Performance Analysis of Advanced Video Coding (H.264)

Nikhilesh R. Deshpande[1], Prof.Ameya K.Naik[2],
[1,2]*Department of Electronics and Telecommunication Engineering, K.J.S.C.E, Mumbai*

***Abstract*— H.264/MPEG-4 Part 10 or AVC (Advanced Video Coding) is a video compression format, and is currently one of the most commonly used formats for the recording, compression, and distribution of video content. The final drafting work on the first version of the standard was completed in May 2003.**

This paper provides performance analysis of the ITU-T H.264 standard, presenting PSNR, MSE values to get better insight about the H.264 applications in industries and day to day life. Various test sequences are used to get better perception of the superiority of H.264 than previous video coding standards.

Keywords—H.264/AVC, PSNR, and MSE.)

I. INTRODUCTION

H.264 Advanced Video Coding is an industry standard for video coding which was first jointly published in 2003 .The standard was developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC joint working group, the Moving Picture Experts Group (MPEG) [1]. The product of this partnership effort is known as the Joint Video Team (JVT) [3]. Recommendation H.264: Advanced Video Coding [2] is the standard document which defines a format or syntax for the compressed video and a method for decoding this syntax to produce a displayable sequence.

The application focus for the initial version of the standard document was broad – from video conferencing to entertainment (broadcasting over cable, satellite, terrestrial, cable modem, DSL etc.; storage on DVDs and hard disks; video on demand etc.) to streaming video, surveillance and military applications, and digital cinema [4]. Only the central decoder is standardized, by imposing restrictions on the bit-stream and syntax, and defining the decoding process of the syntax elements such that every decoder conforming to the standard will produce similar output when given an encoded bit-stream that conforms to the constraints of the standard [5].

Motivated by the rapidly growing demand for coding of higher-fidelity video material, especially in application areas like professional film production, video post-production and high definition TV/DVD, the JVT issued a Call for Proposal for the support of extended sample bit depth and chroma format in the H.264/MPEG4-AVC standard, following which, in September 2004, the Fidelity Range Extensions (FREext) of H.264/MPEG4-AVC was included in version 4 of the standard document [6] [2].

There is a trend towards creating and delivering multiple views of the same video scene. Stereoscopic video, with suitable display technology, gives the impression of a 3D image. Multiple views of a scene can give the users the option of choosing their viewpoints. Free viewpoint video (FVV) can deliver any view of a scene, by synthesizing intermediate views between actual camera positions. The multi-view applications generally require coding of multiple, closely related video signals or views [1]. Multi-view video coding (MVC) was standardized as an extension to H.264, which provides compact representation for multiple views of a video scene, such as multiple synchronized video cameras. It enables inter-view prediction to improve compression capability, as well as support of ordinary temporal and spatial predictions [7].

H.264 is based on hybrid video coding – video is compressed using a hybrid of motion compensation and transform coding. These video coding algorithms compress the video data by reducing the redundancies inherent in video, which fall into four classes, namely, spatial, temporal, perceptual and statistical [8]. Various tools are used by video coding algorithms to reduce these redundancies:

i. Chroma sub sampling, quantization and pre-filtering to remove perceptual redundancies
ii. DCT, intra-prediction, integer transform and variable block size transform to remove spatial redundancies
iii. Block motion estimation, multiple reference frame motion estimation and variable block size motion estimation to remove temporal redundancies
iv. Huffman coding [9], adaptive VLC (variable length coding) [9] and arithmetic coding [9] to remove statistical redundancies.
v. These algorithms differ in which tools are used for reducing the redundancies and in the specific ways these tools are applied [8].

II. PROFILES AND LEVELS

H.264/AVC contains a rich set of video coding tools. Every application doesn't require all coding tools hence, subsets of coding tools are defined; these subsets are called profiles [6]. Profiles and levels specify conformance points that provide interoperability between encoder and decoder. They also provide implementations within applications of the standard and between various applications that have similar functional requirements [10]. A profile defines a set of syntax features that is used for generating conforming bit-streams, whereas a level places constraints on certain key parameters of the bit-stream such as maximum bit rate and maximum picture size.

