

State-Of-The-Art In Video Processing: Compression, Optimization And Retrieval

G. Megala^a, S. Prabu^{b*}

^aResearch Scholar, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. E-mail: megala.g39@gmail.com

^{b*}Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. E-mail: sprabu@vit.ac.in

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Video compression plays a vital role in the modern social media networking with plethora of multimedia applications. It empowers transmission medium to competently transfer videos and enable resources to store the video efficiently. Nowadays high-resolution video data are transferred through the communication channel having high bit rate in order to send multiple compressed videos. There are many advances in transmission ability, efficient storage ways of these compressed video where compression is the primary task involved in multimedia services. This paper summarizes the compression standards, describes the main concepts involved in video coding. Video compression performs conversion of large raw bits of video sequence into a small compact one, achieving high compression ratio with good video perceptual quality. Removing redundant information is the main task in the video sequence compression. A survey on various block matching algorithms, quantization and entropy coding are focused. It is found that many of the methods having computational complexities needs improvement with optimization.

Keywords: Codec, MPEG, HEVC, Video Compression, Block Matching Algorithms.

1. Introduction

Video data is the representation of audio with pictorial scene in a digitized form. It is sampled as spatial and temporally in the digital form. Use of social media has become a necessary activity in day today life to access news, information, and interaction and to make decisions. Facebook, YouTube, Instagram, Tumblr, WhatsApp, WeChat, etc., are the most popular social media platform where a billion of users share video files. Netflix, a media service provider of American technology performs streaming of the videos in an adaptive bitrate. According to the users' broadband network connection conditions and speed, it also adjusts the quality of video and audio. Sharing video information concisely in few seconds captures viewer's interest. Video data representation involves sampling spatially within a frame as a rectangular grid and temporally between the sequence of frames at regular time interval. A complete visual scene is sampled at a point to generate a frame which consists of odd and even number of spatially sampled lines. Sampling is repeated at an interval of 0.04 or 0.03 seconds to generate a moving video signal.

Each pixel in a spatio temporal space is represented as a set consisting of luminance (also known as brightness) and chrominance (also known as color). Rather than RGB, $YCbCr$ color space has more advantage where Y (the luma), C_b (blue chroma) and C_r (red chroma). As the human visual is fewer sensitive to color than luminance, the C_b and C_r components can be represented in low resolutions compared to Y. Thus, the data used to represent the chrominance component can be reduced without affecting the visual quality [3]. $YCbCr$ video data representation with the reduced chroma resolution is similar to RGB having no obvious difference. For less storage space and reduced transmission requirements, it is necessary to convert the RGB images to $YCbCr$. Here image compression is performed effectively to represent the chroma components in a low resolution. The process of compressing the video data is known as video coding. For a smooth transformation of images in a video scene, a higher temporal sampling rate of frames is required. Sequence of images called frames that displays at a frequency (as frames per second or fps) known as frame rate. Video coding standards usually support 24fps and 30fps videos. Video coding standards are developed aimed at having high coding efficiency. Video coding is the process of compression and/or decompression of video signal which encodes/ decodes the video data at lowest bit rate without compromising good video quality. There are two quality metrics used to measure the coding efficiency such as Peak Signal to Noise Ratio (PSNR), the Objective metric and video quality, the subjective metric.

1.1. Motivation

The research in video compression is not new and continuing from last almost four decades. Video on demand communication is increasing tremendously in last decade and therefore the area is still attracting the

researchers. Motion Estimation and compensation process is also an old topic, where there is a need of optimization. For these optimization problems, deep learning algorithms, evolutionary algorithms, bio-inspired and nature inspired algorithms can be explored in new aspect overcoming the regular block matching algorithm patterns and improving accuracy. For many decades, research has been done in video compression to provide high compression ratio with good video quality.

2. Video Compression Basics

Gesture video data consists of ordered sequence of group of pictures. Many surveillance applications share the available network bandwidth infrequently with others. Deriving compression techniques, the bit rate of video file is reduced. When the compression ratio is high, low bandwidth is consumed. Increasing compression may also cause increased degradation called as artifacts.

Compression is done by reducing the image data of the video sequence to a reduce media overheads for distributing. Comparing adjacent frames and reducing color resolution with respect to predominant light intensity, removing imperceptibleslices, outliers and noise are performed in the compression technique basically. It results in significant reduced file size for the video sequences which is achieved with visual quality. There are two types of compression techniques involved such as lossy compression and lossless compression. Lossless compression algorithm hat removes the statistical redundancy and allows to extract original data which can be reconstructed from compressed data. It results in limited data reduction by applying limited amount of techniques available. Lossy compression involves original data reduction to some extent, where the original data cannot be reconstructed in decompression process. Latency is the delay or time taken by the applied algorithm to interpret the video data and view the video on display screen. Latency increases when advancement is done in compression algorithm as it compares the adjacent frames. Main objective of video compression algorithms is to attain best compression ratio with the minimized distortion consequently. Removing spatial, temporal and frequency domain redundancy is possible to compress the data meaningfully to a convinced extent of loss of information. Subsequently compression is accomplished by coding schemes such as arithmetic coding and Huffman coding. Discrete cosine transforms (DCT) followed by the process of Motion compensation (MC) are the widely used video coding techniques involved in compression. A block based motion compensation is the motion compensated DCT video coding which is mostly used in H. 26x and MPEG formats.

2.1. Standard Video Codec Lexicon

Following terms are used for understanding video coding standards.

- Frame – Group of picture consists of three different types of frames viz.: I-frame (Intraframe coded picture/ frame), P frame (Interframe predicted picture/frame) and B frame (Bi-directionally predicted pictures).
- Macroblock – It is region of 16 x 16 size which consists of four - Y (luminance) blocks, one - C_R(red Chroma) block and one - C_B(blue Chroma) block. In Chroma processing, the color accuracy is compared to that of the ten bit YC_BC_R for the formats such as 4:4:4 and 4:2:0 [31].
- Block – A region of size 8 x8 in a picture or frame. Discrete Cosine Transform (DCT) is used code the block.
- Motion vector (MV)– A key element in the process of motion estimation. It represents the macroblock in a picture (frame) based on its position placed in another picture known as reference frame. Motion Vector is found by calculating the correspondence amongst the frames at time t and t-1. Vector diagram depicts the direction, magnitude and velocity of moving object. Motion vectors stores the changes in a block. It is represented like a bi-dimensional pointer communicating the predicted macroblock is located left or right or up or down based on the reference frame to the decoder.
- Motion Estimation – A video compression scheme which analysis the two frames and identifies the macroblocks are changed or not with the motion vectors. It examines the moving objects from one frame to another. It exploits the redundancy of vectors between reference frame and subsequent frame by finding the best prediction of macroblock. Process of compressing a video using motion estimation process is referred as interframe coding.
- Motion compensation – Based on the knowledge of moving objects, motion compensation exploits the high correlation between the successive frames of a video sequence.

