced because this scalability has more flexible prediction module for the application of motion prediction loop. In this technique, the only drawback is drifting effect, in which the motion prediction loop is not consistent on the reference picture at the encoder and decoder sides. This deficiency is solved by taking continuous updates in the base layer, in case of utilizing the enhancement layer for motion estimation and is known as key picture concept. These update helps in synchronizing the motion prediction loop and thus drifting effect is minimized ("EE Times." [Online]).

iii) Fine Grain Scalability (FGS): In this scalability method, motion estimation and motion compensation are applied at the lowest quality layer of the reference picture because the base quality layer is always available at receiver end and helps in the motion prediction technique. In this technique, the bit rate can also be rescaled by dropping certain packets from the enhancement quality layer.

2. PERFORMANCE ANALYSIS OF H.264/AVC

The main characteristic of any codec is its compression performance and video quality. The compression performance of the codec is signified by multiple factors like decoder's computational performance and efficiency. The quality of the video can be assessed by objective method known as PSNR. Microsoft Windows 7 based 64 bit Operating System with Intel Core i3 CPU @2.13GHz and 3GB RAM is used for the purpose of performing experiments and results generation. Three types of video sequences are used in the experiments namely Mobile, Foreman and the Akiyo video sequences. These video sequences are in YUV 4: 2: 0 color format, have 8 bits/color, each with 300 frames and 10 seconds duration. These video sequences were considered on the basis of the criteria that they have different amounts of the information, dynamism and variations in the contents of the videos.

3. EXPERIMENTS AND RESULTS

Experiments are conducted while employing the considered three different types of scalabilities.

Experiments related to Temporal scalability are conducted using different sizes of GOP, that is, 4, 8, 16 respectively. The resolution of all the video sequences in these experiments is 352x288. The decoding time, bit rate and PSNR are observed with varied GOP sizes.

The key factor that affects the decoding efficiency is the GOP size. The graph in the **(Fig.1)** shows the decoding complexity revealed by the three video sequences considered for the experiments. Incase of Mobile video sequence, it is quite visible that highly varied video contents and considerably large amount of information contained in it, highly effects GOP size. The presence of P and B frames increases with the increase in GOP size. It results in the enhanced prediction mechanism which leads to a decreased decoding time. Along with it, it can also be noticed that, with the smallest GOP dimension, i.e. 4, the Mobile video sequence yields the longest time required for decoding due to the generation of least number of B and P frames. So, it can be inferred that the decoding time slowly reduces as the GOP size increases. The Foreman and Akiyo video sequences have least impact on the decoding time as the GOP size increases, due to the lower information and the less varied video contents present in them. As most of information distributed in these frames (P and B frames) is the same with less variable video content present in them. Therefore, in comparison to the Mobile video sequence, even with the increased number of P and B frames generated due to increase in the GOP size, the decoding time of the two video sequences i.e. Foremen Akiyo is not increased.

In Temporal scalability, due to the effect of different GOP size, varying bit rate is observed for the considered video sequences. **(Fig. 2)** demonstrates that in case of Mobile video sequence, the increase in the GOP size increases the bit rate. This video sequence has highest bit rate as compared to the other two video sequences (Foreman and Akiyo). This is due to the generation of increased number of temporal layers which helps in generating more P and B frames and results in increased bit rate of the encoded bit stream. On the other hand, depending on the amount of information and variation in the video contents, the bit rate of Forman video sequence is less as compared to Mobile video sequence and more in case of Akiyo video sequence. Conclusively, increase in the GOP size increases the bit rate but if the video has smaller amount of information and less changing contents like Foreman and Akiyo video sequences than it requires less number of bits to encode the information.

(Fig. 3) shows that as the GOP size increases, the quality of the Mobile video sequence is increased. Specifically in Temporal scalability, as the GOP sizes moves from 4 to 8 and then 8 to 16 the video quality is much improved. The impact of increasing GOP size basically increases the number of Temporal layers and as a result more B frames are generated. This improves the overall video quality as B frames utilize inter layer bi-predictive coding. The increase in GOP size has no major impact on the quality of Foreman video sequence and in case of Akiyo video sequence it becomes least evident. This shows that by increasing the GOP size, the number of B frames are also increased but the quality of those frames are improved that have a bulk of information like Mobile video sequence. But, increase in GOP size and B frames bring no prominent change in the overall quality of the videos that have lesser amount of information and lesser varying video contents such as Foreman and Akiyo video sequences. So from these experiments, it is clear that increase in GOP size that generates more B frames does not always leads to better quality of the video.

