International Journal of Applied Engineering Research ISSN 0973-4562 Volume 17, Number 4 (2022) pp. 427-435

Video Compression Using H.265 (HEVC-Main Profile)

Vijayalaxmi S.Patil

Assistance Professor
Department of Information Science and Engineering
Basaveshwar Engineering College
Bagalkot 587102, India.

Dr.Suvarna Nandyal

Professor and Head
Department of Computer Science
P.D.A College of Engineeing
Kalburgi, India.

Abstract

For a long time, the television industry has worked to improve video resolution. The video business has recently looked at ways to increase representation's brightness and color spectrum. This is accomplished by using High Dynamic Range/ Wide Color Gamut (HDR/WCG) films, which, compared to Standard Dynamic Range (SDR) films, provide end-users with a visual quality of experience similar to real life. Practical compression algorithms must store or transmit video, especially high-resolution video formats while keeping an "acceptable" level of video quality to provide users with high-quality video services. High-Efficiency Video Coding (HEVC) is one such technique (HEVC). It is widely used since it is the most up-to-date video codec with excellent compression characteristics.

1 INTRODUCTION

The High-Efficiency Video Coding (HEVC) [2] standard for video compression was developed by the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). H.264/MPEG-4 AVC was used to transmit HD TV signals through satellite, cable, terrestrial transmission systems, video content capture and editing systems, camcorders, and security applications [5, 6]. It supplanted all previous video compression standards [9]. However, with the increasing number of services, the popularity of HD video, and the introduction of beyond HD formats (e.g., 4k2k or 8k4k quality), a more efficient coding standard than H.264/MPEG-4 AVC [5,] is required. The current state-of-the-art video coding standard HEVC [1] meets these requirements.

HEVC's primary goal is to significantly improve compression performance over existing standards such as H.264/MPEG-4 AVC [4], with a focus on two key issues: higher video resolution formats and increased use of parallel processing architectures, resulting in bit-rate reductions of up to 50% at comparable visual quality [1]. It is best suited for consumer applications because of the limited pixel formats of 4:2:0 8-bit and 4:2:0 10-bit. In addition, HEVC promises to simplify transportation system integration and data loss resistance. It features 4K and 8K resolutions and bit levels of 8, 10, 12, and 16 bits per sample.

Motivation The following are examples of today's multimedia environment:

- The number of HD (1920x1080) and UHD (4Kx2K, 8Kx4K) video materials increased.
- High-Efficiency Video Coding (HEVC) is a cutting-edge video coding technology [2].
- To speed up the data, use more dimensions.
- In 2019 [13], consumer internet video traffic will account for 80% of total consumer internet traffic.

Effective compression methods must be used to save or transmit video, particularly at HD and UHD resolutions, while retaining an 'acceptable' level of video quality to supply high-quality video services to customers. As a consequence, the emphasis of the research is on compressing HD video using the widely used HEVC video codec, which is presently the most sophisticated video codec.

2 BASICS OF VIDEO COMPRESSION

Today's society is surrounded by gadgets that range from simple mobile phones to sophisticated high-end technology capable of recording, processing, transferring, receiving, and presenting movies and multimedia data. Any video shot or created by any device must be stored or transferred to another device for sharing. As a result, digital video processing, which relies on video coding technology, is crucial in our daily lives.

A movie is just a sequence of pictures presented at regular intervals and comprises pixels with brightness and color components [2]. The most important aspects of a video are its height, width, frame rate, and pixel values.

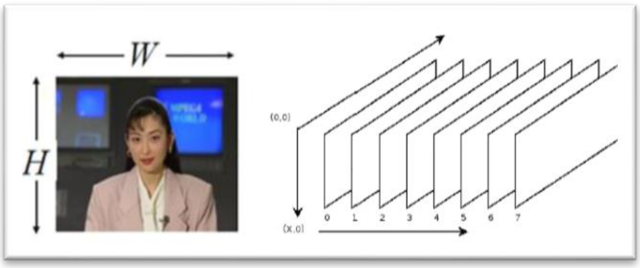


Figure 1: Video and its attributes [1] [6]

The following are the color spaces used in the digital picture and video representation:

- Three values represent the proportions of red, green, and blue colors for each pixel in the RGB color space.
- YCbCr color scheme is the luminance component, a monochrome version of the color image. The weighted average of R, G, and B yields Y:

Y=kr R + kg G +kb B, where k is the weighting factor.

Color disparities, also called chrominance components, communicate color information by expressing the difference between R, G, or B and Y in each chrominance component.

Because color is less sensitive to the human visual system than brightness, YCbCr has an advantage over RGB. The quantity of data required to represent the chrominance component is lowered while the visual quality is maintained [1]. Figure 2 shows a few examples of patterns.