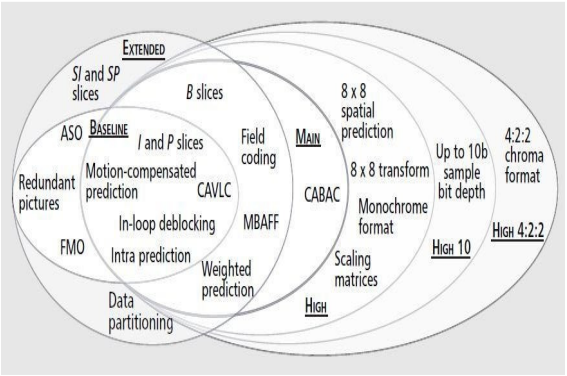

Figure 2.1 Illustration of profiles in H.264/AVC [10]

- Baseline profile: Targeted at low cost mobile applications and videoconferencing applications in which a minimum of computational complexity and a maximum of error robustness are required.
- Main profile: Targeted at standard-definition digital TV broadcast applications that require a maximum coding efficiency, with slightly less emphasis on error robustness.
- Extended profile: Intended for streaming video and designed to provide a compromise between the Baseline and Main profile capabilities with an additional focus on the specific needs of video streaming applications, and further added robustness to errors and packet losses.
- The FREXt amendment [4], which was released in 2004, defines four new profiles in H.264 [4]:
- High (HP) for high definition broadcast and disc storage applications supporting 8-bit video with 4:2:0 sampling.
- High 10 (Hi10P) with support for up to 10 bits of representation accuracy per sample of decoded picture precision.
- High 4:2:2 (Hi422P) with support for 4:2:2 chroma sub sampling and up to 10 bits per sample.
- High 4:4:4 (Hi444P) supporting up to 4:4:4 chroma sub sampling and up to 12 bits per sample and additionally supporting efficient lossless region coding and an integer color transform for coding RGB video while adding color-space transformation error.

For real-time decoders or decoders with constrained memory size, it is important to specify the processing power and memory size needed for implementation. Picture size plays the main role in influencing these parameters. H.264/AVC defines 16 different levels, tied mainly to the picture size [6]. Levels also provide constraints on the number of reference pictures and the maximum compressed bit rate that can be used. In the standard, levels specify the maximum frame sizes in terms of only the total number of pixels/frame. Table 2.1 shows 16 different levels defined for H.264/AVC standard.

III. H.264/AVC ENCODER

H.264 video encoder carries out prediction, transform and encoding processes to produce a compressed H.264 bit stream [1].

A coded video sequence in H.264/AVC consists of a sequence of coded pictures. A coded picture can represent either an entire frame or a single field. A frame of video can be considered to contain two interleaved fields: a top field and a bottom field.The typical encoding operation for a picture begins with splitting the picture into blocks of samples. The first picture of a sequence or a random access point is typically coded in Intra mode.This is done without using any other pictures as prediction references. Each sample of a block in such an Intra picture is predicted using spatially neighboring samples of preciously coded blocks. For all remaining pictures of a sequence or between random access points, Inter (inter-picture) coding is used. Interceding employs inter picture temporal prediction using other previously decoded pictures.

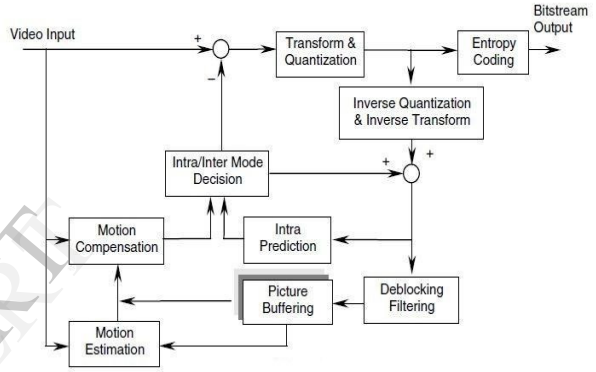

Figure 2.2 Block representation of an H.264/AVC video encoder [11]

The residual of the prediction (either Intra or Inter), which is the difference between the original input samples and the predicted samples for the block, is transformed. The transform coefficients are then scaled and approximated using scalar quantization. The quantized transform coefficients are entropy coded and transmitted together with the entropy-coded prediction information. The encoder contains a model of the decoding process so that it can compute the same prediction values obtained in the decoder for the prediction of subsequent blocks in the current picture or subsequent coded pictures.

IV. H.264/AVC DECODER

The H.264 decoder is illustrated in Figure 3.4. The decoder works similar to the local decoder at the encoder. The decoder receives the compressed H.264 bit stream, decodes each of the syntax elements and extracts the following information:

- Quantized transform coefficients
- Prediction information
- Information about the structure of the compressed data and the compressed tools used during encoding
- Information about the complete video sequence

After entropy (CABAC or CAVLC) decoding, the transform coefficients are inverse scanned and inverse quantized prior to being inverse transformed. To the resulting blocks of the residual signal, an appropriate prediction signal (intra or inter)

is added depending on the macro block type and mode, the reference frame and the motion vectors. The reconstructed video frames undergo de-blocking filtering prior to being stored for future use for prediction. The frames at the output of the de-blocking filter may need to undergo reordering prior.