In regard to Spatial scalability, the two scenarios kept into consideration are Dyadic and Non-Dyadic. Many experiments are performed to find out that what level of decoding time, bit rate and PSNR of video series (picked for tests) are influenced in afore mentioned situation of Spatial scalabilities (Dyadic and Non-Dyadic Scenario). In the Dyadic situation, three layers of dyadic ratio 2:1 are considered. Resolutions of the three layers selected for Dyadic situation are 176x144 (Base Layer), 352x288 (Enhancement Layer) and 704x576 (Enhancement Layer) respectively. In non-Dyadic scenario, resolutions of the three layers are 176x144(Base Layer), 336x144 (Enhancement Layer) and 640x480 (Enhancement Layer)respectively.

The graph in the **(Fig.4)** shows the decoding time of the video series for Dyadic and Non-Dyadic situations. It can be observed thatNon-Dyadic situation has more decoding effectiveness than Dyadic case. Moreover, it can be concluded that the rescaling algorithm (simple bit shift operation) employed in the encoding procedure for producing an encodedbit stream results in decoding load leading to amplified decoding duration. In Non-Dyadic relation, the rescaling algorithm used in the encoding phenomena is recognized as ESS (Extended Spatial Scalability) and is highly effective thus decoding this encoded bit stream leads to less decoding duration (increased decoding efficiency).With Dyadic ratios, Mobile video sequence is having highest decoding time due to high variation and large information content then comes the Foreman video sequence with second highest decoding time and finally Akiyo video sequence lowest decoding time of all due to least amount of information and lowest variation in the video contents. In Non-Dyadic situation, the decoding duration is almost same for all the three video sequences with a small alteration, leading to a complete decoding effectiveness as compared to Dyadic situation.

(Fig. 5) represents the graph depicting the bit rates of all the video sequences (Mobile, Foreman and Akiyo) in Dyadic as well Non-Dyadic situation individually. In case of Non-Dyadic situation, the bit rate for each video sequence is greater than the Dyadic one. Additionally, for Mobile video sequence, the change in bit rates of bit streams in both Dyadic and Non-Dyadic case are maximum while in case of Foreman and Akiyo video sequences it is less obvious. So the differences in bit rates for Dyadic and Non-Dyadic situations are very apparent. The bit stream generated after complex mathematical operation needs more bits for encoding and thus leads to higher bit rates. In Dyadic case, simple bit shift process is used for encoding which obviously requires fewer bits and will have less bit rate when compared to Non-Dyadic case. Further, the amount of information to be encoded in bit stream is also an influencing factor. Therefore, when the video has superior details and extremely changing video contents, the bit rate is considerably amplified (doubled) as in the situation of Mobile video sequence but as these aspects are lowered; the change in the bit rate of the encoded bit stream is hardly noticeable. So it is concluded, that if the video is encoded via easy and simple operation and have less details then the resulting encoded bit stream will possess less bit rate and vice versa.

The graph in the **(Fig. 6)** shows the PSNR in both Dyadic and Non-Dyadic situations for Mobile, Foreman and Akiyo video sequences respectively. Due to the lack of resizing alteration with dyadic-ratios, all video sequences show slightly advanced PSNR than non-dyadic ratios.

Finally, numerous experiments are conducted concerning SNR/Quality scalability and decoding time, bit rates and PSNR are observed respectively. The two techniques CGS and MGS of SNR /Quality scalability are used in the experiments. The resolution of all the video sequences in these experiments is 352x288.

First of all, the decoding duration of the video sequences is witnessed by using two modes of SNR scalability i.e. CGS and MGS. It is evident that decoding duration of all video sequences for MGS is higher than CGS as presented in graph in the **(Fig.7)**. This decrease in the decoding efficiency is due to the key picture concept which is applied in MGS to decrease the drift among motion-compensated forecast loop at encoder and decoder which exerts an extra computational burden on the decoder leading to the amplified decoding duration.