The following are some of the most common sample patterns [1]:

- 4:4:4 - The three components, Y, Cr, and Cb, all have the exact resolution; therefore, there are four Cr and four Cb samples for every four luminance samples.
- There are two Cr and two Cb samples in the horizontal plane for every four luminance samples. For high-quality video color reproduction, this representation is employed.
- 4:2:0 – Horizontal and vertical resolutions are present in each half of the Y, Cr, and Cb. The 4:2:0 sampling pattern is seen in Figure 3. Video conferencing, digital television, and DVD storage all utilize this.

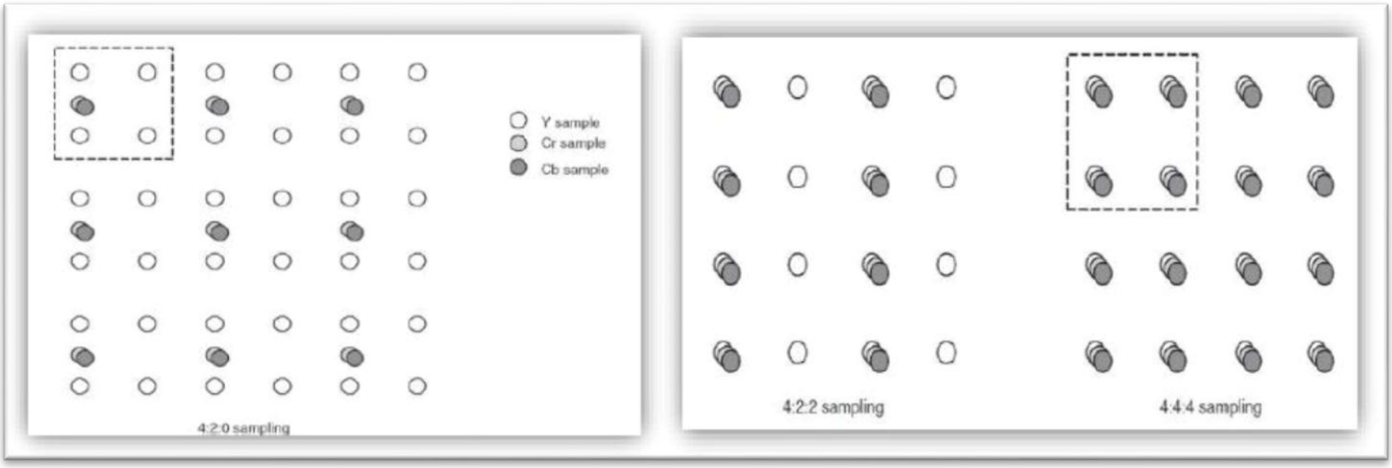


Figure 2: 4:2:0, 4:2:2 and 4:4:4 sampling patterns [1]

Figure 3: 4:2:0 sub-sampling pattern [6]

The digital representation of video material necessitates a large number of bits. Uncompressed video demands higher bit rates because the quantity of data created by digitizing a video stream necessitates a substantial amount of storage space and transport capacity. As a result, video compression is essential. The compression must minimize the original video's size while maintaining its quality. Every compression technology exists in order to provide data in a small format. Data may be delivered in fewer bits by removing extraneous information from the video sequence.

An image or frame is classified as an I-picture, P-picture, or B-picture. In I-images or intra-predicted frames, the picture or current frame is coded without referring to prior photographs or frames. Motion-compensated prediction using reference frames is employed to inter-code P- and B-pictures. Only one reference frame is used in P-pictures (the P- or I-picture that came before the current P-picture), whereas B-pictures utilize two (the P- and I-pictures preceding and following the current frame). Coding lowers the difference between expected and actual frames, resulting in less information being sent. Figure 4 [2] [6] [8] depicts intra and inter images. (P, B, and b)

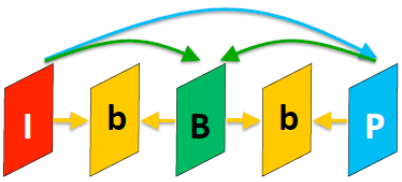


Figure 4: I-, P-, and b- frames [6]

Is compression still necessary now that transmission and storage capabilities have increased to the point where uncompressed video can be accommodated? Both storage and transmission capacity is growing. On the other hand, a welldesigned and efficient video compression system offers considerable visual communication performance benefits at low and high transmission bandwidths [1].

3 VIDEO CODING STANDARDS

Organizations such as the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), the Moving Picture Experts Group (MPEG), and the Joint Collaborative Team on Video Coding have developed video coding standards (JCT-VC). Each new standard is a step forward from the previous one. According to video compression standards, encoders and decoders must work together. There are typically several levels and profiles available to fulfill the demands of various use cases and applications. Every standard has stated that halving the bit rate results in the same video quality. Figure 5 shows how video coding standards have changed throughout time. [12]

