[Parvez*, 5(4): April, 2016] ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

COMPARATIVE STUDY OF H.265 & H.264 VIDEO COMPRESSION

Khan Adil Parvez*, Sanjeev Kumar Shrivastava
Pillai Institute of Information Technology, Engineering and Media Studies & Research New Panvel Navi Mumbai India

DOI: 10.5281/zenodo.48845

ABSTRACT
Video is the sequence of images played with respect to time. The successive images are highly correlated with each other. Video compression algorithms take the advantage of this fact. Only the residual information is transmitted using the technique called as block based motion estimation and motion compensation [2]. MPEG1 (Motion Picture Expert Group), MPEG2, MPEG4, H.261, H.263, H.264 are the ancestors of H.265. Work on the emerging "Advanced Video Coding" (AVC) standard now known as ITU-T(International Telecommunication Union) Recommendation H.264 and as ISO 14496(International Organization For Standards) (MPEG-4) Part 10 has dominated the video coding standardization community. The work has been stimulating, intense, dynamic, and all-consuming for those of us most deeply involved in its design. The time has arrived to see what has been accomplished. The new H.264/AVC & Enhanced version H.265 standard is designed to provide a technical solution, the H.265 is the latest video compression technique in which 50% of bit rate is saved more than H.264 and broad range of applications, including broadcast over cable, satellite, cable modem, and terrestrial. It finds the applications in interactive or serial storage on optical and magnetic storage devices, DVDs (Digital Video Disk). Conventional services over Ethernet, LAN (Local Area Network), wireless and mobile network and mobile.

KEYWORDS: Entropy coding, intra estimation, loop filter, motion estimation, quantization (and inverse quantization), transform (and inverse transform).

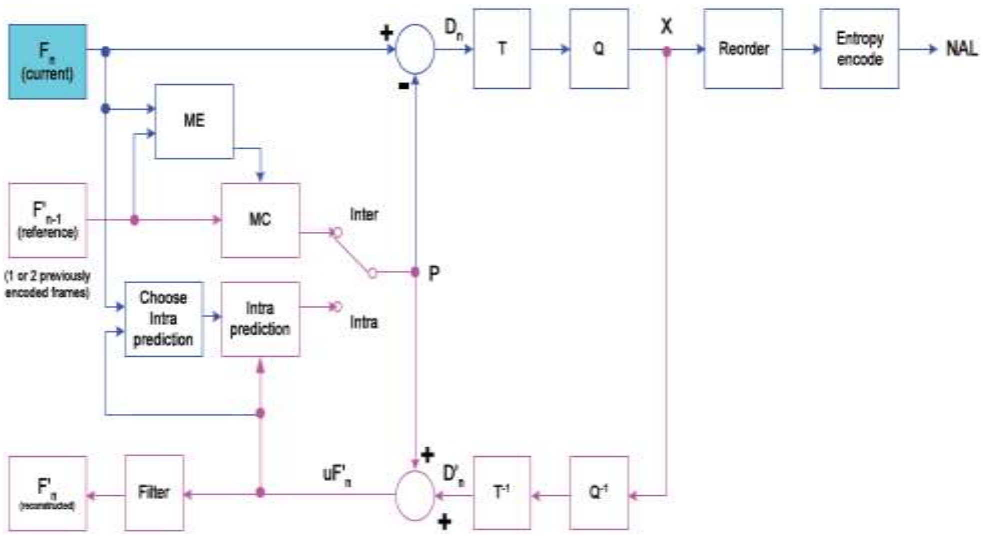
INTRODUCTION
The upcoming H.264 AVC video compression standard promises a significant improvement over all previous video compression standards. In terms of coding efficiency, the new standard is expected to provide at least 2x compression improvement over the best previous standards and substantial perceptual quality improvements over both MPEG-2 and MPEG-4.The standard, being jointly developed by ITU-T and ISO/IEC, will address the full range of video applications including low bit-rate wireless applications, standard-definition and high-definition broadcast television, video streaming over the Internet, deliver y of high-definition DVD content, and the highest quality video for digital cinema applications.