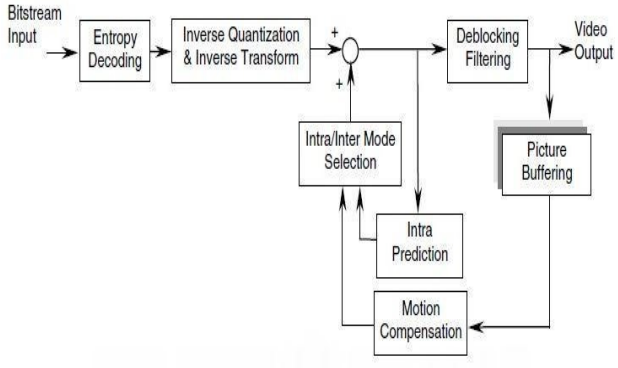

Figure 3.3 H.264/AVC decoder block diagram [11]

V. IMPLEMENTATION

Different video formats are used to perform the analysis and comparative tests, from low quality video to high definition quality video. AVC encoder is used to encode the test sequences, which are present in the AVI format. The simulation results are conducted based on the following configuration and test settings specified below.

The configuration of the H.264/AVC encoder initialization:

1. Frame Start = 1
2. Frame End = 10
3. Quantization Parameter QP = 27
4. The input video test sequence is resized to greyscale video of frame size of specifications:

Width = 128
Height = 128

5. The macro block size for the P frames obtained is set to '16'.

Mat lab 2013 software is used for simulation of encoding and decoding sequences using the H.264/AVC video compression standard as reference.

A. *Formulae*

The MSE (Mean Squared Error) and the related peak signal-to-noise ratio PSNR are popularly used to assess image quality.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad \ldots\ldots\ldots (5.1)$$

The PSNR is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$
$$= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$
$$= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \ldots (5.2)$$

Here, MAXI is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented using linear PCM with B bits per sample, MAXI is 2B−1. For color images with three RGB values per pixel, the definition of PSNR is the same except the MSE is the sum over all squared value differences divided by image size and by three. Alternately, for color images the image is converted to a different color space and PSNR is reported against each channel of that color space, e.g., YCbCr

Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, provided the bit depth is 8 Bit, where higher is better. For 16 Bit data typical values for the PSNR are between 60 and 80 db. [5][6] Acceptable values for wireless transmission quality loss are considered to be about 20 dB to 25 dB [12].

TABLE I
INPUT VIDEO TEST SEQUENCES

No.	Sequence Name	Frame Rate	Resolution	Duration
1	Basketball drive	30	1920x1080	0.70
2	Claire	30	352x264	1.33
3	Coastguard	23.9760	3840x2160	10.0
4	Foreman	25	176x144	12.1
5	Garden	30	352x240	6
6	Kirsten and Sara	59.9401	1280 x720	9.99
7	Marketplace	24	852x480	22.6
8	Sequence Name	23.9760	3840x2160	10.6
9	Basketball drive	30	352x264	1.33
10	Suzie	25	176x144	6
11	Tennis	30	352x240	6
12	Video Traffic	30	2560x1600	5

TABLE II
SEQUENCE NAME: BASKETBALL DRIVE.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	4.52563476562500	41.5740085890144
Frame 2	4.86753929751717	41.2577089443527
Frame 3	5.10976839610218	41.0467914499431
Frame 4	4.97901590117379	41.1593684761093
Frame 5	5.13208846667566	41.0278622675467
Frame 6	5.14151808933211	41.0198899275241
Frame 7	5.41026500852506	40.7986182239326

Frame 8	5.22082513236686	40.9534121378245
Frame 9	5.27363547451257	40.9097025377967
Frame 10	5.17276038635766	40.9935799946885

TABLE III
SEQUENCE NAME: COASTGUARD.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	2.53735351562500	44.0869938154943
Frame 2	2.81296300668760	43.6391634007743
Frame 3	3.01133337353857	43.3432152356380
Frame 4	2.91330528971008	43.4869436347358
Frame 5	2.87036236729553	43.5514363340826
Frame 6	2.79414690626543	43.6683112483663
Frame 7	2.71401550359118	43.7946803666546
Frame 8	2.66481138060827	43.8741388643966
Frame 9	2.61267065710592	43.9599569312918
Frame 10	2.57895836604411	44.0163602979754