3. Existing Video Coding Standards or Compression Formats

There are many compression standards previously developed to compress image and videos. MPEG is the basic compression standard associated with digital audio-visual sequences. Following are the collection of compression standard formats:

- H.120 – It is the first video coding standard established by ITU-T organization with a bit rate of 1544 and 2048 Kbps.
- JPEG-2000 – It uses wavelet transformation instead of DCT transformation. It removes the blockiness of JPEG and is replaced with fuzzy picture.
- Motion JPEG 2000 – It is a still picture compression technique used to represent a video sequence with better compression ratio compared to JPEG. This technique is unsuccessful due to its low viewing experience of video stream.
- H.261 or H.263 – It is designed for the purpose of video conferencing thru low bandwidth over telephone line but not suitable for video encoding.
- MPEG-1 – It is developed to code the motion pictures associated with audio. It also provides media storage of 1.5Mbps.
- MPEG-2 – It is designed to code the visual-audio sequence with good video quality with an increased bit rate compared to MPEG-1.
- H.263 – It is a coding standard developed for video conferencing. This codec performs optimization for low data rate with relative small motion.
- MPEG-4- It is used to represent the audio visual data in terms of objects.
- MPEG-7- It is a content description standard of a multimedia to allow fast searching and efficient retrieval. It uses XML to store metadata and does not perform encoding. The video stream contents and the events can be tagged for intellectual processing in video management.
- H.264 – Its goal is to provide high coding efficacy as high compression ratio according to network environments. It is also used to deliver good video quality by means of low and high latency, at low bit rate and high bit rate for low and high resolution. The applications of H.264 includes streaming service or video on demand service, multimedia messaging service and conversational services.
- H.265– It is also known as High Efficiency Video Coding (HEVC) standard [75], which specifies the decoding format along with the encoded video sequence involved in the video compression process. It also defines the syntax of compressed sequence.

Table 1. Video Coding Standards.

| Standards | Purpose | Bit rate | Luminance resolution | Primary Applications | Pros and Cons |
|-----------|--|-----------------|---|--|--|
| MPEG-1 | Outstanding image quality at CD-ROM data rates | 1.5Mbps | SIF: 352 x 240 | Digital storage media | Pros: Good image quality. Cons: High requirements for playback. Requires Licensing fee |
| MPEG-2 | Basic video codec standard for DVD videos having tremendous image visual quality and resolution | 4 – 6 Mbps | 720 x 480 | Broadcast, DVD, HDTV | Pros: Broadcast quality encoded video with good image quality. Cons: It requires licensing fees. |
| MPEG-4 | Coding of audio visual objects Delivers Interactive multimedia at internet data rate across networks. | 20Kbps to 6Mbps | QCIF: 176x144 CIF : 352x288 720 x 480 | Multimedia compression, wireless video phone, Interactive TV, surveillance, Content based storage/ retrieval | Pros: Provides noble image quality at low data rates. Merge natural and synthetic data, robust bitstream syntax and flexible for |

| | | | | | |
|--|---|---|--------------------------------|--|---|
| H.261 | Standard video conferencing codec. | 128-384 kbps | QCIF: 176x144 CIF : 352x288 | Wired video conferencing | interactivity Pros: Low data rates are optimized. It has strong temporal compression component. Cons: Quality is less |
| H.263 | Standard video conferencing codec. | 20 – 384 Kbps | QCIF: 176x144 CIF : 352x288 | Wireless video conferencing | Pros: Advanced MC by half pel motion vectors, overlapping MB and no loop filtering. Good quality compared to H.261 Cons: It is CPU intensive. |
| H.264 or Advanced Video Coding (AVC) | Motion compensation with variable sizes, Image transform, flexible macroblock ordering, Deblocking filters, adaptive VLC coding, weighted prediction, Intra-prediction and slices | 20 Mbps (low frame rates – 24fps and 30fps) to 30 Mbps (high frame rates - 50 and 60 fps) | 1440 p | Web based delivery such as YouTube, Instagram, Twitter, Facebook, etc. | Pros: Provides the finest image quality, with least file size having lower bit rates. It efficiency is double of MPEG-4 and it is used in large application segment. Cons: It requires time-consuming encoding. Considerable amount of storage and bandwidth is required. |
| H.265 or High Efficiency Video Coding (HEVC) | Stream high quality video in congested network by improving coding efficiency. Enable encoding for 4K material in HDR | 120 fps to 300 fps | 3840 x 2160p 8192x4320p | Web-based delivery | Pros: Encode video at low rate with high image quality by reducing 50 % of storage requirement. It allows 10 bit encoding. |

4. Video Compression Essentials

Video sequence consists of group of pictures or frames which further comprises of three different frames IP and B frames as mention in section 2. The Intra-coded frame (I-frame) exploits the spatial redundancy existence between the neighboring pixels within a frame. The Inter- coded frame (P-frame) exploits the temporal redundancy between the consecutive frames in a sequence. Huffman coding is employed to encode I frame on the quantized version of DCT by itself. Moving objects and motion failures are detected by adaptive thresholding and differencing approach done by [32] along with Particle swarm optimization-based Motion estimation to remove temporal redundancy.

Video codec model performs transformation, quantization, block-based motion estimation and motion compensation and entropy coding. Video coding method exploits both temporal and spatial redundancy for achieving high compression. Temporal model, spatial model and entropy encoding are three functional units of video encoder. The entropy encoder compresses the Motion vectors and coefficients factors of temporal and spatial model. Fig. 1 illustrates the system that performs video compression.

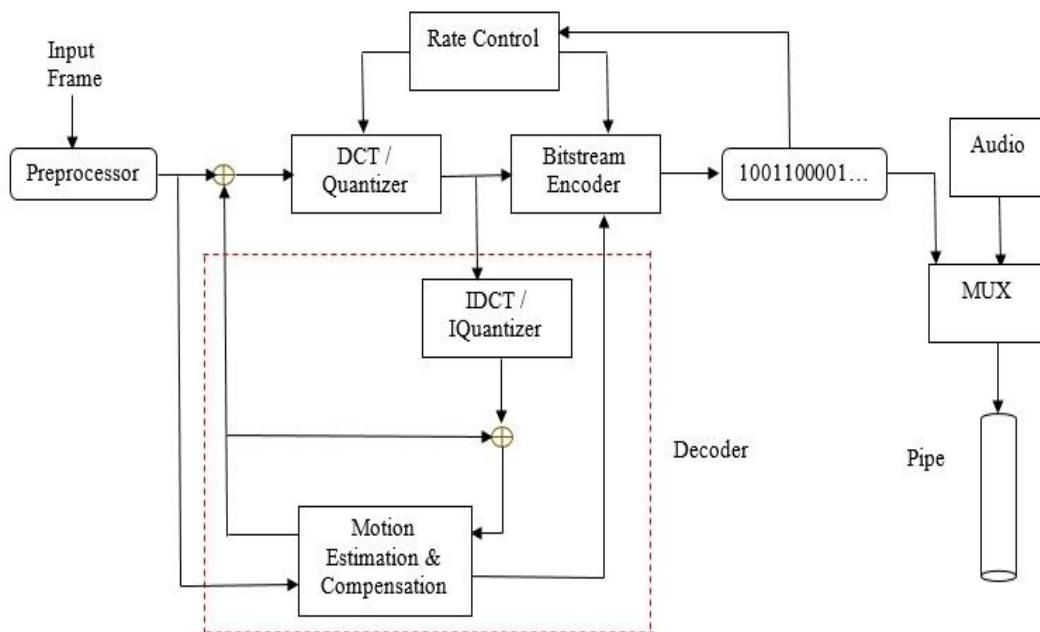


Figure 1. Video Compression System.

Encoder compress the video sequence and creates a compressed form of video bit streams which may be used for storage or transmission. The video encoder involves the steps such as (i) Partition the picture (frame) into multiple units namely, Prediction Unit (PU)[15], Coding Unit (CU), Coding Tree Unit (CTU) (ii) Perform inter and intra prediction by subtracting prediction from unit (iii) Transform and quantize the residual errors (iv) Encode the output of transform, predicted information and header information.