As can be seen "History of Video Standards", the ITU-T and ISO/IEC are responsible for all previous international video compression standards. To date, the most successful of these standards has been MPEG-2, which has gone on to achieve mass-market acceptance in areas such as DVD, digital television broadcast (over cable and satellite),and digital set-top box. The new H.264 standard represents the single largest improvement in coding efficiency and quality since the introduction of MPEG-2. Consequently, over time, it is expected that H.264/MPEG-4 AVC will displace MPEG-2 and MPEG-4 ASP in many existing applications.

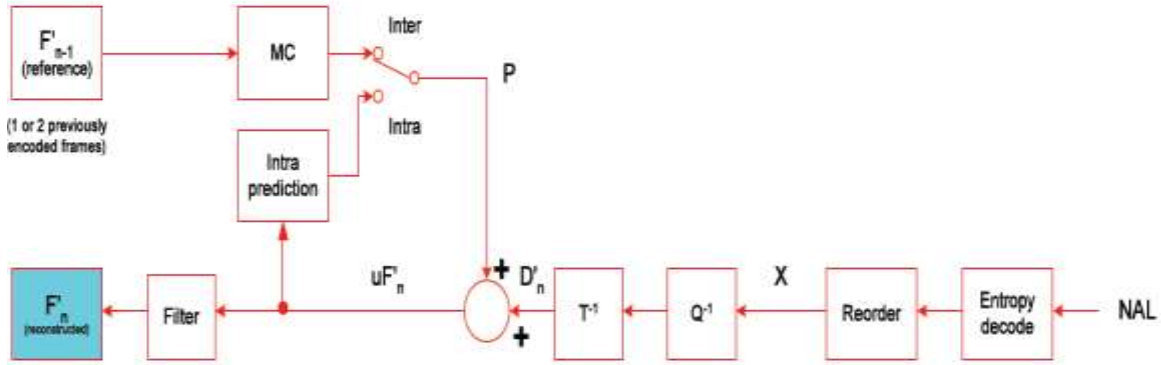
TECHNICAL OVER VIEW
AS CAN BE SEEN IN THE "H.264/MPEG-4 AVC – OVER VIEW BLOCK DIAGRAM", THE NEW STANDARD IS COMPOSED OF SEVERAL PROCESSING STAGES:

Motion Estimation and Intra Estimation
- Transform (and Inverse Transform)

- Quantization (and Inverse Quantization)
- Loop Filter
- Entropy Coding



Encoder of advance video coding algorithm H.264 (Adopted from White paper: A technical introduction to H.264/AVC) Fig. 1:H.264 CODEC



Decoder of advance video coding algorithm H.264 Fig. 2:H.264 DECDER

Video is composed of a stream of individual pictures that can be broken down into individual blocks of 16 pixels by 16 lines called "macroblocks". This practice simplifies the processing which needs to be done at each stage in the compression algorithm. We will explore the purpose and function of each of these processing elements in the next few Sections.

MOTION ESTIMATION AND INTRA ESTIMATION

Motion estimation is used to identify and eliminate the temporal redundancies that exist between individual pictures. When searching for motion relative to a previous picture, the picture to be encoded is called a "P-picture". When searching both within a previous picture and a future picture, the picture to be encoded is called a "B-picture"[2].To improve coding efficiency, the macroblock is broken down into smaller blocks that attempt to contain and isolate the motion as shown in the diagram "H.264 Motion Estimation – Superior Motion Estimation". Then, motion vectors to previous and/or future pictures are used to predict a given block. H.264 introduces smaller block sizes greater flexibility in block shapes, and greater precision in motion vectors.

A) INTRA ESTIMATION

In instances where motion estimation cannot be exploited, intra estimation is used to eliminate spatial redundancies. Intra estimation attempts to predict the current block by extrapolating the neighboring pixels from adjacent blocks in a defined set of different directions. The difference between the predicted block and the actual block is then coded. This approach, unique to H.264, is particularly useful in flat backgrounds where spatial redundancies often exist. An example of this is shown in "H.264 Intra Estimation.