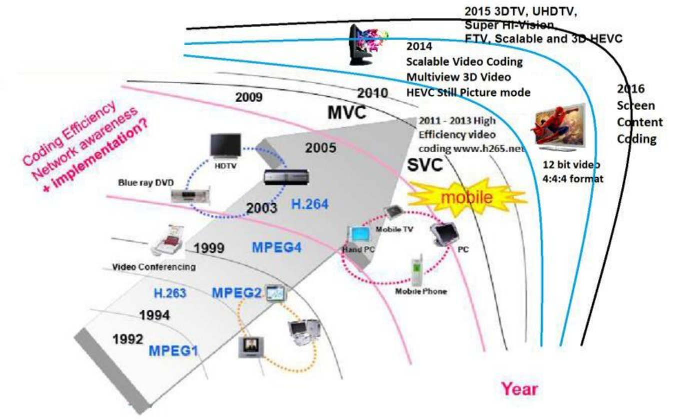


Figure 5: Evolution of video coding standards [12]

4 HIGH-EFFICIENCY VIDEO CODING (HEVC)

4.1 H.265 Encoder and Decoder

Since H.261, all video compression standards have used a hybrid method in their video coding layer (inter-/intra-picture prediction and 2-D transform coding). A video encoder is a device that turns a series of video frames from a source video into a compressed video bitstream. Either store or send the compressed bitstream. The data stream is decompressed by a video decoder, which produces a sequence of decoded frames [10]. Figure 6 depicts the encoder and decoder methods.

The stages taken by the video encoder are as follows:

- Each image is divided into many sections.
- Use inter or intra prediction to forecast each unit, then erase the prediction from the unit.
- Residue transformation and measurement (the difference between the original picture unit and the prediction).
- Transform output, prediction information, mode information, and headers are encoded using entropy.

The stages taken by the video decoder are as follows:

- Entropy is used to decode and retrieve the components of a coded sequence.
- The transform step has been reversed and rescaled.
- Predict each unit and combine the outcome of the inverse transform with the prediction.
- Putting together a video picture that has been decrypted.

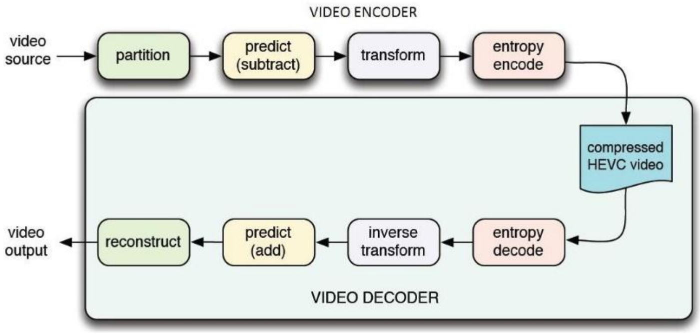


Figure 6: Block Diagram of HEVC CODEC [10]

4.2 Features of HEVC

4.2.1 Picture Partitioning

With HEVC, video sequence segmentation is highly versatile. Each frame comprises square-shaped Coding Tree Blocks (CTBs) ranging in size from 16x16 to 64x64 pixels. The Coding Tree Unit (CTU), HEVC's primary processing unit, comprises one Luma CTB and two chroma CTBs with syntactic components. Each CTB is predicted based on previously coded data. Following prediction, any remaining information is transformed, and entropy encoded.

Figures 7 and 8 demonstrate how each coded video frame or picture is partitioned into coding tree components after being divided into tiles and slices (CTUs). The CTU is a fundamental coding unit that may be as large as 64x64 pixels and is equivalent to the macro-block in previous standards.

The coding tree (CTU) may be broken into square portions called coding units (CUs) using a quad-tree topology. A CU's size may vary from 8x8 to 64x64 pixels. Each CU is broken into Prediction Units (PUs), which are then inter- or intra-predicted and translated using transform units.

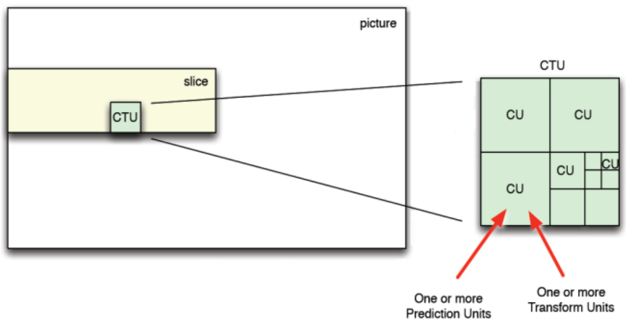
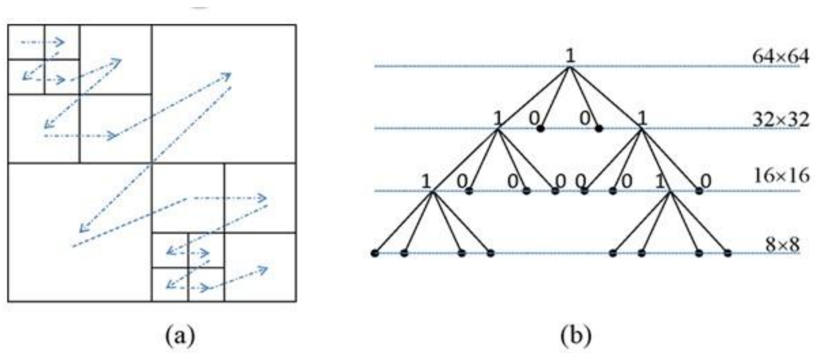


Figure 7: CTU partitioning structure [7]



Example of CTU, partitioning and processing order when size of CTU is equal 64 × 64 and minimum CU size is equal to 8 × 8 (a) CTU partitioning (b) Corresponding coding tree structure.