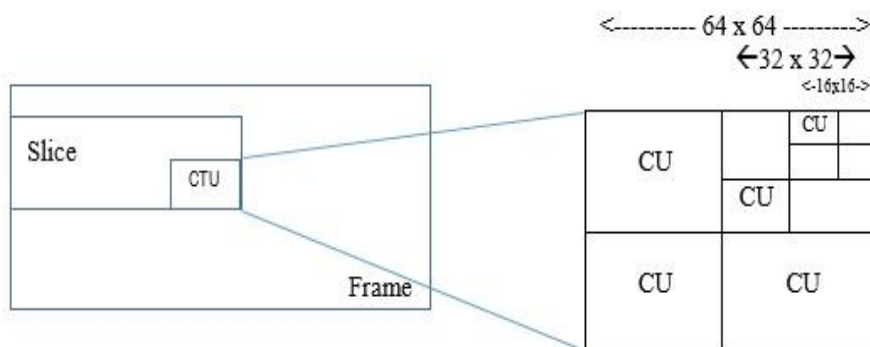


Figure 2. Frame Slice with its Coding Units.

Video decoder decompress the compressed bitstream of video sequences and reconstruct the frames. Steps involved in decoder are (i) Entropy decoding and extraction of sequence elements (ii) Inverse transformation and rescaling (iii) Add prediction to each prediction unit to form the inverse transform output and (iv) reconstruct the decoded sequence.

Each coded video frame, or picture, is partitioned into Slices. Each slices consists of several macroblocks or Coding Tree Units (CTU) having maximum size as 64x64 pixel. It is further divided into Coding Unit (CU) shown in Fig 2. Each CU is further split [9]into single or many Prediction Units (PUs), in which each are predicted as either intra prediction or inter prediction. PU modes and CU combinations [4] are iteratively examined to select the optimal Rate Distortion (RD) cost of CTU [6,17]. Quad tree structure with multi type

recursively partitions the slice and generate flexible block sizes [7,8]. In a video encoder, BjøntegaardDelta - Bit Rate (BD-BR) measurement is accustomed to compare the performance. Its negative value indicates exactly howample the low bit rate is reduced. Whereas the positive value indicates that the bit rate is amplified for thealike PSNR value. CTU rate control [12] estimates the allocation of bits to each CTU by exploiting the correlation between quantization parameters and features. Deep neural network based model improves the efficiency of CTU level rate control video coding.

4.1. Intra Prediction

Each prediction unit is anticipated from its neighboring pixel information within the same frame. The preliminary frame is divided into 8 x 8 blocks in inter prediction coding [20] as shown in Fig. 2. Consequently on each block, the Discrete Cosine Transform (DCT) and quantization is applied to assess the coefficients of digital and alternate current. Then entire the digital and alternate current coefficients are scanned in the zigzag manner to perform scrambling of run length. At latter conversion of entropy is obtained by means of Huffman algorithm. Neural network based intra prediction linear model is built by [19,37] to demonstrate the performance of versatile video coding among conventional methods. Neural network combined with Gray Level Covariance Matrix (GLCM) [5] used for intra prediction with flexible quad tree CU. CU are classified into natural content and scene content based on decision tree model [10] to reduce homogeneous CU splitting complexity. To speed up the process of encoding, affirm deciding CU scheme is adopted; thus skipping the dividing procedure which has smaller threshold value.

4.2. Inter Prediction

Each prediction unit is predicted from its data of neighboring frames using motion compensation. Succeeding frames are then undergone inter prediction coding in order to eliminate the time-based redundancies existence amongst neighboring frames. Frame rate up conversion [11] enhances the original videos' temporal resolutions thus converting the frame rate between different systems by interpolation of frames between consecutive ones. Encoding process of HEVC inter prediction [14,20] computes quadtree partitioning of CTU and evaluates its hierarchy using rate distortion optimization.

4.3. Transformation

DCT and DWT are important multimedia compression schemes using transform techniques. The elementary of conversion to accomplish compression on an image or video is Discrete cosine transform (DCT). This DCT has extensive recognition of signals for compression for its robust dynamism most extensively used to ensure enduring coding in compressing a video. Main purpose of this DCT technique [48] is to transform the M dimension of video or image into N dimension and also converts energy compaction especially short frequency coefficients. Gathering of associated coefficients can also be decorrelated using this cosine transform function.

Video bit streams on applying DCT technique that transforms the residual samples and the resultant factors are quantized. Resultant quantized factors are then traversed in crisscrossed fashionaiming to fragment the coefficients of DCT using macro chunk level dissection into two fold sub streams. Fragmentation of video into frames is done whenever it is sensed by a node [33]. Here every frame in RGB color domain are converted into luminance-chrominance (YCbCr) color domain so as to eliminate the RGB color space redundant components which are storage incompetent and not transmission efficient. An 8 X 8 sized block is selected from the luminance-chrominance color space to undergo DCT transformation by measuring the rows at the end. In this DCT conversion of color space, the original pixels are converted into spatial occurrences for further process. Inverse Discrete Cosine Transform decodes all the spatial frequencies into pixel values. [38] combined the non-sub sampledelineation conversion with Huffman followed by run length coding thus achieving high performance in compression. Distinct shape of pictures are captured by contourlet transform in different directions.

4.4. In Loop Filtering

When the video files are being compressed, there occurs a quality degradation in lossy coding. In-loop filters are utilized in video coding standard towards enriching the eminence of video appearance [63]. To improve video quality, the key frame I frame are sent to the in loop sifting system on an encoder [1]. It reduces the noise distribution in the video frames. Many works have been performed on deblocking filter, CU partitioning, and adaptive filtering. Deep learning techniques [2] are also used to achieve better performance.

4.5. Motion Compensation and Estimation

4.5.1. Motion Compensation

Motion compensation is mainly used to reduce the temporal redundant information of the current frame with reference to reference frame to achieve high compression ratio. MC predicts the current P or B frame based on reference frame [35] and encode the difference in prediction error. Segmentation [16] can also be applied to this motion compensation process in order to accurately predict the object edges in the block.

The stages of motion compensation involves, (i) motion estimation between current frame and previously reconstructed frame (ii) current frame prediction (iii) diverse prediction encoding and original frame error prediction.

4.5.2. Motion Estimation

The process of identifying identical blocks between the consecutive frames (either previous frame or future frames) from the current frame is referred as motion estimation (ME). ME can be performed using various techniques known as inter frame encoding in the video compression procedure. ME is the initial step involved in compression and computes the motion vectors and its displacement values of each pixel. Positive values of motion vectors indicates that the frame moves in right or down. Negative value of MVs indicates that the movement of frames is in left or up.

Block-based ME and motion compensation (MC) are altered according to the structure of video object plane to a random form in MPEG-4 formats. A box like boundary is selected in a video object plane of a frame and the estimation of motion is performed within the MBs. Here the ME of current MB is mostly preferred as block based if the estimation of motion is done within internal MBs. If the estimation of motion is at the boundary then polygon matching, modified block based ME is preferred. Pixels in current MB of the video object plane are used to calculate the falsification measure in this polygon matching. ME is an interframe prediction process which includes pel recursive algorithm and block matching algorithm. Block-matching motion estimation assumes that the objective motion being foreseen is inflexible and non-rotational.

4.5.3. Block based Motion Estimation

Matching the blocks of reference and current frame is the main step involved to detect the temporal and spatial redundancies. Each video sequence consists of several GOPs and each GOP consists of any number of three different frames. I frame indicates the scene change. Frames are split in to macroblocks of size 8x8 or 16x16. The best macroblock of the current frame which perfectly matches with the reference frame are found as shown in Fig. 3 by means of various block based ME algorithms. Quality measures in finding the best matched macroblock is determined via Mean squared Error (MSE), Mean Absolute Difference(MAD), Sum of Absolute Differences(SAD) and Peak Signal to Noise Ratio (PSNR).