B) TRANSFORM

Results from the motion estimation or intra estimation stages are transformed from the spatial domain into the frequency domain. H.264 uses a DCT-like 4x4 integer transform. In contrast, MPEG-2 and MPEG-4 employ a true DCT 8x8 transform that operates on floating-point coefficients the smaller block size of H.264 reduces blocking and ringing artifacts. Integer coefficients eliminate rounding errors inherent with floating point coefficients and that cause drifting artifacts with MPEG-2 and MPEG-4

DISCRETE COSINE TRANSFORM

The Discrete Cosine Transform (DCT) operates on X, a block of N × N samples (typically image samples or residual values after prediction) and creates Y, an N × N block of coefficients. The action of the DCT (and its inverse, the IDCT) can be described in terms of a transform matrix A. The forward DCT (FDCT) of an N × N sample block is given by:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^T \quad \text{---1}$$

And the inverse DCT (IDCT) by:

$$\mathbf{X} = \mathbf{A}^T\mathbf{Y}\mathbf{A} \quad \text{---2}$$

Where **X** is a matrix of samples, **Y** is a matrix of coefficients and **A** is an $N \times N$ transform matrix.
The elements of **A** are:

$$A_{ij} = C_i \cos\frac{(2j+1)i\pi}{2N} \quad \text{where } C_i = \sqrt{\frac{1}{N}}\ (i = 0), \quad C_i = \sqrt{\frac{2}{N}}\ (i > 0)$$

Equation 1 and equation 2 may be written in summation form:

$$Y_{xy} = C_x C_y \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} X_{ij} \cos\frac{(2j+1)y\pi}{2N} \cos\frac{(2i+1)x\pi}{2N}$$

$$X_{ij} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} C_x C_y Y_{xy} \cos\frac{(2j+1)y\pi}{2N} \cos\frac{(2i+1)x\pi}{2N}$$

The output of a two-dimensional FDCT is a set of N × N coefficients representing the image block data in the DCT domain and these coefficients can be considered as 'weights' of a set of standard basis patterns. The basis patterns for the 4 × 4 DCTs is shown in Fig. 3 composed of combinations of horizontal and vertical cosine functions. Any image block may be reconstructed by combining all N × N basis patterns, with each basis multiplied by the appropriate weighting factor (coefficient).

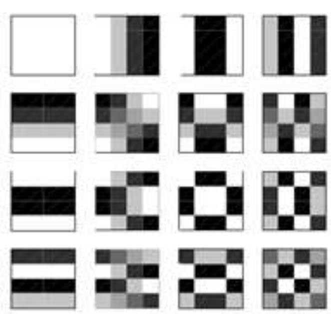


Fig. 3: 4 × 4 DCT basis patterns

QUANTIZATION

The coefficients from the transform stage are quantized, which reduces the overall precision of the integer coefficients and tends to eliminate high frequency coefficients, while maintaining perceptual quality. The Quantizer is also used for constant bit rate applications where it is varied to control the output bit rate.

A) LOOP FILTER

The H.264 standard defines a de-blocking filter that operates on both 16x16 macroblocks and 4x4 block boundaries. In the case of macroblocks, the filter is intended to remove artifacts that may result from adjacent macroblocks having different estimation types (e.g. motion vs. intra estimation), and/or different quantizer scale. In the case of blocks, the filter is intended to remove artifacts that may be caused by transform/quantization and from motion vector differences between adjacent blocks. The loop filter typically modifies the two pixels on either side of the macroblock boundary using a content adaptive non-linear filter[5].

REORDERING

Scanning of the coefficients is called as reordering. Depending on whether these coefficients were originally motion estimated or intra estimated, a different scan pattern is selected to create the serialized stream. The scan pattern orders the coefficients from low frequency to high frequency. Then, since higher frequency quantized coefficients tend to be zero, run-length encoding is used to group trailing zeros, resulting in more efficient entropy coding[1].