Figure 8: CTU partitioning and corresponding coding tree structure [7]

4.2.2 Prediction Schemes

The prediction in HEVC may be intra- or inter-predicted. The usage of intra- or inter-prediction is decided at the CU level. A "CTB slice" is a collection of CTBs. In intra mode, CUs are predicted using neighboring samples that have been recreated inside the same slice. For the CUs in the 'I' slice, only intra-prediction is enabled. CUs may be in either intra or inter-prediction mode in P and B slices.

Intra prediction: This technique, which considers spatial redundancy, predicts blocks by using nearby pixels reconstructed from the same frame. To forecast each PU, DC prediction (average value), planar prediction (fitting a flat surface to the PU), and directional prediction (using surrounding image data from the same picture) are all employed (extrapolating from neighboring data). The 35 intra-prediction modes in HEVC are DC, planar, and 33 angular modes. Horizontal prediction modes vary from 2 to 18, while vertical prediction modes range from 19 to 34. Because intra-coded CUs may only have 2Nx2N or NxN partition modes, intra PUs are always square.

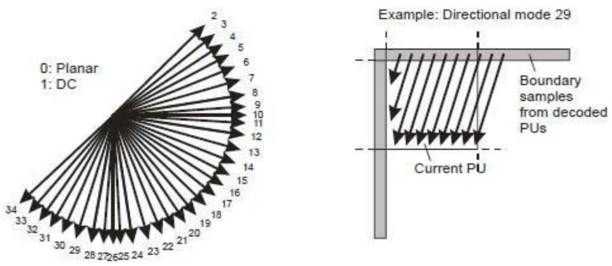


Figure 9: Prediction Schemes

Inter prediction: This is employed at the prediction block (PB) level. It foresees the current frame block by using temporal redundancy between adjacent frames. Inter prediction, like motion-compensated prediction, guesses the current PB using shifted sections of the reference pictures. Motion-compensated prediction and visual input from one or two reference images anticipate each PU (in display order, before or after the present picture). For inter-prediction, the splitting of CB into PB might be symmetric or asymmetric.

One or two reference images might be utilized as the prediction source for motion compensated prediction. The amount of available prediction sources is determined by the slice type to which the PB belongs. P slices may benefit from uni-prediction, which requires just one prediction reference. Two prediction sources and two reference image sets are employed for B slices. Both list 0 and list 1 reference images are used in bi-prediction. In this case, as shown in Figure 10, either uni-prediction or bi-prediction may be applied.

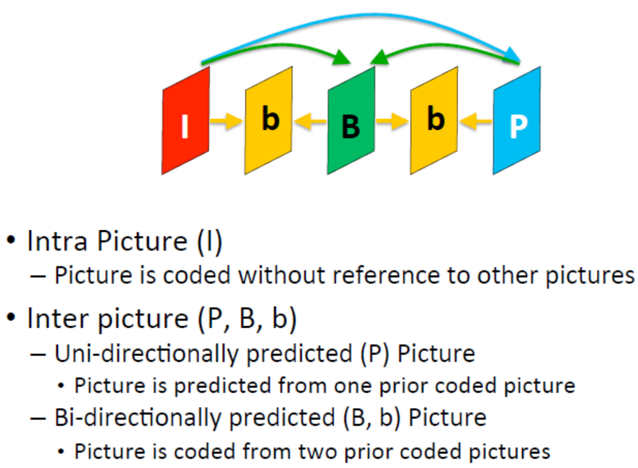


Figure 10: Picture Coding Types [6]

4.2.3 Motion Estimation and Motion Compensation [1]

Motion estimation (ME) is a technique for estimating motion vectors and presenting object transitions in a series of video frames (ME). Motion estimation serves two essential purposes in video applications: minimizing temporal redundancy in video codecs and presenting real-time object motion. The motion estimation algorithm in HEVC encodes the same perceived quality with less than half the coding complexity or time.

Based on information from one or more previously encoded frames, motion estimation creates a model of the current frame (referred to as "reference frames"). These frames of reference might be from the past or the future. A motion estimation approach aims to characterize the current frame properly (which enhances compression performance) while reducing computing complexity. Figure 11 depicts the notion of Motion Estimation.