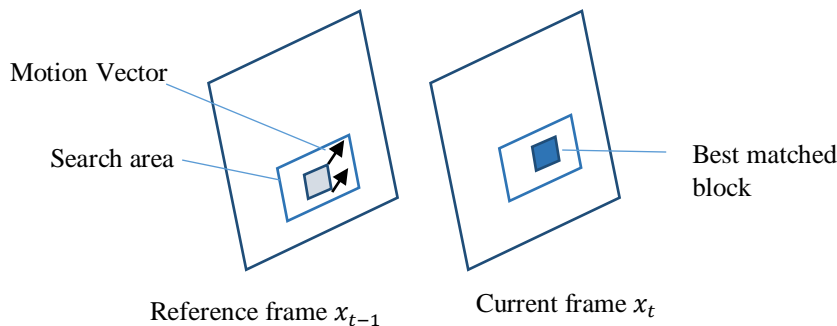


Figure 3. Block Matching.

The similarity measures between the regions of macroblocks of the frame x_t and x_{t-1} are computed as

$$E(d_1, d_2) = \sum_{m_1, m_2 \in N} \varphi(x_t(n_1 + m_1, n_2 + m_2), x_{t-1}(n_1 + m_1 + d_1, n_2 + m_2 + d_2)) \quad (1)$$

Where E is the expected value, (m_1, m_2) is the pixel and (d_1, d_2) is the displacement value.

Search region size of a frame is $(2p + 1) \times (2p + 1)$.

Mean Squared Error (MSE) is the square value of the differences computed between the reference and current frame. The quality is high only when the MSE value is low and reduces the error.

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (f_o(i, j) - f_{rc}(i, j))^2 \quad (2)$$

Here the squared difference between the pixel value (i, j) of the original frame f_o and the reconstructed frame f_{rc} is computed. PSNR is used to measure the quality of reconstructed picture at the decoder. Ratio of the squared maximum possible pixel value (Max_p (i.e. 256)) and Mean Squared Error is calculated. The logarithm of this value is known as PSNR value which is expressed in terms of decibel (db).

$$PSNR = 10 \cdot \log_{10} \left(\frac{Max_p^2}{MSE} \right) \quad (3)$$

Higher the PSNR value indicates high quality is achieved in reconstructed frame. Data Compression Fraction is well-defined as the fraction of uncompressed original video data size and the compressed size of video data. Data representation is reduced to relevant size in compression. Streaming of audio and video data is termed as

$$Compression\ ratio = \frac{Uncompressed\ Data\ rate}{Compressed\ Data\ rate} \quad (4)$$

4.6. Quantization

Quantization is the process of mapping input source symbols to a small set of possible output values in such a way that the reconstructed picture in lossy compression is same as the original. DCT outputs the DCT coefficients to be quantized. Quantization process uses quantization parameter (QP) which decreases the DCT coefficient values. Scalar quantization and vector quantization are the two types of quantization used in compression techniques. [36] Defining the set size of quantization in video coding standard is the important part in providing compression efficiency. When the step size is coarse, compression ratio is high with less quality of reconstructed one. Compression efficiency is less when the step size is small. Scalar quantization is the process of mapping one input signal value to single quantized value of output which is performed along with the transform coding. A scalar quantiser may be uniform (same step size) or non-uniform quantiser (different step size). Vector quantization is the process of mapping vector (set of input values) to a set of quantized values. A survey on various techniques used for quantization is shown in Table 2.

Table 2. Analysis on Quantization Process

| Reference | Concentrated on / Adopted in | Technique | Observation | Merits / Demerits | Limitation & Future work |
|-----------|--|--|--|---|--|
| [21] | High Efficiency Video Coding (HEVC) | Rate Distortion Optimized Quantization (RDOQ) | Quantized coefficient is selected in a low complexity for cost estimation | All zero block detection is performed to reduce complexity | Not implemented in Versatile Video Coding (VVC) |
| [30] | Joint Exploration Model (JEM) | Trellis-coded quantization | Binary partitioning of video frame with intra prediction, Inter-prediction, diffusion filtering and DCT based nonlinear transformation. | Selection of chunk by the encoder is an inspiring chore. Diverse fragment fractions are permissible or restricted in the top level syntax | High computation. |
| [47] | Video Coding Experts Group (VCEG) and MPEG | Classic block-based hybrid video coding design on joint Call for Proposals | Block partitioning, prediction, in loop filtering, transformation, Scaling and entropy coding are performed | Bit rate is decreased to 30% associated to HEVC based on BD-BR metric | Difficult to interpret chroma measures |
| [48] | VCEG and MPEG | Multiparty response towards call for proposals | Prediction, in loop filtering, and entropy encoding is basically performed in compression | Achieves 40% bit rate equivalent on Static Dynamic Range (SDR). Limited usage of kernels for intra prediction of CUs. | MV are not exactly predicted by the coarse representation. |
| [58] | H.264 | Pixel Motion CNN (PMCNN) | Spatio temporal coherence is computed to perform predictive coding. Probabilistic quantization is used to perform binarization. Quantization is replaced with convolutional encoders | PMCNN outperforms spatial prediction and temporal prediction with 48.4% BD-rate | Lack of entropy encoding. Computation complexity is high. |

4.7. Entropy Encoding

Encoding is performed with new schemes to reduce the energy transmission followed by quantization operation. Each frames of the video sequences are encoded so as to reduce the block size and statistical redundant information. Entropy coding determines the least number of bits essential towards representing the information deprived of any distortion. It converts the moving vectors and transform coefficients into compressed stream of bits for storage or transmission. High dimensional multimedia data consumes more time for encoding and decoding process. Variable Length Coding (VLC), Huffman coding and arithmetic encoding

are broadly used entropy encoders. Context adaptive VLC (CAVLC) and context adaptive binary arithmetic coding saves the bit rate in compression. H.264 standard uses CAVLC entropy encoding technique proving low cost complexity. A fast and efficient video compression mechanism is need for video codecs. Machine learning algorithms can be used for encoding high and differing resolutions. Multi-reconstruction recurrent residual network (MRRN) [62,64] extracts the features of artifacts and fed into CNN reconstruction model to achieve content related restoration with different denoised ratio. ML algorithms involves high computation cost, enormous storage space and communication overhead. Thus considering constraints, researchers are motivated to design an enhanced encoding methods. When there is a proliferation in quantized bit stream size in the process of encoding [33], then there is a reduction in average number of bits to be trimmed which in turn obstructs the efficiency of transmission. Therefore in order to increase the average truncation bits count during encoding, those quantized stream of bits can be partitioned as either into two or three or many parts. In this scheme, half of the total length is considered as the quantized bit stream size and so threshold value can be fixed.

Huffman encoding and arithmetic coding are introduced in hardware to parallelize and optimize coding but still bottlenecks such as complex computation exists. A Huffman tree is built using the heap data structure and stored in the memory block which is further used to generate codeword lengths for every symbol. Variable length codes are concatenated with the encoded stream of bits in Huffman decoding which is more difficult. The maximum length of codeword is 255 and shares the prefix codes with canonical code. The main drawback of classical approach is that it looks up the memory several times the canonical codes are extracts the communal prefix code. Various entropy coding algorithms in compression is comparatively analyzed and shown in Table 3.