ENTROPY CODING

The entropy coding stage maps symbols representing motion vectors, quantized coefficients, and macroblock headers into actual bits. Entropy coding improves coding efficiency by assigning a smaller number of bits to frequently used symbols and a greater number of bits to less frequently used symbols.

Variable Length Coding (VLC) and Context Adaptive Binary Arithmetic Coding (CABAC) can be used. CABAC offers superior coding efficiency over VLC by adapting to the changing probability distribution of symbols, by exploiting correlation between symbols, and by adaptively exploiting bit correlations using arithmetic coding. H.264 also supports Context Adaptive Variable Length.Coding (CAVLC) which offers superior entropy coding over VLC without the full cost of CABAC.

INTRODUCTION TO H.265

The High Efficiency Video Coding (HEVC) standard is the most recent joint video project of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) standardization organizations, working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC). HEVC is among us. On January 25, 2013, the ITU announced the competition of the first stage approval of the H.265 video codec standard and in the last 1 year several vendors/entities have started to work on the first implementations of H.265 encoders and decoders. Theoretically HEVC is said to be from 30 to 50% more efficient than H.264 (especially at higher resolutions). First of all let's start with a technical analysis of H.265 compared to AVC and then, in the next blog post, we will take a look at the current level of performance that is realistic to obtain in today's H.265 encoders.

A) TECHNICAL OVERVIEW

HEVC re-uses many of the concepts defined in H.264. Both are block based video encoding techniques so have the same roots and the same approach to encoding:

- Subdivision of picture in macroblocks, eventually sub-divided in blocks
- Reduction of spatial redundancy using intra-frame compression techniques
- Reduction of temporal redundancy using inter-frame compression techniques (motion estimation and compensation).
- Residual data compression using transformation & quantization.
- Reduction of final redundancy in residuals and motion vectors transmission and signaling using entropy coding.

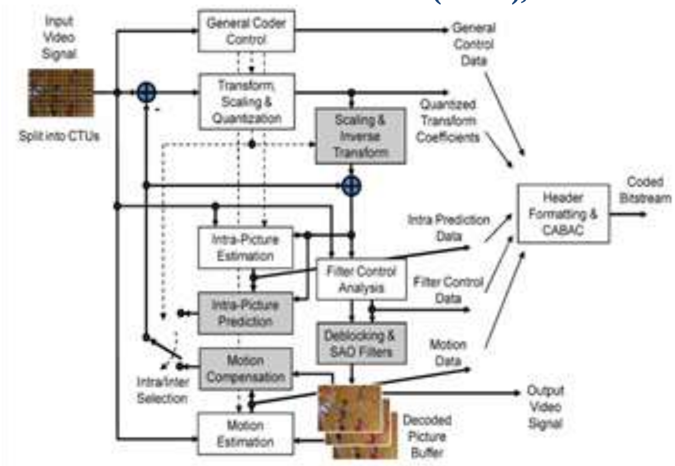


Fig. 4:H.265 Encoder

The video coding layer of HEVC employs the same hybrid approach (inter-/intrapicture prediction and 2-D transform coding) used in all video compression standards since H.261. Fig. 1 depicts the block diagram of a hybrid video encoder, which could create a bitstream conforming to the HEVC standard. An encoding algorithm producing an HEVC compliant bitstream would typically proceed as follows. Each picture is split into block-shaped regions, with the exact block partitioning being conveyed to the decoder. The first picture of a video sequence (and the first picture at each clean random access point into a video sequence) is coded using only intrapicture prediction (that uses some prediction of data spatially from region-to-region within the same picture, but has no dependence on other pictures). For all remaining pictures of a sequence or between random access points, interpicture temporally predictive coding modes are typically used for most blocks. The encoding process for interpicture prediction consists of choosing motion data comprising the selected reference picture and motion vector (MV) to be applied for predicting the samples of each block.