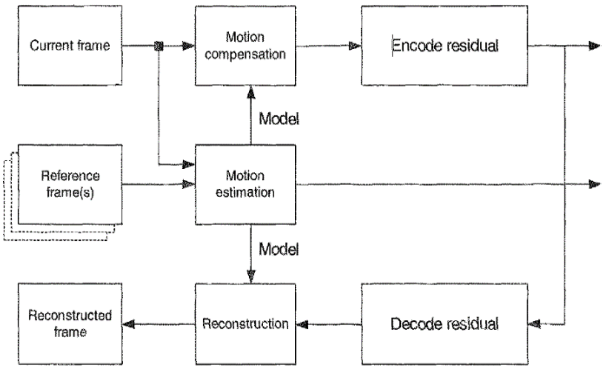


Figure 11: Motion estimation and Motion compensation block diagram [1]

The most extensively used motion estimation approach for video coding is block-based motion estimation (BBME) [17]. The concept of motion estimation is that the patterns associated with objects and background in a video stream change from frame to frame. The current frame is divided into a matrix of "macroblocks" compared to the matching block and its surrounding neighbors in the previous frame to generate a vector that depicts a macroblock's migration from one position to another. The overall motion computed for all macroblocks that comprise a frame is the estimated motion in the current frame. The previous frame's search zone for a good macro block match is limited to "p" pixels on all four sides of the relevant macroblock. The search parameter is represented by the letter "p." The block-based motion estimation approach is shown in Figure 12. Higher motions need a more significant p, and the greater the search parameter, the more computationally intensive the motion estimating method. The macroblock usually is a 16-pixel square with a p search parameter of 7 pixels.

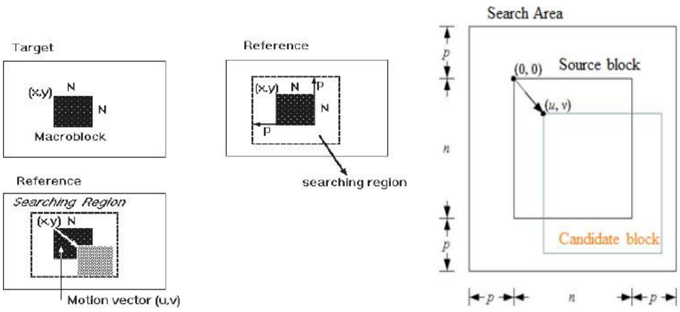


Figure 12: Block based motion estimation process [14]

Motion vector predictor and motion vector difference are two methods for determining the proper motion vectors for compensation. The motion vector predictor is picked from two alternatives based on the geographic and temporal neighbors of the current block. An approach for selecting predictors is advanced motion vector prediction. Furthermore, motion information such as motion vector and reference index may be collected without recording a motion vector difference by selecting a specified range of possibilities. A merge mode is a method for calculating motion vectors from many candidates [5].

4.2.4 Transforms

In HEVC, the transformation process is specified by fixed point integer operations with maximum output and intermediate values of 16 bits. After prediction, any residual data is modified using a block transform based on the Discrete Cosine Transform (DCT) or the 4x4 intra-Y Discrete Sine Transform (DST) (DST).

The residual data in each CU is updated in one or more 32x32, 16x16, 8x8, and 4x4 block sizes. To generate two-dimensional transformations, one-dimensional transforms are applied to both the horizontal and vertical axes. Each transform block (TB) comprises four square sub-blocks (coefficient groups). The last significant coefficient is processed first in the reverse scanning sequence, followed by the DC coefficient. The higher accuracy and bigger capacity of the transforms are two main reasons why HEVC surpasses H.264/MPEG-4 AVC [6][7][12].

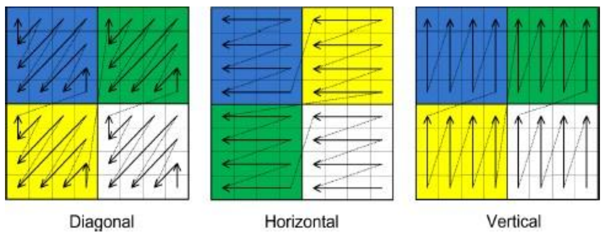


Figure 13: Coefficient scanning methods in HEVC [12] [15]

4.2.5 Quantization [12]

Following conversion, the data is quantized [6]. HEVC quantizes using a scalar technique known as uniform-reconstruction quantization (URQ), controlled by a quantization parameter (QP). Quantization is the primary cause of information loss in lossy video compression, such as HEVC. The QP scale has a scale of 0 to 51.

4.2.6 Entropy Coding [12]

A coded HEVC bitstream comprises quantized transform coefficients, prediction information such as modes and motion vectors, partitioning information, and header data. These items are encoded using Context Adaptive Binary Arithmetic Coding (CABAC) [8].

CABAC is an arithmetic coding strategy that uses primary coding data to update probability models. CABAC achieves high compression rates by selecting probability models for each syntactic element based on context, then updating probability estimates via local statistics and arithmetic coding.

Coding the data symbols involves the following steps [7]:

(1) "binarization" transforms syntactic components into binary symbols (0 or 1).

(2) The probability of bins is calculated via context modeling.

(3) Based on a probability calculation, arithmetic coding converts bins to bits.