Table 3. Comparative Analysis of Algorithm on Entropy Encoding or Compression.

| Reference | Algorithm | Strategy adopted | Advantages / Drawbacks |
|-----------|---|--|--|
| [13] | Adaptive compression algorithm and residual restoration by several hypothesis algorithm | Adaptive sampling method is used to eliminate redundancies by utilizing limited samples and normalized Bhattacharyya co efficiencies (NBC). | The definite whole amount of samples is stable within the target, and the quantity of samples for each block do not surpass the block sampling maximum. |
| [23] | Video Coding for Machines (VCM) | Video coding optimization and coding the features Optimization of video and feature coding for machine and human vision. | Domain prediction generalization and adaptation are to be improved. |
| [24] | Weighted Entropy encoding with optimized quantization matrix | Weighted encoding and quantization operation is performed to optimize the quantization matrix | Default intra QM is performed in HEVC due to frequency dependent scaling. |
| [25] | Neural network-based compression | Recurrent Neural Network Based Coding, Random Neural Network Based Coding, Generative Adversarial Network Based Coding | Compression is performed by combining the visual info with semantics thus formulating high effective representation of signals. High computational complexity and more memory consumption |
| [26] | Dual CNN networks with shared parameters | Type of interaction is recognized. Deep temporal or spatial features are extracted and learned spatial and the temporal models with Long Short Term Memory (LSTM) networks | Prediction of interaction is limited within a group. Emotion recognition is not done. |

| | | | |
|------|--|--|---|
| [27] | High Dynamic Range video compression algorithm with optimization | Perceptual transfer function is used to code the dynamic range frame and for reproduction of chroma accurately, error minimization scheme is used. | Normalization factor is the maximum pixel value which cause issues with high luminance pixel |
| [28] | Video compression framework based on spatiotemporal resolution adaptation | CNN super resolution for up sampling of spatial resolution. Decision based on quantization resolution. Decoder reconstructs complete resolution video | High complexity in computation |
| [29] | Up graded fusion encoding with cuckoo search algorithm | Local search and the global search is balanced as a result of avoiding local optima with the help of mutation operator | Slow convergence rate with less accuracy exists |
| [34] | tabled Asymmetric Numeral Systems (tANS) compression algorithm | Parallel architecture performs high when compared to sequential in decoding involving stack, decoding table and symbol decoding with the header information of each chunk | For its high efficiency it is mostly used by face book and apple. Has achieved throughput 200 Mbps with less resources. But it requires more memory and larger decoding tables. |
| [44] | Temporal fluctuation reduced video encoding and spatial texture preserved video encoding | Encoding temporal fluctuation of stable background and spatial domain fluctuation of dynamic foreground | Improves accuracy of detecting objects in inter and intra frames with less distortion |
| [49] | Generative Adversarial Network | Down sampling of the blocks are performed in prediction. After coding the low bit rate the signals are then upsampled in decoder to convert to match original resolution producing good video quality. | Higher network training performance and flexibility Feature extraction is more complex and time consuming in concluding efficient features. |
| [51] | Spatiotemporal Knowledge distillation | Distills the inherent knowledge from a complex to simple on low resolution dynamic saliency estimation | Redundancies at inter and intra level are removed step by step producing high accuracy with low computational cost. Temporal cues are computationally expensive. |

4.8. Optimization Techniques

Optimization can be defined as the procedure of resolution refinement so as to find and achieve the best efficient methodology. Evolutionary algorithms, nature inspired algorithms [41], swarm intelligence algorithms are the most preferred optimization techniques.

Optical movement, matching the identical blocks [42,43], recursive methods, various transform techniques and adaptive PSO techniques are used for attaining motion assessment optimization. ME based video compression saves bit by sending less entropy encoded images to a fully coded frame. ME process is the most computational cost expensive and resource extensive operation. Hence a fast and computationally inexpensive

algorithms for ME is needed. Harmony search, simulated annealing and ray optimization techniques are physics or chemistry based algorithms. Table 4 shows the comparison of different block matching algorithms with its search points and PSNR value for the inputted Claire video sequence.

Table 4. Comparative Analysis of Block Matching Algorithms.

| Algorithm | Description | Average search points | PSNR value | Advantages / Drawbacks |
|---|--|-----------------------|------------|--|
| Full search | Searches $(2P+1) \times (2P+1)$ position to result finest match | 184.6 | 38.94 | Unsuitable for real time video coding as it is time consuming |
| New Three Steps Search | Center biased pattern search with minimum MAD | 15.1 | 38.94 | Small motions are not efficiently identified |
| Simple and Efficient TSS SESTSS | Searches a quadrant of minimum error location | 16.13 | 38.89 | Less PSNR value |
| Diamond search | Large and small Diamond search pattern are used | 11.6 | 38.94 | Two search patterns are utilized |
| CUCKOO search block matching algorithm [43] | ME uses levy flight and breeding behavior of flies to generate new solutions from the existing | 8.8 | 44.02 | ME is performed based on the cuckoo and levy flight behavior of finding optimal. It might signify worst solution if cuckoos egg is found not healthier |
| Harmony Search Differential Evolution | Fitness approximation is incorporated to estimate MVs | 5.1 | 43.9 | Crossover and mutation operator increases diversity and tried to resolve differential evolution drawbacks. |
| Differential Evolution based Block Matching | Minimizes Sum of Absolute Differences (SAD) error and computes structural similarity (SSIM) | 16.6 | 42 | No stable exploration pattern |
| Artificial Bee Colony-Differential Evolution based BM | Developed for acceptable block matching solution with minimum number of SAD evaluations. | 5.5 | 43.85 | Combines search space of DE with ABC |
| Four Step Search | Employs center biased search and halfway stop | 14.8 | 38.94 | Searching techniques are half way stop procedures |
| Adaptive Rood Pattern Search ARPS | Shares similar MVs of a macroblock for temporal coherence prediction | 5.2 | 38.94 | Increasing macroblocks size decreases the PSNR value |
| Artificial Bee Colony-Differential Evolution based BM | Population based minimal bee foraging model searches with minimal MSE | 5.5 | 43.8 | Minimum number of SAD evaluations. Fails to consider spatio-temporal correlation. Searches food source several times |
| PSO Based ME Algorithm | High correlation between MVs are exploited | 18.8 | 34.81 | Searching is fast. Exploit high correlation between MVs and update it. |

5. Video Retrieval

Video can be represented as moving objects in spatial and time; which consists of pictures (frames), audio, metadata and captions. Automatic indexing and retrieving the video unit is the recent topic in the multimedia research. Splitting a video into several potential units is performed by segmentation to improve indexing. High definition video compression, video summarization can be performed using video Object Segmentation [22] and Video object Tracking mechanism. [61,64] categorized content based video indexing and retrieval as four concepts such as; Video segmentation, indexing, dimension decrease and machine learning (ML) techniques. Segmentation of video into shots, the fundamental unit is performed to extract the necessary information

describing the frames and shot boundary detection [46]. Redundant frames are eliminated and keyframes are identified or video summarization. Segmentation is performed on approaches such as cut based, machine learning, color based and entropy based. The feature are extracted [45] from color based approach, visual approach, motion based and texture approach. A few algorithms such as k-means, Principal Component Analysis (PCA) and document frequency based are used to reduce the dimensionality. Many machine learning algorithms used for CBVIR are Nearby neighbors based techniques, RF, Softmax, SVM, NN, MV based methods, regression, learning based, tree based, density based, probability boosting methods, querying and labelling methods, clustering techniques such as Fuzzy and hierarchical methods.

5.1. Dataset

There are many widely accessible datasets such as VideoVis, a biomedical educational video dataset (website- <https://patentq.njit.edu/oer>), TRECVID meetings dataset available in website (<http://www-nlpir.nist.gov/projects/trecvid/trecvid.data.html>). TRECVID 2005 (news), KTH (human action recognition), UCF (Youtube and sports) are the dataset were the CBVIR method is most applied.