The encoder and decoder generate identical interpicture prediction signals by applying motion compensation (MC) using the MV and mode decision data, which are transmitted as side information. The residual signal of the intra- or interpicture prediction, which is the difference between the original block and its prediction, is transformed by a linear spatial transform. The transform coefficients are then scaled, quantized, entropy coded, and transmitted together with the prediction information. The encoder duplicates the decoder processing loop (see gray-shaded boxes in Fig. 1) such that both will generate identical predictions for subsequent data. Therefore, the quantized transform coefficients are constructed by inverse scaling and are then inverse transformed to duplicate the decoded approximation of the residual signal. The residual is then added to the prediction, and the result of that addition may then be fed into one or two loop filters to smooth out artifacts induced by block-wise processing and quantization.

The final picture representation (that is a duplicate of the output of the decoder) is stored in a decoded picture buffer to be used for the prediction of subsequent pictures. In general, the order of encoding or decoding processing of pictures often differs from the order in which they arrive from the source; necessitating a distinction between the decoding order (i.e., bitstream order) and the output order (i.e., display order) for a decoder. Video material to be encoded by HEVC is generally expected to be input as progressive scan imagery (either due to the source video originating in that format or resulting from DE interlacing prior to encoding). No explicit coding features are present in the HEVC design to support the use of interlaced scanning, as interlaced scanning is no longer used for displays and is becoming substantially less common for distribution. However, a metadata syntax has been provided in HEVC to allow an encoder to indicate that interlace-scanned video has been sent by coding each field (i.e., the even or odd numbered lines of each video frame) of interlaced video as a separate picture or that it has been sent by coding each interlaced frame as an HEVC coded picture. This provides an efficient method of coding interlaced video without burdening decoders with a need to support a special decoding process for it.

PICTURE PARTITIONING

Instead of 16×16 macroblocks like in AVC, HEVC divides pictures into "coding tree blocks (CTBs)". Depending by an encoding setting the size of the CTB can be of 64×64 or limited to 32×32 or 16×16. Several studies have shown that bigger CTBs provide higher efficiency (but also higher encoding time). Each CTB can be split recursively, in a quad-tree structure, in 32×32, 16×16 down to 8×8 sub-regions, called coding units (CUs). See the picture below for an example of partitioning of a 64×64 CTB (numbers report the scan order). Each picture is further partitioned in special groups of CTBs called Slices and Tiles.

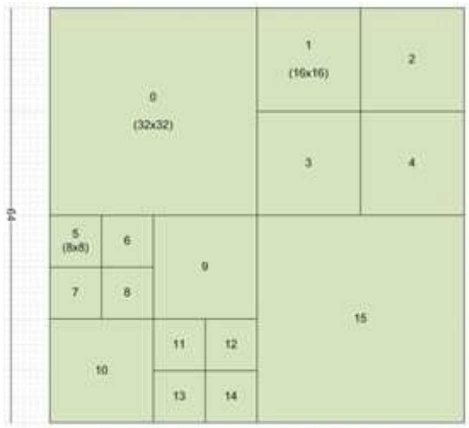


Fig. 5:H.265 Picture Partitioning

TRANSFORM SIZE

Each CU can be recursively splitted in Transform Units (TUs) with the same quad-tree approach used in CTBs. Differently from AVC that used mainly a 4×4 transform and occasionally an 8×8 transform, HEVC has several transform sizes: 32×32, 16×16, 8×8 and 4×4. From a mathematical point of view, bigger TUs are able to encode better stationary signals while smaller TUs are better in encoding smaller "impulsive" signals. The transforms are based on DCT (Discrete Cosine Transform) but the transform used for intra 4×4 is based on DST instead (Discrete Sine Transform) because several tests have evidenced a small improvement in compression.

Transformation is performed with higher accuracy compared to H.264. The adaptive nature of CBT, CU and TU partitions plus the higher accuracy plus the larger transform size are among the most important features of HEVC and the reason of the performance improvement compared to AVC. HEVC implements a sophisticated scan order and coefficient signaling scheme that improves signaling efficiency. Note that unlike H.264 there's neither Hadamard nor 2×2 chroma (min chroma transform size is 4×4). HEVC drops also the support for MBAFF or similar techniques to code interlaced video. Interlaced video can still be compressed but there's no separation between fields and frames (only frames).