4.2.7 In-loop filtering [2] [12]

Figures 2-1 depict the de-blocking and Sample Adaptive Offset (SAO) filters used by HEVC as the following in-loop filters. In the inter-picture prediction loop, a de-blocking filter similar to that used in H.264/MPEG-4 AVC. For parallel processing, the decision-making and filtering methods have been streamlined and improved[2]. Because the coding approach is block-based, it transforms blocks and applies them to prediction edges to reduce the number of visible block structures. This filter uses adaptive strength and length to operate on the block edges [5].

Sample Adaptive Offset (SAO) is a sample-based filtering algorithm based on CTUs. The slice's samples are filtered, not only the block edges. The output of the de-blocking filter is treated with the SAO treatment, which reduces ringing artifacts.

The DCT in HEVC works well on flat surfaces but fails when the signal has noise, contours, or other irregularities. It works great with huge blocks but not so well with little ones. When the transform size is increased, the artifacts become more visible. While de-blocking may help reduce artifacts inside the TB's borders, SAO is the only option to obliterate artifacts. Consequently, SAO may be employed on high-dimensional (32x32) transformations [7].

4.3 Profiles, Levels, and Tiers [2] [3] [16]

A profile is a set of coding tools for converting video sequences to bitstreams. A profile's encoder may use any coding tools as long as it creates a compliant bitstream, but its decoder must support all of them. The HEVC standard specified three profiles in its initial form: Main, Main 10, and Main Still image. The convergence of formerly distinct services such as broadcast, mobile, and streaming to the point that most devices should be able to handle them justifies reducing the number of profiles.

- *Principal Profile:* The most popular video format for consumer devices is the Main profile, which has an 8-bit per sample bit depth and 4:2:0 chroma sampling.

- *The Main 10 profile:* The chroma sampling is 4:2:0, with each sample having an 8 to 10-bit depth. To support the Main 10 profile, HEVC decoders must be able to decode bitstreams from both the Main and Main 10 profiles. A more significant bit of depth opens up a wider variety of color possibilities. A sample with 8 bits per bit may contain 256 shades per primary color, but a sample with 10 bits per bit may have 1024 shades. Because the color transition is smoother with higher depth, color banding is less of a problem. Because it can hold information with a greater depth than the Main profile, the Main 10 profile produces better video quality. Furthermore, compared to the Main profile, the Main 10 profile's 8-bit video may be coded with a 10-bit bit depth, resulting in improved coding efficiency.

- *The Main Still Picture profile:* This enables the encoding of a single still image while still adhering to the constraints of the Main profile. A subset of the Main profile, the Main Still Picture profile has an 8-bit per sample bit depth and 4:2:0 chroma sampling.

The levels correspond to the limits of decoding and buffering. All options are the maximum image size, coded and decoded picture buffer sizes, picture number of slice segments and tiles, maximum sampling rate, and maximum bitrate. It also gives the compression ratio for a stream with a bit-minimum size. The HEVC standard is divided into thirteen layers.

Various application types require different bitrate ranges. Hence tiers are used to differentiate them. As a result, the maximum bitrate and CPB size differ depending on the level. There are two compression layers in HEVC. The Main tier is for personal usage, while the High tier is for commercial use.

4.4 High-Level Syntax Architecture [2] [3]

A network abstraction layer (NAL) unit and a series of data units compose an HEVC bitstream. In HEVC, there are two types of NAL units: VCL NAL units and non-VCL NAL units. Non-VCL NAL units contain control information that translates to multiple coded images, while VCL NAL units carry just one slice of coded picture data. NAL units are available in 64 different configurations. An HEVC access unit comprises a coded image and non-VCL NAL components.

"High-level syntax" in HEVC [3][5] refers to the syntax components that regulate the bit-stream structure. High-level syntax includes parameter sets, reference image management syntax, and Supplemental Enhancement Information (SEI) messages, which may be applied to numerous pictures or coded block parts inside a picture.

HEVC specifies three parameter sets. The video, sequence, and image options (VPS, SPS, and PPS). VPS provides information on a coded video stream's many levels and sublayers. The SPS file contains all video sequence slices. PPS transmits information that changes from picture to image, such as QP tiles. Variables that may vary among slices include POC, slice type, prediction weights, de-blocking settings, and tile entry locations.