TABLE IV
SEQUENCE NAME: CLAIRE.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	1.58258056640625	46.1371453244391
Frame 2	1.66687388662399	45.9117761791763
Frame 3	1.76059350241209	45.6742126613749
Frame 4	1.77271968685505	45.6444029310487
Frame 5	2.10173101354048	44.9050322805781
Frame 6	2.26864418627886	44.5731397430563
Frame 7	2.45862185994550	44.2238862219320
Frame 8	2.42582295781603	44.2822125899904
Frame 9	2.40030624177674	44.3281370639418
Frame 10	2.39762347400507	44.3329937893626

TABLE V
SEQUENCE NAME: FOREMAN.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	7.35363769531250	39.4657813196311
Frame 2	7.51229095826398	39.3730796064055
Frame 3	7.20387097465225	39.5551443510719
Frame 4	7.10551779875945	39.6148462920962
Frame 5	7.25075947117084	39.5269686231567
Frame 6	7.09348885003637	39.6222047044592
Frame 7	6.89440927560001	39.7458329983214
Frame 8	6.91590912527494	39.7323108283788
Frame 9	6.56178390971641	39.9605843661048
Frame 10	6.56990836986311	39.9552104834623

TABLE VI
SEQUENCE NAME: GARDEN.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	8.70581054687500	38.7327114892432
Frame 2	8.18815581136619	38.9989426266485
Frame 3	8.00551366727792	39.0969115756609
Frame 4	8.00189731425117	39.0988738694918
Frame 5	8.46052213899301	38.8568319462723
Frame 6	8.01422839448044	39.0921864579979
Frame 7	8.10884571230668	39.0412132372876

Frame 8	8.95925861859263	38.6080828772093
Frame 9	8.13202835695057	39.0288147648409
Frame 10	8.09020135636866	39.0512103000081

TABLE VII
SEQUENCE NAME: MARKET PLACE.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	5.97235107421875	40.3693503182254
Frame 2	5.94795496016741	40.3871268962338
Frame 3	5.96065748482943	40.3778619403295
Frame 4	5.68986585249855	40.5797833352493
Frame 5	5.68126070319929	40.5863564211291
Frame 6	5.98231982090621	40.3621073372695
Frame 7	5.91126851579034	40.4139967350728
Frame 8	5.95056715708498	40.3852199993705
Frame 9	5.80136262080560	40.4955034847859
Frame 10	5.65993183978160	40.6026915967698

TABLE VIII
SEQUENCE NAME: KIRSTEN AND SARA.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	0.481323242187500	51.3064352742644
Frame 2	0.383382230997086	52.2944838054377
Frame 3	0.371479183293559	52.4314587879834
Frame 4	0.371479183293559	52.4314587879834
Frame 5	0.371479183293559	52.4314587879834
Frame 6	0.371479183293559	52.4314587879834
Frame 7	0.371479183293559	52.4314587879834
Frame 8	0.371479183293559	52.4314587879834
Frame 9	0.371479183293559	52.4314587879834
Frame 10	0.371479183293559	52.4314587879834

TABLE IX
SEQUENCE NAME: NEWS.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	1.1589	47.4904
Frame 2	1.0780	47.8045
Frame 3	1.0822	47.7879
Frame 4	1.1489	47.5279
Frame 5	1.1502	47.5232
Frame 6	1.1288	47.6046
Frame 7	1.1405	47.5597
Frame 8	1.1419	47.5544
Frame 9	1.1443	47.5452
Frame 10	1.1419	47.5544

TABLE X
SEQUENCE NAME: SALESMAN.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	7.7047	39.2633
Frame 2	7.7240	39.2524
Frame 3	7.6048	39.3199
Frame 4	7.6160	39.3135
Frame 5	7.9029	39.1529
Frame 6	8.6683	38.7515
Frame 7	8.9877	38.5943
Frame 8	8.9813	38.5974

Frame 9	8.9231	38.6257
Frame 10	8.8832	38.6451

TABLE XI
SEQUENCE NAME: TENNIS.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	16.7581	35.8886
Frame 2	16.8541	35.8637
Frame 3	16.8474	35.8655
Frame 4	16.8020	35.8772
Frame 5	16.8125	35.8745
Frame 6	16.1828	36.0403
Frame 7	16.4002	35.9823
Frame 8	16.9950	35.8276
Frame 9	16.6301	35.9218
Frame 10	16.7368	35.8941