In SNR/Quality scalability, the important aspect which greatly impacts the bit rate of the encoded bit stream is the quantization parameter. Experiments are conducted to note the effect of varying the quantization parameter of the base and enhancement layer. The Mobile video sequence possessing highly varying video contents with a lot of details and information has been selected for the experiments. The graphs in the **(Fig.8)**. and **(Fig.9)**. respectively reveals the resulting bit rates and PSNR of the Mobile video sequence by varying the quantization parameter value of the base layer and enhancement layer in CGS and MGS mode. The values of the quantization parameter kept into consideration for the base layer QP_l (possessing inferior quality) are 36, 38, 40 and the values of the quantization parameter for the enhancements layer QP_u(possessing better quality) are26, 28, 30 respectively. The groupings of superior and inferior layer quantization parameter values taken are QP_l/QP_u: 36/26, 38/26, 40/26, 36/28, 38/28, 40/28, 36/30, 38/30, 40/30 respectively.

It is evident from the graphs in the Fig. 8 and Fig. 9 that when the values of the quantization parameter of the base layer (QP_l) are 36,38,40 respectively while keeping a constant value of quantization parameter of superior layer (QP_u) as 26, highest bit rate and PSNR are observed. Furthermore, in this circumstance, MGS also makes advanced bit rate and PSNR comparatively to CGS. But as the value of quantization parameter of improvement layer ($QP_{u)}$ is further stretched to 28 and 30 correspondingly, a significant weakening is seen in the bit rates and PSNR value. As the enhancement layer adds to upgrading of the superiority of whole encoded bit stream. Thus, with higher quantization values of the improvement layer(QPu) specifically, it is more probable not to assume its part to enhance the quality of the encoded bit stream. In addition, another perceptible component in this case is that CGS yields higher estimation of bit rate and PSNR than MGS.

It can also be gathered that the elevated quantization parameter values of upgrade layer (QPu), prompts to lower bit rates and quality of the encoded bit stream. Moreover, utilizing MGS with genuinely lower values of quantization parameter of upgrade layers (QPu), the bit rates and quality is significantly higher. As more references in the improved forecast calculation are accessible for better quality in MGS when contrasted with CGS, however its execution and productivity is to a great degree traded off with higher quantization parameter values of the upgrade layer (QPu). Therefore, the resulting encoded bit stream yields lower PSNR esteem. Clearly, when an encoded bit stream has lower quality it will definitely require fewer bits for the encoding procedure.

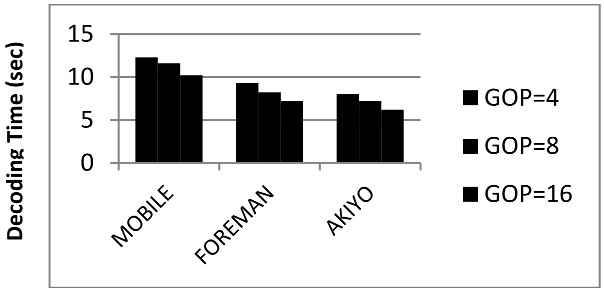

Fig. 1.Decoding Time of all video sequences with different GOP sizes in Temporal Scalability.

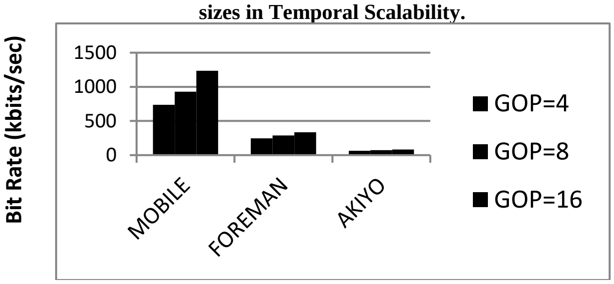

Fig.2. Bit Rates of all video sequences with different GOP sizes in Temporal Scalability.

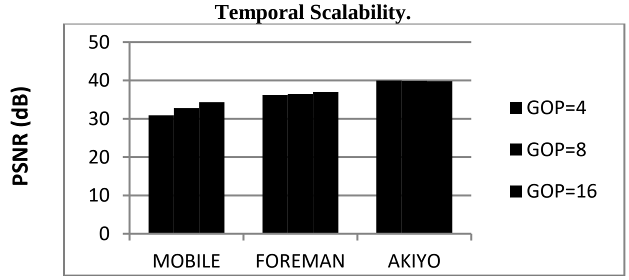

Fig.3.PSNR of all video sequences with different GOP sizes in Temporal Scalability.