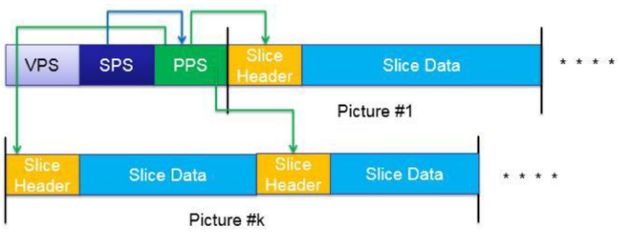


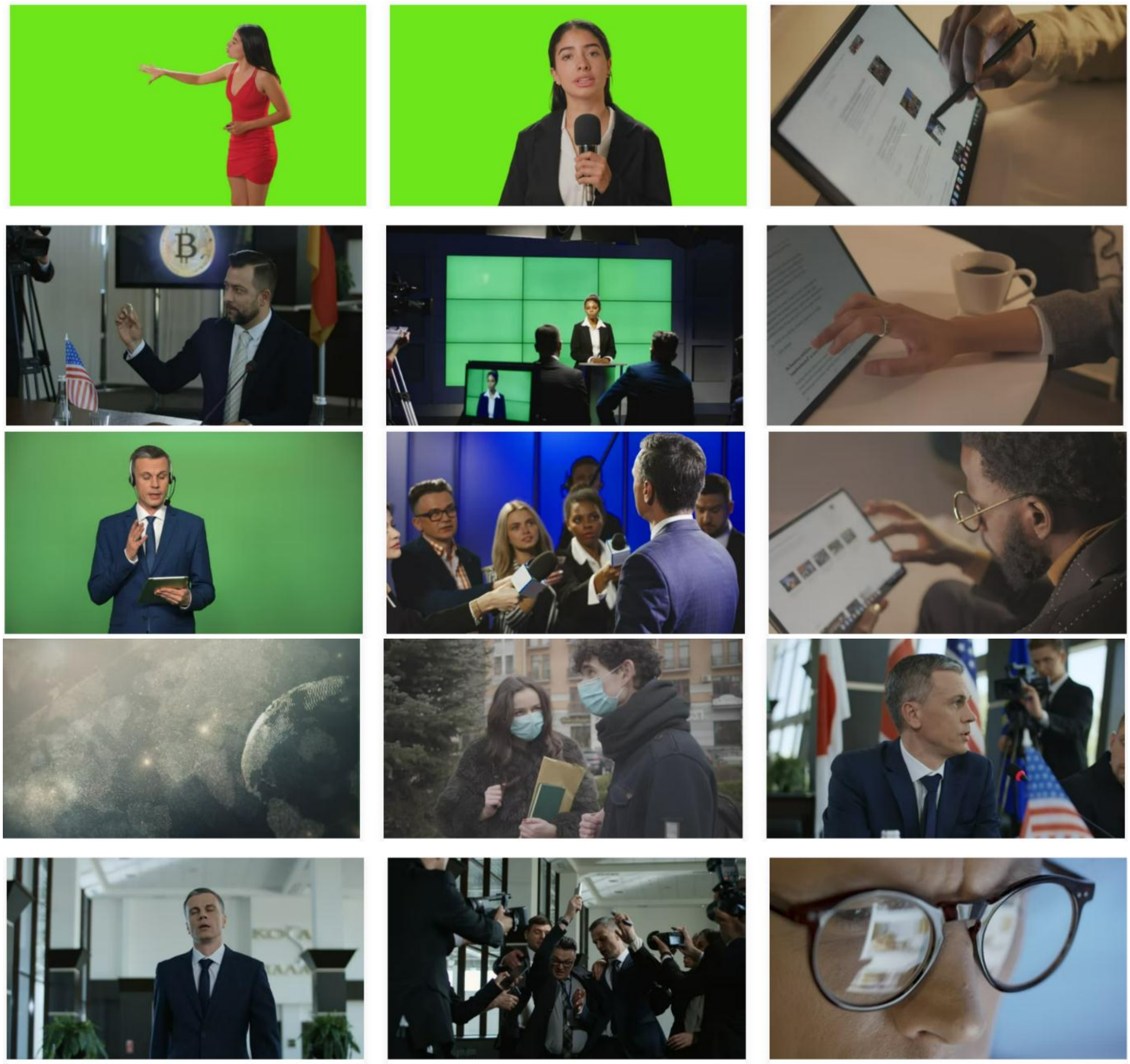
Figure 14: High-Level Syntax of HEVC [7]

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 17, Number 4 (2022) pp. 427-435

5 DATABASE

For this inquiry, we gathered free professional NEWS videos with a resolution of SD (720p) and HD (1080p) and a runtime of 1 minute. There are 50 movies (see Figure 15 for an example). The movie description may be found in Table 1.

Video Parameters	**Values**
No of Videos	50
Duration	60 Seconds
Frame Rate	25 fps
Resolution	1080p, 720p
No of Frames	1500
Format	YUV (4:2:0)