Quality of Experience (QoE) [72,74] is measured based on the user satisfaction of the retrieved videos from the social cloud. QoE can be assessed in two ways such as subjective and objective QoE. Video degradation falls under subjective QoE, whereas technical data loss impact in video quality comes under objective QoE. Videos of.wmv and.mp4 formats with 240P and 360P respectively can be downloaded from YouTube website <https://www.tubeoffline.com/> [73]. When these videos are uploaded to a social cloud, it is automatically compressed to decreasing quality with reduced file size for storage. Parameters of videos such as frame width, frame height, data rate, bit rate, frame rate, audio bit rate, storage size, length, type of codec and video file type are used to describe a video file. [40] measured the video quality parameters thru divergent compacted and distorted video sequences.

Social clouds such as tumblr and Qzone video provided good quality of service when compared to youtube with increased QoE and less noise, less blur in low transmission bit rate. Cloud resources performs transcoding as a service [39] to reduce the complexity at user level. Cloud server transcoder splits the videos into chunks of different size having GOPs, process each chunks individually and merge them to a compressed one. Mobile video stream platform [50] can also be created based on collaboration with cloud environment to process the video. A large amount of data transmits when the resolution of the frames are high. Mobile devices and edge devices having dynamics transmission rate are handled by adaptive algorithm implemented in Netvision[55]. It processes the on demand video scheduling problem, query and optimize the response time. [57] built a lookup table, history tree and map table to improve query processing response time by maintaining space complexity. A summary on video compression and retrieval approaches with its basic quality metrics used are shown in Table 5.

Table 5. Summary of Video Compression and Retrieval Approaches.

| Reference | Codec | Preprocessing | ME & MC | Quantization | Rate Control | Streaming | Evaluation metrics | Quality measures | Benchmark |
|-----------|--------------|---------------|---------|--------------|--------------|-----------|-----------------------|---|-------------|
| [52] | H.265 | ✓ | | | | | KLD, LCC, RSS and ROC | PSNR, AWS-PSNR | Salient 360 |
| [53] | H.264/AVC | | ✓ | | | | BD rate | PSNR | MPEG |
| [54] | H.265 / HEVC | | | ✓ | ✓ | | RDO bit rate | Block size sustained intra mode detection (BSSIM D) | self |
| [66] | H.266 | | ✓ | | | | BD BR complexity | PSNR, WS PSNR | JVET |
| [67] | H.265 | ✓ | ✓ | | | | BD-BR | SPSNR | JVET |

| | | | | | | | | | |
|------|---------------|--|---|---|---|---|---------------------------------------|---------------|---------------------|
| [68] | H.265 | | ✓ | | | | BD-BR | WS-PSNR | JVET |
| [69] | H.265 | | ✓ | | | | BD-BR | WS-PSNR | JVET |
| [70] | H.265 | | ✓ | | | | Others | PSNR | Self |
| [71] | H.265 | | | ✓ | | | BD-BR, RD curve | PSNR, WS PSNR | JVET |
| [56] | AVC and H.265 | | | | ✓ | | R-D curve | Quad log cost | JPEG |
| [58] | H.264 | | | ✓ | | | BD-Rate | PSNR/MS-SSIM | MPEG/VEG |
| [59] | HEVC | | | | | ✓ | BD-Rate | PSNR and SSIM | Vimeo-90k |
| [60] | 3D-HEVC | | | | | ✓ | R-D curve | | 3D-HEVC |
| [65] | HEVC | | | | | ✓ | Spearman Rank correlation coefficient | SSIM, MS-SSIM | LIVE video database |

6. Challenges and Future Directions

6.1. Challenges

As additional video traffic will concern more sequences at 4K resolution, there exists some challenges.

- Pre-sequence transcoding pecking: Fixing of resolutions and bitrates of the media by the providers is performed here. Single size does not fit for all media services. Movies have different bit rates combined with perceived quality metrics and reducing bit rate of encoded video is the major research area to provide video services without affecting the quality.
- Espousing new standards: Newly developed video standards provide high compression ratio compromising the video quality. Compression performance of new standards triggers transcoding demand.

6.2. Future Directions

Though many papers are reviewed, many of the methods found unable to identify the redundant features, correlated frames from the original frame to perform transformations. Machine learning algorithms are more useful in content based video retrieval preserving the geometric structures and features. Semantic labelling video indexing can be considered which could explore the correlations thus improving annotation performance.

It is predicted that deep learning based compression of image or video might be a pivotal good quality videos represented in less bit rate. Following are the provoked issues necessarily be further examined.

- Decision making during encoding or motion estimation is a complicated problem during the compression standard generation. To tackle this problem, learning based algorithms such as active learning, reinforcement learning, deep learning, ensemble learning and transfer learning can be applied thus improving the performance of coding operation.
- Memory and computation efficient video codec design, semantic-kind oriented video compression, compacting the feature descriptors and visual content with deep learning based framework. CNN model compression is also a multi-variable optimization problem, which should be optimized jointly considering computational cost, CNN performance [18,19] and rates utilized for CNN transmission.
- Automatic compression of media file in social clouds adds noise, blocking effects, blurring, information loss during quantization and drop of frames in a sequence of video. Degradation of video decreases the quality of experience of the videos at the user side.

Many NIA algorithms are implemented to perform optimization in motion estimation. This paved a path to do research on combination of NIA algorithms along with learning concepts to enhance the performance by computing motion vectors. In future, Artificial bee colony and cuckoo search based DE hybridization can be started as the base work of ME optimization using nature inspired algorithms.

7. Conclusion

This paper provides the study on popular video compression standards, fundamental needed and operations performed to compress a video. Quality comparative analysis with its evaluation metrics are discussed. Lot of aspects in prediction, variable block size coding, compression artifacts, perceptual and semantic based processing and optimization of block estimation can be further improved for high resolution video sequences. Deep learning techniques can be effectively used for computing temporal coherence in predictive coding but time and memory complexity are high. Various block matching algorithms in motion estimation are studied. Artificial bee colony and cuckoo with differential evolution are found to provide best matching in motion estimation with high PSNR values. These techniques alleviates the resources, degradation of quality and time consumption. In summary block matching optimization along with deep learning based motion estimation and compensation is the major exploration topic considering the computational complexity and time complexity.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