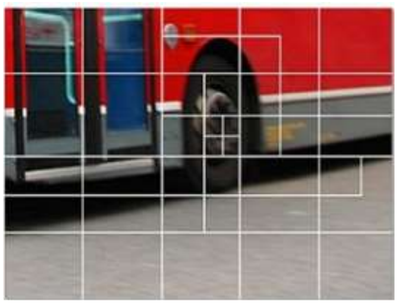


Fig. 6: Transform Size

PREDICTION UNIT & FILTERING

A) PREDICTION

We have introduced the new transform sizes just after the picture partitioning to exploit the analogy between CU and TU trees, but before transform and quantization there's the prediction phase (inter or intra). A CU can be predicted using one of eight partition modes.

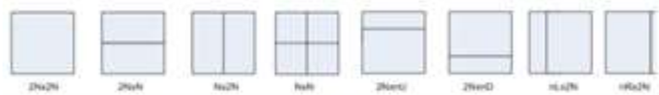


Fig.7: Modes of Partition

Even if a CU contains one, two or four prediction units (PUs), it can be predicted using exclusively inter-frame or intra-frame prediction technique, furthermore Intra-coded CUs can use only the square partitions 2Nx2N or NxN. Inter-coded CUs can use both square and asymmetric partitions. A number of other limitations are applied to simplify signaling. For example no 4×4 prediction is allowed in inter-prediction and 4×8 and 8×4 are allowed only in forward prediction (so not in b-frames). Tangentially inter-prediction stops at 8×8 level.

B) DEBLOCKING FILTER

Unlike h264 where deblocking was performed on 4×4 blocks, in HEVC deblocking is performed on the 8×8 grid only. This allows for parallel processing of deblocking (there's no filter overlapping). All vertical edges in the picture are deblocked first, followed by all horizontal edges. The filter is similar to AVC.

C) SAO (SAMPLE ADAPTIVE OFFSET)

After Deblocking there's a second optional filter. This filter is called Sample Adaptive Offset, or SAO. Similarly to Deblocking filter, it is applied in the prediction loop and the result stored in the reference frames list. The objective of the filter is to fix mis-predictions, encoding drift and banding on wide areas subdividing the colors in "bands" and applying adaptive offset to them.

ENTROPY CODING

In HEVC there's only CABAC for entropy coding. CABAC in HEVC is almost identical to CABAC in AVC with minor changes and simplifications to allow a parallel decoding.

Context-adaptive binary arithmetic coding (CABAC) is a form of entropy encoding used in the High Efficiency video Coding (HEVC) standard. It is a lossless compression technique, although the video coding standards in which it is used are typically for lossy compression application. CABAC is notable for providing much better compression than most other entropy encoding algorithm used in video encoding and it is one of the key elements that provide the H.265 encoding scheme with better compression capacity than its predecessor.

A) PARALLEL PROCESSING

Since HEVC decoding is much more complex than AVC, several techniques to allow a parallel decoding have been implemented. The most important are: Tiles and Wave front. The picture is divided into a rectangular grid of CTBs (Tiles). Motion vector prediction and intra-prediction is not performed across tile boundaries. With Wave front Each CTB row can be encoded & decoded by its own thread. Multiple rows encoding / decoding are synchronized (entropy coding state) guarantying that each "wave front" CTB is surrounded by specific CTB during encoding and decoding.