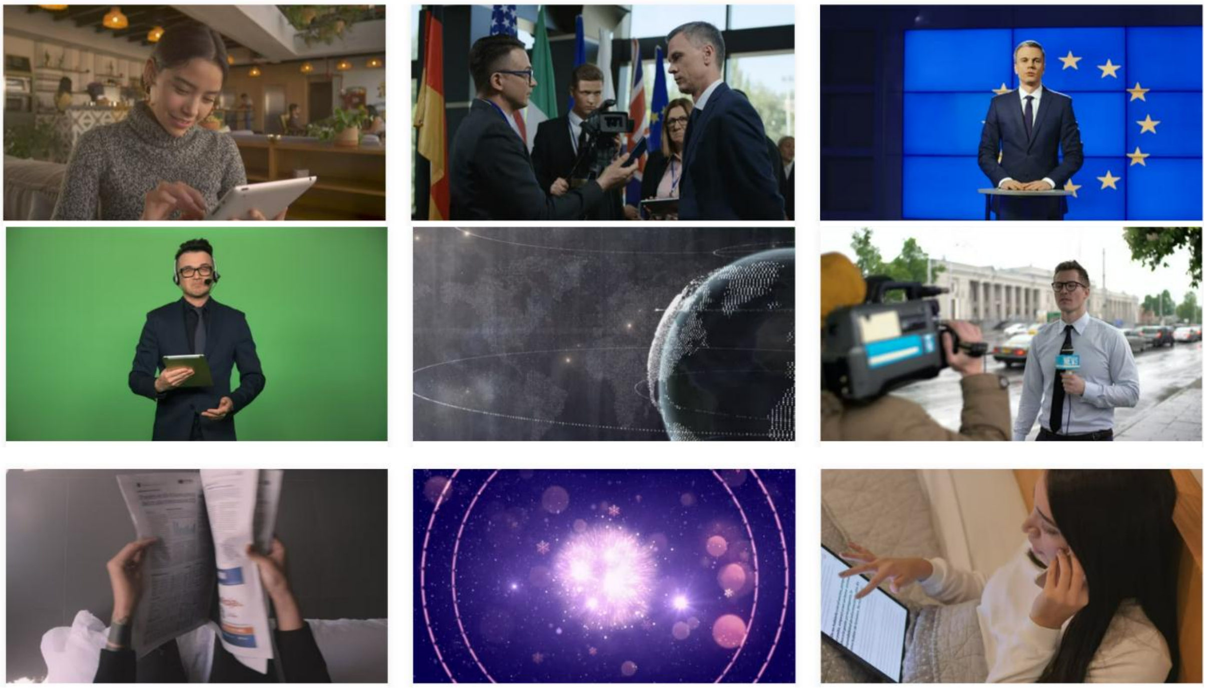


Figure 15: Sample Dataset

6 CONCLUSION AND FUTURE WORK

This project aims to research and develop various NEWS HDR video compression methods, enabling increased visual quality, an integral part of concentration, while complying with the Objective criteria. PSNR improves dramatically with a little increase in bitrate.

REFERENCES

[1] I.E.G. Richardson, "Video Codec Design: Developing Image and VideoCompression Systems," Wiley, 2002.

[2] G. J. Sullivan et al., "Overview of the High-Efficiency Video Coding (HEVC) Standard," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.

[3] V. Sze, M. Budagavi and G. J. Sullivan, "High-Efficiency Video Coding (HEVC) – Algorithms and Architectures," Springer, 2014.

[4] JVT Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264-ISO/IEC 14496-10 AVC), Mar. 2003, JVT- G050 [Online] Available: http://ip.hhi.de/imagecom_G1/assets/pdfs/JVT-G050.pdf

[5] M. Wien, "High-efficiency video coding: Tools and specification," Springer, 2015.

[6] V. Sze and M. Budagavi, "Design and Implementation of Next-Generation Video Coding Systems (H.265/HEVC Tutorial)", IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne, Australia, June 2014 [Online] Available: http://www.rle.mit.edu/eems/publications/tutorials/

[7] Tutorial on HEVC by S. Riabtsev, "Detailed Overview of HEVC/H.265". [Online] Available: https://app.box.com/s/rxxxzr1a1lnh7709yvih

[8] K.R. Rao, D.N. Kim, and J.J. Hwang, "Video Coding Standards: AVS China, H.264/MPEG-4 Part 10, HEVC, VP6, DIRAC and VC-1", Springer, 2014.

[9] A. Luthra and P. Topiwala, "Overview of the H.264/AVC video coding standard", Proceedings of SPIE –The International Society for Optical Engineering, Vol. 5203, pp. 417-431, Applications of Digital Image Processing XXVI, Aug.2003.

[10] HEVC tutorial by I.E.G. Richardson: http://www.vcodex.com/h265.html

[11] HEVC white paper – Elemental Technologies https://www.elementaltechnologies.com/resources/white-papers/hevc-h265-demystified-primer

[12] Access the MPL website: http://www.uta.edu/faculty/krrao/dip/Courses/EE5359/

[13] Cisco White paper on "Visual Networking Index: Forecast and Methodology, 2014–2019",http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[14] M. J. Jakubowski and G. Pastuszak, "Block-based motion estimation algorithms – a survey," Opto-Electronic Review, Vol. 21, pp 86-102, March 2013.

[15] J. Sole et al., "Transform Coefficient Coding in HEVC,"

IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1765-1777, Dec. 2012.

[16] HEVC reference [Online] Available: https://en.wikipedia.org/wiki/High_Efficiency_Video_Coding

[17] M. J. Jakubowski and G. Pastuszak, “Block-based motion estimation algorithms – a survey,” Opto-Electronic Review, Vol. 21, pp 86-102, March 2013.