References

1. Sun, W., He, X., Chen, H., Sheriff, R.E., & Xiong, S. (2020). A quality enhancement framework with noise distribution characteristics for high efficiency video coding. *Neurocomputing*, 411, 428-441.
2. Li, T., Xu, M., Zhu, C., Yang, R., Wang, Z., & Guan, Z. (2019). A deep learning approach for multi-frame in-loop filter of HEVC. *IEEE Transactions on Image Processing*, 28(11), 5663-5678.
3. Yang, R., Xu, M., Liu, T., Wang, Z., & Guan, Z. (2018). Enhancing quality for HEVC compressed videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(7), 2039-2054.
4. Liu, X., Li, Y., Liu, D., Wang, P., & Yang, L.T. (2017). An adaptive CU size decision algorithm for HEVC intra prediction based on complexity classification using machine learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1), 144-155. <https://doi.org/10.1109/TCSVT.2017.2777903>
5. Huang, C., Peng, Z., Chen, F., Jiang, Q., Jiang, G., & Hu, Q. (2018). Efficient CU and PU decision based on neural network and gray level co-occurrence matrix for intra prediction of screen content coding. *IEEE Access*, 6, 46643-46655.
6. Yan, T., Ra, I.H., Wen, H., Weng, M.H., Zhang, Q., & Che, Y. (2020). CTU layer rate control algorithm in scene change video for free-viewpoint video. *IEEE Access*, 8, 24549-24560.
7. Yang, H., Shen, L., Dong, X., Ding, Q., An, P., & Jiang, G. (2019). Low complexity CTU partition structure decision and fast intra mode decision for versatile video coding. *IEEE Transactions on Circuits and Systems for Video Technology*.
8. Fan, Y., Sun, H., Katto, J., & Ming, E.J. (2020). A Fast QTMT Partition Decision Strategy for VVC Intra Prediction. *IEEE Access*, 8, 107900-107911.
9. Zhang, Q., Wang, Y., Huang, L., & Jiang, B. (2020). Fast CU Partition and Intra Mode Decision Method for H. 266/VVC. *IEEE Access*, 8, 117539-117550.
10. Badry, E., Inoue, K., & Sayed, M.S. (2020). Decision Tree Models and Early Splitting Termination in Screen Content Extension of High Efficiency Video Coding. *IEEE Access*, 8, 143437-143452.
11. Yoon, S.J., Kim, H.H., & Kim, M. (2018). Hierarchical extended bilateral motion estimation-based frame rate upconversion using learning-based linear mapping. *IEEE Transactions on Image Processing*, 27(12), 5918-5932. doi:10.1109/TIP.2018.2861567.
12. Marzuki, I., Lee, J., & Sim, D. (2020). Optimal CTU-Level Rate Control Model for HEVC Based on Deep Convolutional Features. *IEEE Access*, 8, 165670-165682. doi: 10.1109/ACCESS.2020.3022408.
13. Xu, Y., Xue, Y., Hua, G., & Cheng, J. (2020). An Adaptive Distributed Compressed Video Sensing Algorithm Based on Normalized Bhattacharyya Coefficient for Coal Mine Monitoring Video," in *IEEE Access*, vol. 8, pp. 158369-158379, doi: 10.1109/ACCESS.2020.3020140.
14. Erabadda, B., Mallikarachchi, T., Kulupana G., Fernando A. (2020). iCUS. Intelligent CU Size Selection for HEVC Inter Prediction. *IEEE Access*, vol. 8, pp. 141143-141158, doi: 10.1109/ACCESS.2020.3013804.

15. Fu, B., Zhang, Q., & Hu, J. (2020). Fast prediction mode selection and CU partition for HEVC intra coding. *IET Image Processing*, 14(9), 1892-1900. doi: 10.1049/iet-ipr.2019.0259.
16. Wang, Z., Wang, S., Zhang, X., Wang, S., & Ma, S. (2019). Three-Zone Segmentation-Based Motion Compensation for Video Compression. *IEEE Transactions on Image Processing*, 28(10), 5091-5104.
17. García-Lucas, D., Cebrián-Márquez, G., & Cuenca, P. (2020). Rate-distortion/complexity analysis of HEVC, VVC and AV1 video codecs. *Multimedia Tools and Applications*, 1-18.
18. Ma, D., Zhang, F., & Bull, D.R. (2020). Video compression with low complexity CNN-based spatial resolution adaptation. In *Applications of Digital Image Processing XLIII* (Vol. 11510, p. 115100D). International Society for Optics and Photonics.
19. Liu, C., Sun, H., Katto, J., Zeng, X., & Fan, Y. (2020). A Convolutional Neural Network-Based Low Complexity Filter. *arXiv preprint arXiv:2009.02733*.
20. Reuze, K., Philippe, P., Deforges, O., Hamidouche, W. (2016). Intra prediction modes signalling in HEVC," *2016 Picture Coding Symposium (PCS)*, Nuremberg, pp. 1-5, doi: 10.1109/PCS.2016.7906387.
21. Wang, M., Fang, X., Tan, S., Zhang, X., & Zhang, L. (2020). Low Complexity Quantization in High Efficiency Video Coding. *IEEE Access*
22. Yao, R., Lin, G., Xia, S., Zhao, J., & Zhou, Y. (2020). Video Object Segmentation and Tracking: A Survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(4), 1-47.
23. Duan, L.Y., Liu, J., Yang, W., Huang, T., & Gao, W. Video coding for machines: A paradigm of collaborative compression and intelligent analytics. *arXiv preprint arXiv:2001.03569*.
24. Kumar, S., Manjunath, A. S., & Christopher, S. High efficient video coding using weighted entropy with optimized quantization matrix. *Journal of King Saud University-Computer and Information Sciences*. (2017).
25. Ma, S., Zhang, X., Jia, C., Zhao, Z., Wang, S., & Wang, S. Image and video compression with neural networks: A review. *IEEE Transactions on Circuits and Systems for Video Technology*. (2019)
26. Afrasiabi, M., Khotanlou, H., & Gevers, T.: Spatial-temporal dual-actor CNN for human interaction prediction in video. *Multimedia Tools and Applications*, 1-20. (2020)
27. Mukherjee, R., Debattista, K., Rogers, T. B., Bessa, M., & Chalmers, A.: Uniform color space-based high dynamic range video compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(7), 2055-2066. (2018).
28. Afonso, M., Zhang, F., & Bull, D. R.: Video compression based on spatio-temporal resolution adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1), 275-280. (2018)
29. Feng, Y., Jia, K., & He, Y.: An improved hybrid encoding cuckoo search algorithm for 0-1 knapsack problems. *Computational intelligence and neuroscience*. (2014)
30. Pfaff, J., Schwarz, H., Marpe, D., Bross, B., De-Luxán-Hernández, S., Helle, P., & Nguyen, T.: Video Compression Using Generalized Binary Partitioning, Trellis Coded Quantization, Perceptually Optimized Encoding, and Advanced Prediction and Transform Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(5), 1281-1295. (2019)
31. Azimi, M., & Pourazad, M. T.: A Novel Chroma Processing Scheme for Improved Color Accuracy of HDR Video Content. *IEEE Transactions on Broadcasting*. (2019)
32. Sengar, S. S., & Mukhopadhyay, S.: Motion segmentation-based surveillance video compression using adaptive particle swarm optimization. *Neural Computing and Applications*, 1-15. (2019)
33. Banerjee, Rajib., Sipra Das Bit. Low-overhead video compression combining partial discrete cosine transform and compressed sensing in WMSNs. *Wireless Networks*, 25, 5113-5135. (2019)
34. Najmabadi, S.M., Tran, T., Eissa, S. An Architecture for Asymmetric Numeral Systems Entropy Decoder - A Comparison with a Canonical Huffman Decoder. *J Sign Process Syst* 91, 805-817. <https://doi.org/10.1007/s11265-018-1421-4> (2019)
35. Silveira, D., Povala, G., Amaral, L., Zatt, B., Agostini, L., & Porto, M.: Efficient reference frame compression scheme for video coding systems: algorithm and VLSI design. *Journal of Real-Time Image Processing*, 16(2), 391-411. <https://doi.org/10.1007/s11554-015-0551-1>(2019)
36. Hussain, A. J., & Ahmed, Z. A survey on video compression fast block matching algorithms. *Neurocomputing*, 335, 215-237. (2019)
37. Santamaria, M., Blasi, S., Izquierdo, E., & Mrak, M. Analytic Simplification of Neural Network Based Intra-Prediction Modes for Video Compression. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-4). IEEE. (2020)
38. Anandan, P., Manikandan, A., Sabeenian, R. S., & Bharathidhasan, D. Nonsampled Contourlet Transform based Video Compression using Huffman and Run Length Encoding for Multimedia Applications. *International Journal*, 9(3), (2020)
39. Koziri, M. G., Papadopoulos, P. K., Tziritas, N., Loukopoulos, T., Khan, S. U., & Zomaya, A. Y. Efficient cloud provisioning for video transcoding: Review, open challenges and future opportunities. *IEEE Internet Computing*, 22(5), 46-55. (2018)