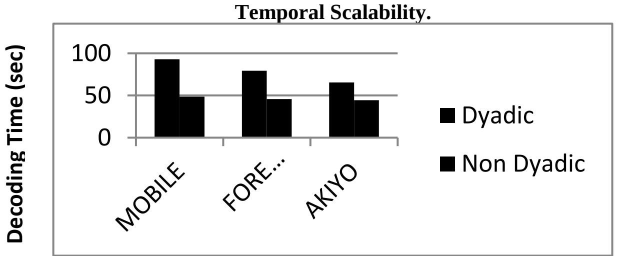

Fig. 4.Decoding Time of all video sequences with Spatial Scalability in Dyadic and Non-Dyadic Scenario.

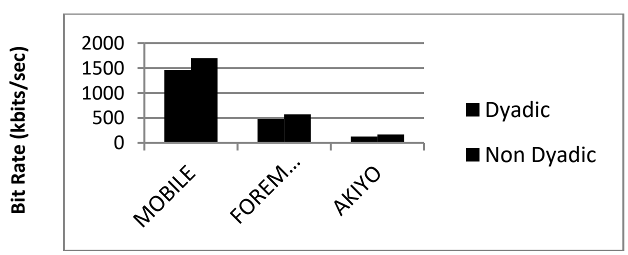

Fig. 5. Bit Rates of all video sequences with Spatial Scalability in Dyadic and Non-Dyadic Scenario.

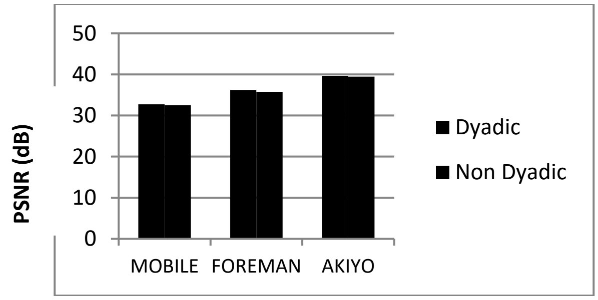

Fig.6.PSNR of all video sequences with Spatial Scalability in Dyadic and Non-Dyadic Scenario.

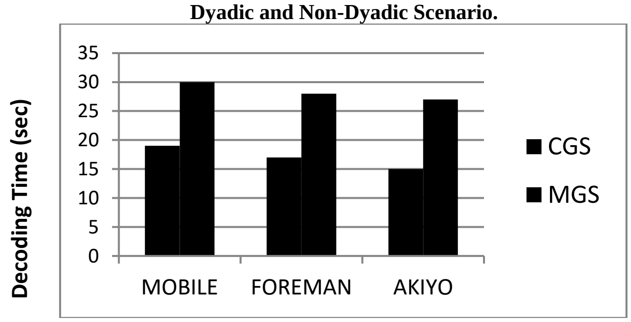

Fig. 7.Decoding Time of all video sequences with SNR scalability in CGS and MGS mode.

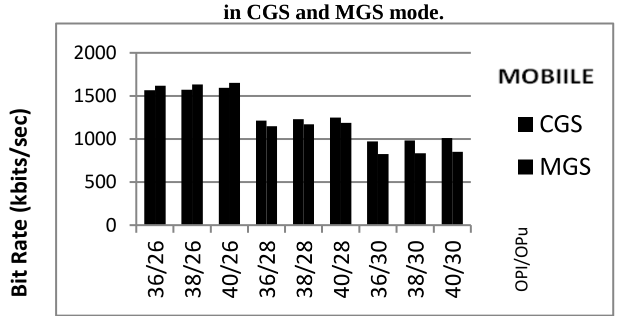

Fig. 8. Bit Rates of video sequence Mobile with SNR scalability in CGS and MGS mode.

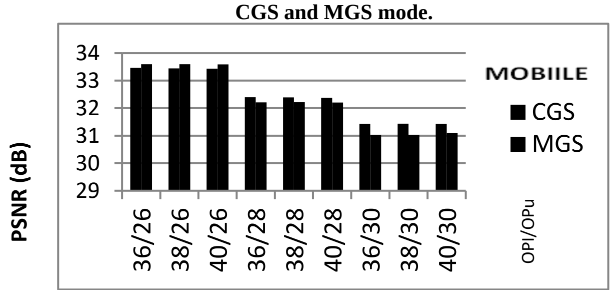

Fig.9.PSNR of video sequence Mobile with SNR scalability in CGS and MGS mode.