TABLE XII
SEQUENCE NAME: SUZIE.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	5.5741	40.6691
Frame 2	5.6652	40.5987
Frame 3	5.7216	40.5556
Frame 4	6.1230	40.2611
Frame 5	6.4866	40.0107
Frame 6	6.4885	40.0094
Frame 7	6.6605	39.8958
Frame 8	6.9904	39.6858
Frame 9	6.9416	39.7162
Frame 10	6.9872	39.6878

TABLE XIII
SEQUENCE NAME: TRAFFIC.AVI

Frame Number	Mean Squared Error	Peak Signal to Noise Ratio
Frame 1	2.6032	43.9757
Frame 2	2.6236	43.9419
Frame 3	2.6014	43.9788
Frame 4	2.6021	43.9775
Frame 5	2.6726	43.8614
Frame 6	2.6603	43.8815
Frame 7	2.7654	43.7132
Frame 8	2.7658	43.7126
Frame 9	2.8285	43.6153
Frame 10	2.8064	43.6492

VI. CONCLUSION

There is a *tradeoff* between Mean Squared Error and Peak Signal Ratio, as the MSE values increase in the magnitude PSNR values show degradation in there magnitudes.

As the resolution (width and height) of the video increases H.264 performs better than compared to the videos for lower resolution. On comparing TableXIII and XI, the we realize that *traffic* sequence which has the resolution 2560 x 1600 has lower MSE values, which means better PSNR (Increase in PSNR $\Rightarrow$ Better image quality in terms of intensity), whereas if we glimpse PSNR values of the *Tennis* sequence they are comparatively lower than traffic sequence. The same case is evident in the other tables as well. This shows that "*As the resolution of the video increases, encoding and decoding becomes more effective leading to better Peak Signal to Noise Ratio*".

In the case of *lower resolution videos*, the magnitude of the MSE *increases* slightly for each increment in the frames and the PSNR correspondingly *decreases.* Table V, X, XI, XII provide the necessary evidence for the above observation. This case is not applicable to the *high resolution videos*; MSE remains *constant* for most of the frames leading to *steady* value of PSNR. Table IV, VII, IX & XIII provide the proof of the previous statement.

The acceptable value of PSNR is approximately 30 to 40dB for lossy image and video compression. Whereas for lossless it ranges from 40dB to 50dB.

The above inferences show that, H.264 *outperforms* in the case of high resolution videos whereas its efficiency decreases in terms of PSNR and MSE if we consider low resolution videos.

REFERENCES

[1] I.E. Richardson, the H.264 advanced video compression standard, 2nd Edition, Hoboken, NJ: Wiley, 2010.

[2] Advanced video coding for generic audiovisual services, ITU-T Rec. H.264 / ISO / IEC 14496-10, Jan. 2012.

[3] S. Kwon et al, "Overview of H.264/MPEG-4 part 10", Journal of Visual Communication and Image Representation, vol. 17, no. 2, pp. 186-216, April 2006.

[4] G.J.Sullivan et al, "The H.264/AVC advanced video coding standard: overview and introduction to the fidelity range extensions". SPIE conference on Applications of Digital Image Processing XXVII, vol. 5558, pp. 53-74, Nov. 2004.

[5] Open Source Article http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC

[6] T. Wiegand et al, "H.264/MPEG4-AVC fidelity range extensions: tools, profiles, performance, and application areas", IEEE ICIP 2005, vol. 1, pp. 593-596, Sep. 2005.

[7] A. Vetro et al, "Overview of the stereo and multi-view video coding extensions of the H.264/MPEG-4 AVC standard". Proceedings of the IEEE, vol. 99, pp. 626-642, Apr. 2011.

[8] H.Kalva, "The H.264 video coding standard". IEEE Multimedia, vol. 13, no. 4, pp.86-90, Oct. 2006. 56

[9] K.Sayood,"Introduction to data compression",Elsevier, Third edition, 2005.

[10] D. Marpe et al, "The H.264/MPEG-4 AVC standard and its applications", IEEE Communications Magazine, vol. 44, pp. 134-143, Aug. 2006

[11] S. Kwon et al, "Overview of H.264/MPEG-4 part 10", Journal of Visual Communication and Image Representation, vol. 17, no. 2, pp. 186-216, April 2006.

[12] M. Pinson and S. Wolf. "A New Standardized Method for Objectively Measuring Of Video Quality," IEEE Trans. on Broadcasting, vol. 50, no. 3, pp. 312-322, Sep. 2004.