40. Karim, S., He, H., Laghari, A. A., Memon, K. A., Khan, M., & Magsi, A. H.: The evaluation video quality in social clouds. *Entertainment Computing*, 35, 100370. (2020).
41. Choudhury, H. A., Sinha, N., & Saikia, M. Nature inspired algorithms (NIA) for efficient video compression—A brief study. *Engineering Science and Technology, an International Journal*. (2019)
42. Bhattacharjee, K., Kumar, S., Pandey, H. M., Pant, M., Windridge, D., & Chaudhary, A. An improved block matching algorithm for motion estimation in video sequences and application in robotics. *Computers & Electrical Engineering*, 68, 92-106. (2018)
43. Bhattacharjee, K., & Kumar, S.: A novel block matching algorithm based on Cuckoo search. In *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)* (pp. 1-5). IEEE. (2017)
44. Kong, L., & Dai, R.: Efficient video encoding for automatic video analysis in distributed wireless surveillance systems. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(3), 1-24. (2018)
45. Xiang Zhang, Siwei Ma, Shiqi Wang, Xinfeng Zhang, Huifang Sun, and Wen Gao. A joint compression scheme of video feature descriptors and visual content. *IEEE Transactions on Image Processing* 26, 2 (2017), 633–647. (2017)
46. Lingchao Kong and Rui Dai. Object-detection-based video compression for wireless surveillance systems. *IEEE Multi Media* 24, 2 (2017), 76–85. (2017)
47. Bross, B., Andersson, K., Bläser, M., Drugeon, V., Kim, S. H., Lainema, J., & Yu, R.: General Video Coding Technology in Responses to the Joint Call for Proposals on Video Compression with Capability Beyond HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(5), 1226-1240. (2019)
48. Choi, K., Chen, J., Park, M. W., Yang, H., Choi, W., Ikonin, S., & Choi, N.: Video Codec Using Flexible Block Partitioning and Advanced Prediction, Transform and Loop Filtering Technologies. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(5), 1326-1345. (2020)
49. Zhang, Y., Kwong, S., & Wang, S.: Machine learning based video coding optimizations: A survey. *Information Sciences*, 506, 395-423. (2020)
50. Sun, H., Yu, Y., Sha, K., & Lou, B. mVideo: Edge Computing Based Mobile Video Processing Systems. *IEEE Access*, 8, 11615-11623. (2019)
51. Li, J., Fu, K., Zhao, S., & Ge, S. Spatiotemporal knowledge distillation for efficient estimation of aerial video saliency. *IEEE Transactions on Image Processing*, 29, 1902-1914. (2019)
52. Xu, M., Li, C., Zhang, S., & Le Callet, P. State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing*, 14(1), 5-26. (2020)
53. Kim, J., Im, J., Rhyu, S., & Kim, K. 3D Motion Estimation and Compensation Method for Video-Based Point Cloud Compression. *IEEE Access*, 8, 83538-83547. (2020)
54. Srinivasan, A., & Rohini, G. Performance based algorithms for video bit transmissions. *Cognitive Systems Research*, 56, 179-191. (2019)
55. Lu, Z., Chan, K., Uргаonkar, R., Pu, S., & La Porta, T. NetVision: On-Demand Video Processing in Wireless Networks. *IEEE/ACM Trans. Netw.*, 28(1), 196-209. (2019)
56. Haghghat, M., Mathew, R., Naman, A., & Taubman, D. Illumination estimation and compensation of low frame rate video sequences for wavelet-based video compression. *IEEE Trans. Image Process*, 28(9), 4313-4327. (2019)
57. Aldwairi, M., Hamzah, A. Y., & Jarrah, M. Multi PLZW: A novel multiple pattern matching search in LZW-compressed data. *Computer Communications*, 145, 126-136. (2019)
58. Chen, Z., He, T., Jin, X., & Wu, F. Learning for video compression. *IEEE Trans Circuits Syst Video Technol*, 30(2), 566-576. (2019)
59. Lu, G., Zhang, X., Ouyang, W., Xu, D., Chen, L., & Gao, Z. Deep Non-Local Kalman Network for Video Compression Artifact Reduction. *IEEE Trans. Image Process*, 29, 1725-1737. (2019)
60. Li, L., Li, Z., Zakharchenko, V., Chen, J., & Li, H. Advanced 3D motion prediction for video-based dynamic point cloud compression. *IEEE Trans. Image Process.*, 29, 289-302. (2019)
61. Spolaôr, N., Lee, H. D., Takaki, W. S. R., Ensina, L. A., Coy, C. S. R., & Wu, F. C. A systematic review on content-based video retrieval. *Engineering Applications of Artificial Intelligence*, 90, 103557. (2020)
62. L. Yu, L. Shen, H. Yang, L. Wang and P. An, "Quality Enhancement Network via Multi-Reconstruction Recursive Residual Learning for Video Coding," in *IEEE Signal Processing Letters*, vol. 26, no. 4, 557-561, doi: 10.1109/LSP.2019.2899253. (2019)
63. Leo Willyanto Santoso, Bhopendra Singh, S. SRajest, R. Regin, Karrar Hameed Kadhim, "A Genetic Programming Approach to Binary Classification Problem" *EAI Endorsed Transactions on Energy*, Vol.8, no. 31, pp. 1-8, (2021). DOI: 10.4108/eai.13-7-2018.165523
64. Feng, Y., Zhou, P., Xu, J., Ji, S., Wu, D.: Video big data retrieval over media cloud: A context-aware online learning approach. *IEEE Trans. Multimed.* 21 (7), 1762–1777. (2019)

65. Liqun Lin, Shiqi Yu, Liping Zhou, Weiling Chen, Tiesong Zhao, Zhou Wang, PEA265: Perceptual assessment of video compression artifacts, *IEEE Trans. Circuits Syst. Video Technol.* (2020)
66. R. Ghaznavi-Youvalari and A. Aminlou, "Geometry-based motion vector scaling for omnidirectional video coding," in *Proc. IEEE Int. Symp. Multimedia*, pp. 127–130. (2018)
67. Y. He, Y. Ye, P. Hanhart, and X. Xiu, "Motion compensated prediction with geometry padding for 360 video coding," in *Proc. IEEE Visual Commun. Image Process.*, pp. 1–4. (2017)
68. L. Li, Z. Li, X. Ma, H. Yang, and H. Li, "Advanced spherical motion model and local padding for 360° video compression," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2342–2356. (2019)
69. B. Vishwanath, T. Nanjundaswamy, and K. Rose, "Rotational motion model for temporal prediction in 360 video coding," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, pp. 1–6. (2017)
70. F. De Simone, P. Frossard, N. Birkbeck, and B. Adsumilli, "Deformable block-based motion estimation in omnidirectional image sequences," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, pp. 1–6. (2017)
71. X. Xiu, Y. He, and Y. Ye, "An adaptive quantization method for 360-degree video coding," *Proc. SPIE*, vol. 10752, Art. no. 107520X. (2018)
72. http://www.compression.ru/video/quality_measure/video_measurement_tool.html
73. <https://www.tubeoffline.com/>
74. <https://www.videohelp.com/software/MediaInfo>
75. <https://www.vcodex.com/hevc-an-introduction-to-high-efficiency-coding/>