4. CONCLUSION

The conducted study constituted of several experiments in order to evaluate the performance of H.264/SVC standard. Three types of video sequences have been analyzed using different scalability types and the effect on the decoding time, bit rate and PSNR have been recorded. The experimental results in Temporal scalability demonstrated a clear understanding of the impact of increase in GOP size on the complexity of the codec, its bit rate and the ultimate quality of the video sequence, as explained in Section 3. Furthermore, the performance analysis of the two scenarios in Spatial scalability i.e. Dyadic and Non-dyadic case was also carried out. Finally, the experiments related to CGS and MGS types of SNR scalability were performed and the impact of quantization parameter on the performance of the video codec was observed. It is concluded that H.264/SVC standard with diverse coding capabilities and support for heterogeneous networking scenarios and varying nature of video processing devices is the optimal choice for the recent video transmission and communication systems.

REFERENCES:

Ambalgi, S. (2015) "A Comparative Study of Video Compression Techniques," vol. 1, no. 3, 174–177. "ISO/ITU." [Online]. Available: https://www.hhi.fraunhofer.de/fileadmin/PDF/VCA/IVG/file514.doc. [Accessed: 13-Oct-2016].

Ben, O. R. and L. C.Fourati, (2013) "Routing protocols performance analysis for scalable video coding (SVC) transmission over mobile ad-hoc networks," 197–202. "difference." [Online]. Available: http://www.differencebetween.net/technology/difference-between-mpeg2-and-mpeg4/.[Accessed:18-Oct-2016].

"EE Times." [Online]. Available: http://www.eetimes.com/document.asp?doc_id=1275532. [Accessed: 14-Oct-2016].

Huang, H. C., W. H. Peng, T. Chiang, and H. M. Hang, (2007) "Advances in the scalable amendment of H.264/AVC," IEEE Commun., vol. 45, no. 1,. 68–76.

Hewage, (2009) "Scalable video coding," vol. 11201, no. 1–13,.

Kessentini, A., I. Werda, A. Samet, M. Ali, B. Ayed, and N. Masmoudi, (2011) "H . 264 / SVC Performance and Encoder Bit- stream Analysis," vol. 36, 8,. 30–37.

Kamaci N. and Y. Altunbasak, (2003) "Performance comparison of the emerging H. 264 video coding standard with the existing standards," 2003 Int. Conf. Multimed. Expo, ICME'03. Proc., vol. 1,. 345–348.

Moon, Y. H., G. Y. Kim, and J. H. Kim, (2005) "An improved early detection algorithm for all-zero blocks in H.264 video encoding," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 8, 1053–1057,.

Magli, E. and J. R. Casas, (2014) "Master Thesis (Projecte Final De Carrera) Performance Comparison of Video Compression Algorithms for Digital.

Ohm, J. (2007). "Scalable Video Coding SVC – Technical Solution," no. October.

Rieckh J. and M. Rupp, (2008) "Scalable video for peer-to-peer streaming," 51Pp.

Richardson, I. (2011) The H. 264 advanced video compression standard.

Schierl, T., T. Stockhammer, and T. Wiegand, (2007) "Mobile Video Transmission Using Scalable Video Coding," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9. 1204–1217,.

Sanz-Rodríguez, S., M. Álvarez-Mesa, T. Mayer, and T. Schierl, (2015) "A parallel H.264/SVC encoder for high definition video conferencing," Signal Process. Image Commun., vol. 30, no. January, 89–106.

Schwarz, H., D. Marpe, and T. Wiegand, (2007) "Overview of the Scalable Video Coding Extension of the H . 264 / AVC Standard," IEEE Trans. Circuits Syst. Video Technol., vol. 17, 9, 1103–1120.

Schafer R. and T. Sikora, (1995) "Digital video coding standards and their role in video\ncommunications," Proc. IEEE, vol. 83, no. 6.

Team, Joint Video. (2008) "JSVM Software Manual(version JSVM 9.12)." ITU-T document .

Unanue, I., I. Urteaga, R. Husemann, J. Del Ser, V. Roesler, A. Rodriguez, and P. Sanchez, (2011) "A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and PerformanceTitle," Recent Adv. Video Coding, 3–26,.

Winkler, S. (2005) Digital video quality: vision models and metrics. John Wiley\ & Sons. "wiki." [Online]. Available: http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC. [Accessed: 18-Oct-2016].

.