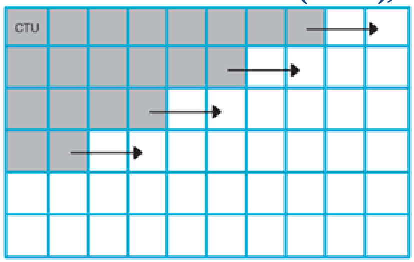


Fig.8: Parallel Processing

COMPARISON OF H.264 & H.265

Sr.no	Parameter	H.264	H.265
1	Partition size	Macroblock 16x16	(Large) Coding Unit 8x8 to 64x64
2	Partitioning	Sub-block down to 4x4	Prediction Unit Quad-tree down to 4x4 Square, symmetric and asymmetric (only square for intra)
3	Transform	Integer DCT 8x8, 4x4	Transform Unit square IDCT from 32x32 to 4x4 + DST Luma Intra 4x4
4	Intra prediction	Up to 9 predictors	Up to 35 predictors
5	Motion prediction	Spatial Median (3 blocks)	Advanced Motion Neighbor (3 blocks) Vector Prediction AMVP (spatial + temporal)
6	Motion precision	½ Pixel 6-tap,. ¼ Pixel bi-linear	¼ Pixel 7or 8 tap 1/8 Pixel 4-tap Chroma
7	Entropy coding	CABAC, CAVLC	CABAC
8	Partition size	Macroblock 16x16	(Large) Coding Unit 8x8 to 64x64

CONCLUSION

Years of effort by many hundreds of researchers and developers have led to the standardization of H.264/MPEG-4 and new generation H.265/MPEG-HEVC. The performance comparison of H.265 and H.264 encoders was presented. The study H.264 has taken a more pragmatic, focused approach to addressing the problems and needs of current and emerging multimedia applications. The coding efficiency of H.264 is inferior to H.265 with an average bit-rate overhead at the same objective quality of 40%.

REFERENCES

[1] Iain E.G. Richardson, "H.264 and MPEG – 4 Video Compression", Video Coding for Next-generation Multimedia, John Wiley & Sons LTD, 2003.
[2] Khalid Sayood, "Introduction to Data compression", Morgan Kaufmann Publishers, 2008.
[3] Video Compression – from Fundamentals to H.264 and H.265 Standards, international Journal of Engineering and Computer Science ISSN: 2319-7242. Volume 4 Issue 7 July 2015, Page No. 13468-13473
[4] Introduction to the Issue on Video Coding: HEVC and Beyond, ieee journal of selected Topics in signal processing, vol. 7, no. 6, December 2013.
[5] "Performance comparison of H.265/MPEGHEVC, VP9, and H.264/MPEG-AVC encoders," Grois, D., Marpe, D., Mulayoff, A., Itzhaky, B., and Hadar, O. in [PCS], 394–397 (2013).
[6] Comparison of Video Compression Standards S. Ponlatha and R. S. Sabeenian International Journal of Computer and Electrical Engineering, Vol. 5, No. 6, December 2013.

[7] Overview of the High Efficiency Video Coding (HEVC) Standard Gary J. Sullivan, Fellow, IEEE, Jens-Rainer Ohm, Member, IEEE, Woo-Jin Han, Member, IEEE, and Thomas Wiegand, Fellow, IEEE. Dec 2012.
[8] The Comparison of the Coding efficiency of video coding standards–including "high efficiency video coding" (HEVC) Circuits and Systems for Video Technology, Ohm, J., Sullivan, G. J., Schwarz, H., Tan, T. K., and Wiegand, IEEE Transactions on 22(12), 1669–1684 (2012).
[9] White paper: "A technical introduction" to H.264/AVC, Iain E.G. Richardson, 2011.
[10] New Techniques for Next Generation Video Coding Li Liu, Robert Cohen, Huifang Sun, Anthony Vetro, Xinhua Zhuang TR2010-058 April 2010.
[11] "Statistical Learning Based Intra Prediction in H.264", Cheolhong An and Truong Q. Nguyen, IEEE, 2008.
[12] "The SP and SI Frames Design for H.264/AVC" M. Karczewicz and R. Kurceren, IEEE Transactions on Circuits and Systems for Video Technology, 2003.
[13] Context-Based Adaptive Binary Arithmetic Coding the H.264/AVC Video Compression Standard" D. Marpe, H. Schwarz and T. Wiegand, ", IEEE Transactions on Circuits and Systems for Video Technology 2003.
[14] "In-loop Deblocking Filter for H.264/AVC Video" by Gulistan Raja and Muhammad Javed Mirza.